



MARINE
BIODIVERSITY
RESEARCH

Prediction and Management of
Australia's Marine Biodiversity

2011

CERF Report on Transition and Extension Program



Marine Biodiversity Hub

Prof Nic Bax, University of Tasmania

May 2011

nic.bax@csiro.au

www.marinehub.org



TABLE OF CONTENTS

WORKPLAN.....	1
HARVEST PROGRAM 1:	5
Taking advantage of existing products and knowledge generated by the Hub to improve uptake in support of improved management	5
TASK: Report on national biodiversity maps including advice on number of clusters needed in each region.....	5
HARVEST PROGRAM 2:	21
Further development of market based instruments to support structural re-adjustment options.....	21
TASK: Fishers adaptation strategies.....	21
TASK: Threatened species distribution	24
TASK: Further develop market based instruments to support structural re-adjustment options	27
EXTENSION PROGRAM 1:	33
Extending national data holdings in preparation for future marine biodiversity research to support implementation of marine bioregional plans	33
TASK: Harvesting of polychaete diversity data	33
TASK: Paleogenetic analyses.....	36
TASK: Comparative investigation of phylo-geographic patterns of connectivity	38
TASK: Sponge biodiversity data for north-west Australia	40
TASK: Extension of biological analysis for Carnarvon Shelf	42
TASK: Gap analysis	44
TASK: New biological and physical data layers	53
EXTENSION PROGRAM 2:	57
Extend statistical methods developed for marine biodiversity mapping to support further application to meet SEWPaC needs	57
TASK: Uncertainty for biodiversity prediction	57
TASK: Extension of species archetypes method	61
TASK: Extension of hybrid GLM/TREE approach	63

WORKPLAN

HARVEST PROGRAM - Development of further products from existing research

Summary and outcomes

Harvesting for the Marine Biodiversity Hub (MBH) consisted of:

- a summary report and guidelines on national biodiversity maps that were out of scope of the initial agreement but were compiled following a request from the Department of Sustainability, Environment, Water, Population and Communities (SEWPaC).
- further development of Market Based Instruments (MBIs) to support alternatives to structural adjustment as requested by SEWPaC staff
- joint MBH/SEWPaC workshops to bring expertise from Hub partners and beyond to bear on management questions defined by SEWPaC (Marine, ERIN and Heritage)
- continuing communication through 1-2 newsletters focused on relevant issues
- additional opportunities as determined through our knowledge broker and SEWPaC staff

Task	Leader	Deliverable	Page
HARVEST PROGRAM 1			
Taking advantage of existing products and knowledge generated by the Hub to improve uptake in support of improved management			
Report on national biodiversity maps including advice on number of clusters needed in each region.	Roland Pitcher (CSIRO)	Summary report on national biodiversity mapping	5

HARVEST PROGRAM 2			
Further development of market based instruments (MBIs) to support structural readjustment options			
Fishers' adaptation strategies	Chris. Wilcox (CSIRO)	Application of existing model to scenarios provided by the East planning team <ul style="list-style-type: none"> • investigation of fishers' adaptation strategies to MPA designations in the East Planning Region and their implications for economic impacts on the fishing industry 	21
Threatened species distribution	Chris. Wilcox (CSIRO)	Refinement and validation of existing model to provide predictions of at-sea distributions of loggerhead turtles <ul style="list-style-type: none"> • using novel data sources to predict the distributions of threatened marine vertebrates 	24
Further develop MBIs to support structural adjustment	Chris. Wilcox (CSIRO)	Targeted development as requested by SEWPaC <ul style="list-style-type: none"> • further develop market-based instruments to support development of structural adjustment options 	27

EXTENSION PROGRAM – Research aligned to future CERF program directions

Summary and outcomes

The Marine Biodiversity Extension research program focused on two projects:

- Bringing together new national datasets to prepare for improved national maps of biodiversity,
- Extending statistical methods to enable improved predictions of biodiversity and single species distributions.

Task	Leader	Deliverable	Page
EXTENSION PROGRAM 1			
Extending national data holdings in preparation for future marine biodiversity research to support implementation of marine bioregional plans			
Harvesting of polychaete diversity data	Robin Wilson (MV)	Develop North Australia polychaete dataset from Hub and other data sources and based on a consistent taxonomy. (Eunicidae, Nereididae, Phyllodocidae, Polynoidae, perhaps Spioinidae)	33
Paleogenetic analyses Extinction–speciation	Madeleine van Oppen (AIMS)	Report on utility of approach to develop hotspots of speciation and extinction <ul style="list-style-type: none"> • extend paleogenetics work on cryptic species identified by the MBH to identify hotspots of speciation and extinction 	36
Comparative investigation of phylo-geographic patterns of connectivity	Tim O'Hara (MV)	Extend genetic analyses of regional ophiuroid populations analyses to tropical areas, including the Coral Sea to delineate populations and identify levels of connectivity	38
Sponge biodiversity data for North-West Australia	Brendan Brooke (GA)	Develop a sponge database that is consistent at the OTU (Operational Taxonomic Unit) level for sponge collections in Northwest Australia to support detailed biodiversity mapping in this area	40
Extension of biological analysis of Carnarvon Shelf samples to include four additional phyla (ascidians, crustaceans, echinoderms & molluscs).	Peter Doherty (AIMS)	Extend Carnarvon Shelf survey analyses to include additional animal phyla (ascidians, crustaceans, echinoderms, and molluscs).	42
Gap analysis Gap identification for marine biodiversity in both physical and biological space	Roland Pitcher (CSIRO)	Report on gaps in datasets provided to SEWPaC and marine scientists in Australia <ul style="list-style-type: none"> • identify gaps in regional and national physical and biological datasets that are utilised to support implementation of marine bioregional plans 	44
New biological and physical data layers Scope & collate additional readily available surrogate layers - new satellite and modelled layers	Roland Pitcher (CSIRO)	Incorporation of new data layers into Hub publically available data directory for future biodiversity mapping <ul style="list-style-type: none"> • identify and collate new available surrogate layers for use in predicting national biodiversity, in particular new satellite and derived layers 	53

EXTENSION – Research aligned to future CERF program directions

Contd

Task	Leader	Deliverable	Page
EXTENSION PROGRAM 2			
Extend statistical methods developed for marine biodiversity mapping to support further application to meet SEWPaC needs			
Uncertainty for biodiversity prediction	Roland Pitcher (CSIRO)	Report on method and uncertainty for random forest approach to producing national biodiversity maps <ul style="list-style-type: none"> develop a method for representing uncertainty for the predictions of compositional biodiversity patterns from the modified random forest approach used to generate national biodiversity maps in support of marine bioregional planning 	57
Extension of species archetypes method	Piers. Dunstan (CSIRO)	Production of R code and report on example analysis and critical comparison of groups from a variety of sources <ul style="list-style-type: none"> extend the modelling framework that underpins the creation of species-archetype groups to produce species-archetypes from abundance and biomass data types, and improve variable selection methods, with potential application to threatened species mapping 	61
Extension of hybrid GLM/TREE approach	Bill Venables (CSIRO)	Production of annotated R code for statistical community <ul style="list-style-type: none"> extend the new hybrid GLM/Tree approach to support use of varied datasets in biodiversity mapping and other analyses 	63

Hub partners (abbreviations):

AIMS – Australian Institute of Marine Science

CSIRO – Commonwealth Scientific and Industrial Research Organisation

GA – Geoscience Australia

MV – Museum Victoria

UTAS – University of Tasmania

HARVEST PROGRAM 1:

Taking advantage of existing products and knowledge generated by the Hub to improve uptake in support of improved management

This program will enable continued delivery and communication of the knowledge and expertise developed by the Marine Biodiversity Hub, through communication, workshops, knowledge brokering and a summary report.

TASK: Report on national biodiversity maps including advice on number of clusters needed in each region.

This report describes the approaches taken by the CERF Marine Biodiversity Hub to predict biodiversity from environmental surrogates and provide a national coverage of biodiversity maps to the former SEWPaC for Australia's large marine planning regions. Two complementary aspects of biodiversity were mapped: structure and composition, using two new methods developed by the Marine Biodiversity Hub. In the case of composition, SEWPaC requested that the outputs be clustered, rather than a continuous representation of composition, raising the question of how many clusters (predicted assemblages) there should be in each marine region. This question is addressed in this report. In addition, patterns of biodiversity in the South-East Marine Region were analysed and mapped.

Introduction

Spatial Marine Management is primarily concerned with the management and conservation of national marine assets. A significant proportion of these assets can be defined as some aspect of biodiversity. However, despite significant research around Australia, data on most aspects of biodiversity are sparse making comprehensive spatial management impossible. Biodiversity surrogates (i.e. the physical environment) have been proposed and used to offset the lack of biological data. However, species and assemblages respond to environmental gradients in different ways, meaning that raw surrogates do not represent biodiversity well. Providing improved inputs to management requires the prediction of biodiversity attributes and modelling the biological response of species and assemblages to the physical surrogates. This allows the development of models that can be used to predict into areas that have not been sampled, using the more extensive data available for physical surrogates to inform the models.

Biodiversity is more than the sum of its parts. It is the number, variety and unique identity of species living in an area — the variety of life. By its definition it is all-encompassing, making rigorous scientific analysis difficult. Key to the development of methodologies to

describe biodiversity is understanding what aspects of biodiversity are interesting and informative for management.

There are a large number of potential attributes of biodiversity (Figure 1). Some of these have been

regional maps of biodiversity for the (then) Department of Environment, Water, Heritage and the Arts (DEWHA, now SEWPaC) and interested NGOs (eg WWF). The Marine Biodiversity Hub used two of the methods that it had been developing to produce the national coverage of maps. These methods analysed biodiversity structure, using Rank Abundance Distributions (RAD; Foster and Dunstan 2010, Dunstan and Foster, in press),

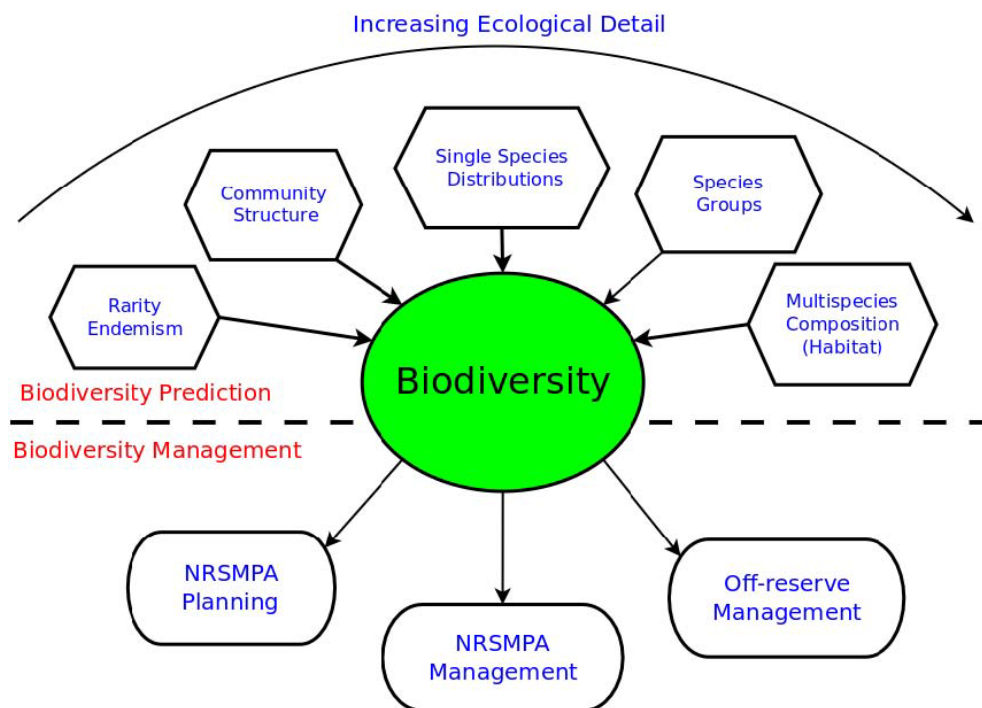


Figure 1: Aspects of biodiversity

relatively well studied (i.e. single species distributions) but other attributes have received little or no attention. The CERF Marine Hub has been actively engaged in developing methods that will allow a more expansive view of biodiversity to be used in spatial marine management (Figure 1). Each attribute of biodiversity will require a different approach and will potentially describe different ecological responses of species and assemblages. As a key output, the Marine Hub produced a series of

and biodiversity composition, using Gradient Forests (Ellis et al 2010, Pitcher et al 2010). Predictions and national maps were delivered for 4 of the 5 marine regions around Australia, the South-West, North-West, North and East Marine Planning regions.

National Mapping – rank abundance distributions

Identifying priority areas for conservation is a key process in managing marine ecosystems. One of the fundamental attributes of biodiversity is species richness (i.e. number of species), and spatial patterns of species richness have been used extensively to identify biodiversity 'hotspots'. In the marine context, species richness is difficult to use as the only indicator of

hotspots. This is due to the fact that most marine surveys collect an overwhelming number of rare species. Williams et al. (2010) found that over 50% of species found in samples from Western Australia were found in only one sample. In this case it is unlikely that concepts of complementarity can be used to manage the rare species since it would be optimistic to assume that rare species would be distributed in a similar way to the more common species. Biodiversity hotspots are typically defined as regions with higher than average species richness (eg Pearman 2007, Guilhaumon 2008, Parviainen 2009). However, richness by itself is only one component of community structure. We have used two of the predicted components of biodiversity, species richness and evenness (η) to identify regions that have both large numbers of rare species and regions that have rare combinations of species richness and evenness. *(Species evenness is estimated from the relative abundance of each species in an area or community.)*

We have taken a recently developed statistical method that analyses and predicts Rank Abundance Distributions (RADs, Foster and Dunstan 2010; Dunstan and Foster in press) and used this to identify areas that are hotspots for biodiversity. Rank Abundance Distributions are a ubiquitous feature of all biological samples where abundances are counted. They allow the characterisation of

total abundance, species richness and relative abundance/evenness. We have developed methods to analyse RADs with respect to environmental gradients and predict them over broad geographic areas. Environmental gradients refer to changes in physical/chemical characteristics of the marine environment through space, such as temperature, salinity, sediment grain size or depth. Rank Abundance Distributions (RADs) are analysed by decomposing the RAD curve into three attributes of biodiversity, total abundance (M), species richness (S) and relative abundance (η). We calculate a measure of community evenness (η), derived from the relative abundances. Relative abundances are modelled as a logarithmic decreasing curve and the derivative of this curve at species rank one indicates how uneven the sample is. Values of η can range between 0 (all species have the same abundance) to $-\infty$ (one species only, i.e. complete dominance). A single value representing relative abundance allows mapping of spatial predictions. Models are fitted to each of these attributes using the methods outlined in (Foster and Dunstan 2010) and then predicted onto a grid of points using environmental covariates at each point (Dunstan and Foster, in press). We have used the CERF Marine Biodiversity Hub datasets and covariates to develop predictive models in four of the marine planning regions, the South-West, North-West, North and East Marine Planning Regions. In each region we fitted models to predict total abundance, species richness and relative abundance and predicted to all the points within the region of interest.

Identifying regions with high numbers of rare species is achieved by identifying predicted locations with both high species richness and low evenness (i.e. $0 \gg \eta$). The RAD curves in these locations have a few highly abundant species and many species with only 1 or 2 individuals. The 'rare' species identified may not be truly rare (i.e. range restricted or small populations), and could be found more abundantly in other locations or by other sampling methods. However, in general there is a strong correlation between species richness

and species that were only ever seen in a single sample, particularly for invertebrates which had more species overall.

The other type of hotspot identified is regions with an infrequent combination of richness and evenness. To identify regions like this, predictions of richness and evenness must be generated for the entire region of interest and split into categories. This is to simplify the complex distributions into outputs that can be used for spatial planning. To categorise the bivariate distributions of richness and evenness for fish and invertebrates, the predictions are independently partitioned into 5 classes for both richness and evenness. Both richness and evenness were split into categories at every 20th percentile (ie breaks at 0, 20, 40, 60, 80, 100th percentile of the distributions). This establishes 25 (5 from richness and 5 from evenness) categories for community structure that can be spatially plotted and are easier to interpret than the full bivariate distribution. Rare combinations of richness and evenness can be identified as the categories that have the smallest number of points in the region of interest. For an example for the North-West Marine Region, see Figure 2.

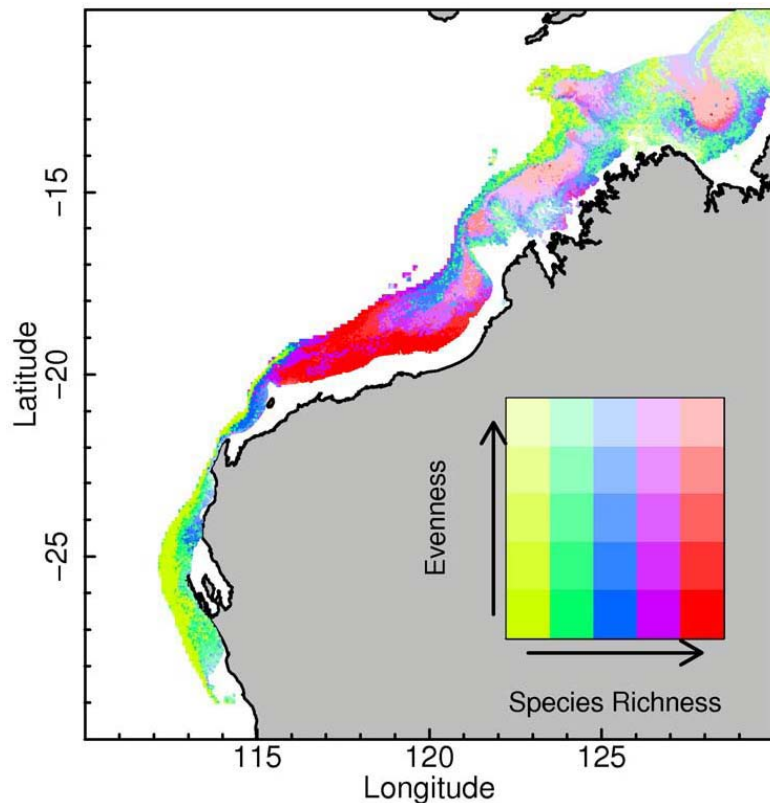


Figure 2: Spatial distribution of RAD categories in North West Australia

National Mapping – Gradient Forests

In order to more fully explore changes in species composition along *gradients* of environmental variables, we applied a newly developed method ("Gradient Forest", Ellis *et al.*, 2010), based on Random Forests (Breiman, 2001) to 4 Australian bioregions: the South-West, North-West, North and East Marine planning regions, and more recently the fifth marine region, the South-East. The overall extent to which environmental variables can predict distribution patterns, and the relative importance of variables, can be quantified by the standard Random Forest method. Gradient Forest adds a highly flexible, non-linear method to explore the shape or thresholds, and magnitude of changes in composition of multiple species along continuous gradients of environmental variables. Moreover, because the gradient response is quantified in comparable units (R^2), information from analyses of multiple datasets and different sampling devices can be combined and compared. Thus, Gradient

Forest addresses several shortcomings in the use of surrogates.

The statistical approach was based on the 'machine learning' method Random Forests (Breiman 2001). This is a partitioning tree method that finds the best splits on the predictors that minimises the sums-of-squares of the species abundance in the child nodes (i.e. maximises the fit improvement), but instead of a single decision tree, the R (R Development Core Team 2008) package *randomForest* (Liaw & Wiener 2002) forms a forest of trees (500 trees, in our case) to avoid the instability of single trees and provide a smoother more stable result. Each tree in the forest is formed from an in-bag sample of about $\frac{2}{3}$ (0.632) of the observations at random, each split is selected from a subset of $\frac{1}{3}$ of the predictors at random, and the fit of each tree is tested against the remaining out-of-bag samples. The analysis was run on each species, in each survey dataset, that had sufficient frequency of occurrence (i.e. having >5 unique abundances). The overall fit over the forest, and the predictive performance was indicated by the proportion of data variance explained (R^2).

The importance of each variable for prediction accuracy was assessed; however, quantifying predictor importance is a vexed issue, as often the environmental predictors are correlated. The standard approach in random forests assesses *marginal* importance by randomly permuting

each predictor in turn, across all sites in the dataset, and calculating the degradation prediction performance of each tree. Unfortunately, this overestimates the importance of correlated predictors, for both influential and spurious variables (Strobl *et al.*, 2008). This problem was addressed by implementing *conditional* permutation for assessing predictor importance in Gradient Forest (Ellis *et al.*, 2010), following the strategy outlined by Strobl *et al.*, (2008). In conditional permutation, the predictor to be assessed is permuted only within blocks of the dataset defined by splits in the given tree on any other predictors correlated above a certain threshold ($r = 0.5$ in our case) and up to a maximum number of splits ($=\text{floor}(\log_2(n*0.368/2))$) in our case, where n =number of sites). This constrains the permutation of a given predictor, making it conditional on other correlated predictors in the tree and reducing the potential for degrading the prediction performance and hence also reducing the importance measure.

The key new modification to the method in Gradient Forest (Ellis *et al.*, 2010) collated the numerous split values along each predictor gradient and their associated fit improvement, for each predictor in each tree, each forest and each species. For each species, the split improvements were first standardised by the density distribution of the survey's observed values of the predictor gradient, to indicate the potential 'true' importance of splits had the gradient been sampled with uniform density. The standardised splits were then normalised to predictor importance, and predictor importances were normalised to R^2 — thus, each split is expressed in proportion of variance explained, and each species contributes to the quantification of compositional change in proportion to its variance explained by the environmental predictors. The normalised splits for all species within a survey dataset were combined along each gradient. The normalised splits frequency distributions quantified where and how much compositional change occurred along each gradient. These normalised splits were aggregated as cumulative distributions of compositional response (in R^2 units) along gradients of each environmental

predictor. Graphically, these represented the cumulative change along each gradient, for each species and for aggregated composition.

The results from the various within-region surveys were combined, to provide an overall cumulative importance curve for each predictor, in each region. This was done by taking a weighted average of the normalised importances within narrow ranges of the predictor (called bins), the weight being proportional to the number of survey sites near the bin and inversely proportional to the variance of the estimate in that bin. The resulting average was then aggregated as cumulative distributions of regional compositional change along each gradient. The overall cumulative importance curves from all regions were also plotted together, for each predictor.

Mapping biodiversity composition

The regional combined cumulative importance curves were used as biologically-informed transformations of the available full-coverage (interpolated) environmental data layers. The transformed layers were mapped to show expected patterns of continuous compositional change of marine biodiversity that are related to multiple gradients of the physical environment. Such maps can be used as a surrogate representation

of expected patterns of biodiversity composition in the absence of complete biological survey data. The transformation converts the different arbitrary scales on which the environmental predictors are measured, which have no direct relevance to biology, to a common biological importance scale (in R^2 units) where the magnitude and non-linear re-scaling within the range, are determined by the responses of the biota in the available datasets and thus more closely reflect the way species-assemblages respond to the gradients. To account for any differences in the ranges and density distributions of predictors in the region to be mapped relative to that of the informing datasets, if required, the cumulative importance curves were extended to the full range using the average slope of the curve, and the overall importance was adjusted by the square-root of the ratio of the 95-percentile ranges of the regional predictor density over the observed density. Once transformed by the cumulative importance curves, the principal components of the transformed environmental predictors were computed, to provide a multi-dimensional space that represents variation in composition, and which is constrained by relationships between the species and their environment. This multi-dimensional compositional space can be mapped in geographic space, most simply by using the first three dimensions as an RGB colour palette, or (in the case of the map products delivered to SEWPaC) by fitting a colour wheel to the first two dimensions of the biologically transformed environment space — either provide continuous representation of patterns of biodiversity composition. The colour key is plotted as a biplot, with vectors showing the direction of the major environmental drivers, to facilitate interpretation of the corresponding geographic map. If required, as in the case of the SEWPaC map products, the continuous compositional space can be clustered to represent expected species-assemblage groups, which can also be mapped in geographic space (for an example for the North-West Marine Region, see Figure 3 on page 11). An issue for the clustered output is determining the appropriate number of clusters, or predicted

assemblages (see *Determining the number of clusters*, below).

the biologically informed environment space, after transformation using the Gradient Forest cumulative

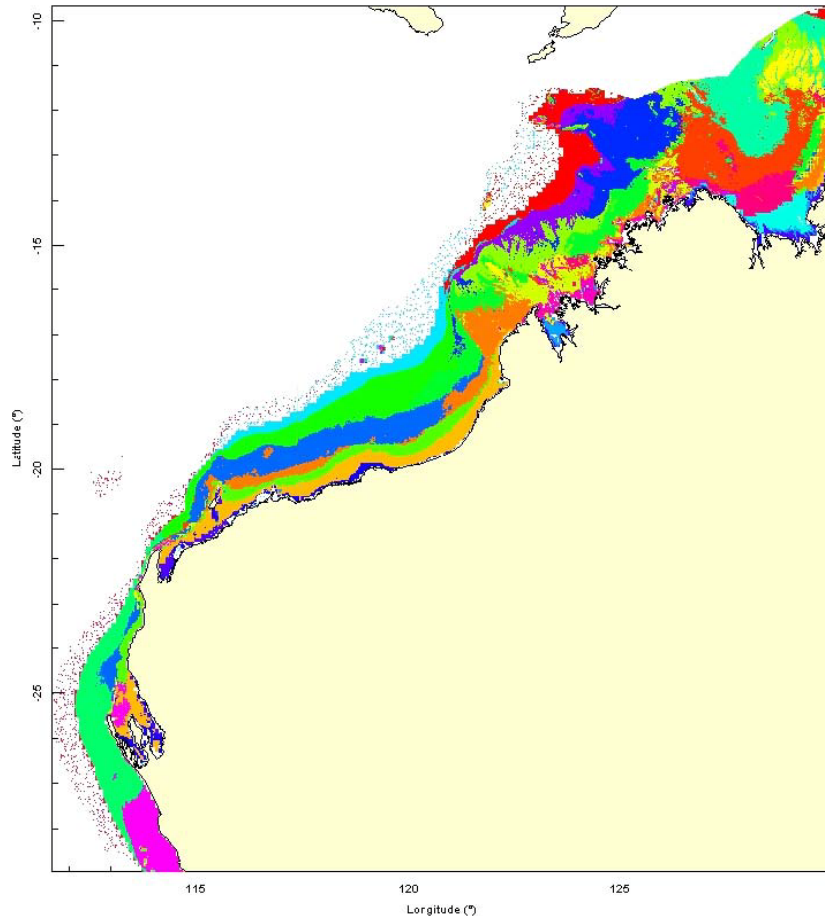


Figure 3: Spatial distribution of 25 gradient forest clusters in North West Australia (note, arbitrary cluster colours used in this figure).

The potential improvement of the “biologically informed” approach to mapping biodiversity, using Gradient Forest, over a purely physical approach to marine bioregional mapping was assessed for sled and trawl datasets from the GBR. As a “benchmark”, the commonly used Bray-Curtis (BC) dissimilarity metric (Bray & Curtis, 1957) was taken as a measure of compositional differences among the sampled sites. These were compared with the Euclidean distances (ED) between the same sites located in

importance distributions. A performance measure was then calculated, analogous to the stress diagnostic of non-metric Multi-Dimensional Scaling (Kruskal, 1964), for the fit of a monotonic regression between the ED in biological space and the BC dissimilarities. Similarly for the untransformed environmental predictors, the BC dissimilarities were compared with the EDs between the same sites located in the physical environment space, after normalising the raw environmental variables to unit variance (i.e. the variables have neutral weight and no re-scaling

within their range). The fits of the monotonic regressions and the stress

diagnostics of the “biologically informed” approach were ~30% and ~50% better than those of the uninformed predictors, for the GBR sled and trawl datasets respectively. This indicated that the biologically informed approach provided an improved representation of the patterns of biodiversity composition, as constrained by relationships with environmental surrogates, compared with unconstrained raw BC dissimilarities.

Determining the number of clusters

In the context of regional biodiversity maps, the main output of the Gradient Forest method is a transformation of environmental variables or

surrogates so that they represent how species react to them (biological space). Every location for which environmental surrogates are available can be mapped in biological space. Locations that are close in biological space are expected to have similar biodiversity composition whereas those further apart are expected to have more different biodiversity composition. Biological space can be used to define a continuous colour scale, which can then be used as a key to a geographical map. Places on a map that are similar in colour are close in biological space and therefore have similar expected biological composition. Though such a continuous colour scale is very useful, some applications require representation of biological space with a discrete number of clusters. This raises the question: "How many clusters?"

This number should be guided by the original biological survey data, rather than attributes of transformed biological space. The solution is not established, particularly for the case of several contributing surveys each having only partial coverage of a region, and several approaches were attempted, starting with the South-East Marine Region.

Initially, a suitable approach was applied to each biological survey separately, i.e. multivariate regression trees (MRT). This can be recommended because an objective number of clusters (i.e. terminal

nodes) can be obtained by partitioning on environmental variables using cross-validation. The resulting number of terminal nodes sets a minimum constraint on the number of clusters in biological space; i.e. the number of clusters in the whole region must be greater than the maximum for any one survey (max_MRT_nodes) because no one survey provided complete coverage of a region.

The first approach involved clustering the regional biological space using *clara* in the R package *cluster*, with number of clusters ranging from \leq max_MRT_nodes up to 60. This set of clusters was matched with each sampled site in each survey. The similarity of each clustering with the cross-validated terminal node assignment of each survey site was then assessed using a range of external (between) cluster measures (similarity index, dot product, Rand, Jaccard, Folkes-Mallows) as available in the R package *clv*. These measures would maximise if and when the regional clustering best matched the MRT nodes for each survey. The result was a saw-toothed pattern with no clear optimum match, but a series of sub-optimal matches for several cluster numbers in the range examined (see Appendix 1, Figure A-1). Further, there was no clear number of clusters that best matched all surveys.

The second approach assessed which regional clustering, taken as a factor, accounted for most variation in the constituent biological survey datasets. This involved linking each candidate clustering back to the biological data using distance-based redundancy analysis (db-RDA, Legendre and Anderson, 1999) as implemented by the function *capscale* in the R package *vegan*. The db-RDA method consists of two stages: first, it performs ordination on the Bray-Curtis dissimilarity matrix between all site pairs in a survey; second, the resulting site configuration undergoes redundancy analysis with sites constrained to the clusters. The db-RDA generates a multivariate F ratio statistic, a large value of which would indicate strong evidence that the clustering has captured structure in the site configuration. This analysis was carried out over the same series of clusterings of the regional biological

space as used in the first approach. The clustering with the largest F ratio is to be preferred on biological grounds. However, for each region there are multiple surveys contributing to the analysis, each one providing a separate F ratio, and so the survey results must be combined in some way. Typically the F ratios differed by a factor of 10 or more, and so, to guard against any one survey dominating the result, the geometric mean was used as the diagnostic.

In this case, for the SEMR, there was a maximum F -ratio at 15 clusters of the regional biological space (see Figure 4). Potentially, this may be a

than the number of MRT nodes in any constituent survey.

Nevertheless, 15 clusters was not a standout result and several other cluster numbers could be selected (eg 12,18,23...) depending on the level of complexity or information content required. Further, the results were less clear for the other marine regions (South-West, North-West, North, East), and so a number of other guides was also examined.

Other guides included auto-associative MRT where the regional biological space is used as both the dependent and independent data matrix, and splitting of the tree is stopped by setting the complexity parameter $cp=0.01$. The latter setting typically corresponds closely with the level of complexity where cross-validation indicates that splitting should stop.

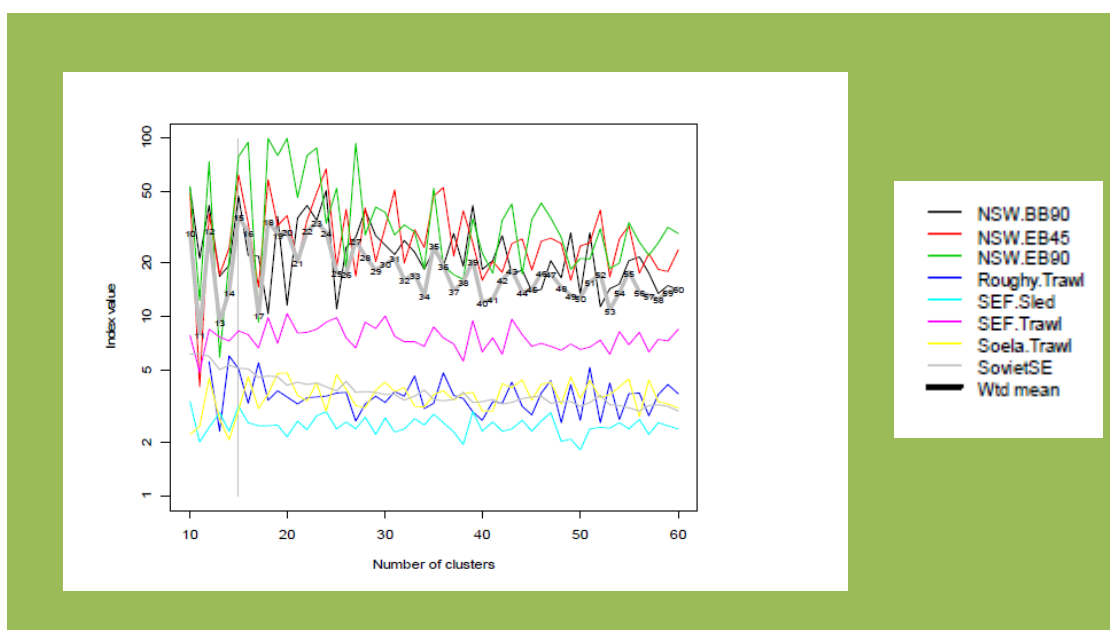


Figure 4: db-RDA F-ratio results for the SEMR surveys, showing a maximum at 15 clusters of the regional biological space

statistically objective procedure for recommending a number of clusters for a region: take the number of clusters that maximises the geometric mean of the F ratios over all surveys subject to being greater

While less statistically objective, this approach has the benefit of providing a similar level of complexity in splitting of the biological space for all marine regions. In the case of the SEMR, auto-associative MRT suggested 20 terminal nodes.

A final ad-hoc guide involved identifying the number of regional clusters that simply provided the best numerical match to the number of MRT nodes for each dataset. In the case of the SEMR, the closest overall match was 26 or 31 clusters, although neither provided a close match to the number of MRT nodes for any individual survey.

least diverse data, with fewer species; mostly of larger fish trawl species that tend to have less spatial structuring.

Comparing predictions

The results of Gradient Forest and RAD analyses clearly indicate that biodiversity cannot be viewed through a single lens. The analyses provide complementary information: one, (RAD), gives an indication of how species vary spatially, and the other, (Gradient Forest),

Table 1. Summary results from various approaches to provide a guide to the appropriate number of clusters for each marine region. The selected number of clusters is shown in bold.

Region	dataset max MRT nodes	max F-ratio	other F-ratio peaks	F-ratio peak > max MRT	autoMRT cp=0.01	closest n clust match n MRT nodes
EMR	11	10	10, 15 ,18,20	15	9	18,20
SEMR	11	15	12, 15 ,18,23,27,31	15	20	26,31
SWMR	6	5	5, 8 ,10,13	8	8	4,5
NWMR	5	7	7,14,17,20,26	7	10	7,8
NMR	7	5	5,7, 13 ,18,20,22,24	13	11	4,5

The results for other regions are shown in Table 1. In each case, a number of clusters was selected (shown in bold) for mapping of each region (see Appendix 1, Figure A-2), although there was not strong evidence that the selected number represented a clear optimum and in each case, alternative numbers of clusters could be selected. The SWMR and the NWMR were the least clear and also had the lowest number of indicated clusters. This could reflect the level of information content in the constituent surveys. For example, the SWMR had the

gives an indication of how the species that are present will allocate resources and respond to potential disturbances. There are similarities between the gradient forest clusters and the RAD categories. For example, in the NWMR, RAD category 1 and gradient forest cluster 5 overlap on 37% of the total points in both analyses (see Figure 5 on page 15). There are a number of strong overlaps between the gradient forest and RAD analyses and a visual comparison suggests gross overall similarity (compare Figures 2 & 3). To a certain extent there should be some overlap in both analyses as the same biological and physical data were used to generate each of the results.

There are also substantive differences. The gradient forest analysis addresses beta diversity – how the species composition of communities differ from each

other. Species composition will change along the coast with a wide variety of environmental gradients. Beta diversity is useful to examine how differences in large groups of species emerge. In contrast, RADs analyse how a community is structured and how this structure differs with environmental gradients. It quantifies how many species might be found in an area and how their abundances vary.

In some circumstances it may be expected that a particular group of species will generate a unique structure; in those circumstances we should expect that RAD and beta diversity will be spatially congruent. However, the same compositional assemblage of species could show a variety of RAD responses, depending on the environment. In some locations one species may be favoured and be highly dominant, in others the abundances will be more similar. Visual comparison of the predictions for north-western Australia show areas of similarity and differences. Regions in the Timor Sea and west of Exmouth show similar patterns, but there are large areas between these regions where the overlap is weaker.

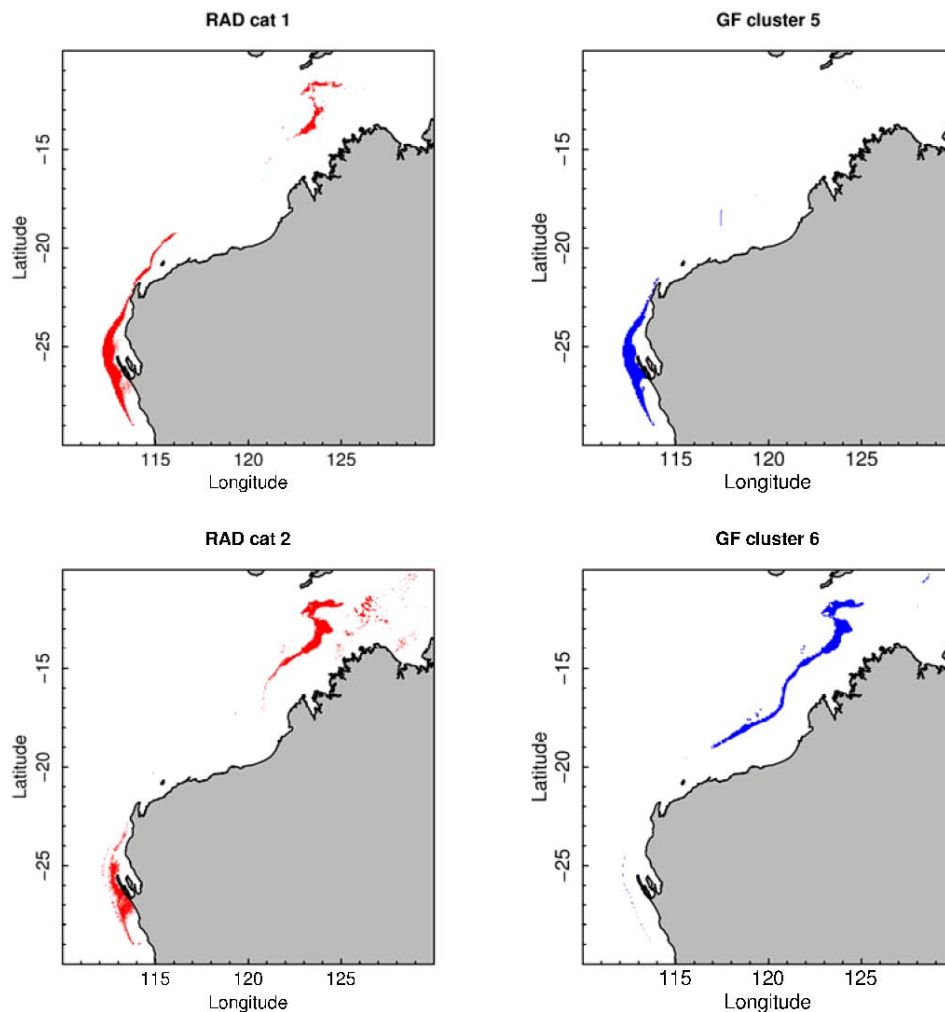


Figure 5: Spatial distribution of RAD categories 1 and 2 and Gradient Forest Clusters 5 and 6. The RAD categories are from low richness low evenness assemblages and the gradient forest clusters represent different species composition mixes.

MANAGEMENT IMPLICATIONS

These two analyses provide complementary information on biodiversity. The Gradient Forest method gives differences in identity and RADs give differences in structure. Consideration of both analyses should allow more comprehensive and representative management decisions to be made. The key question remaining to be addressed is how these methods and others should be integrated to provide an integrated picture of biodiversity that can be used to manage Australia's natural assets.

Biodiversity is a complex concept. Different management questions will need to focus on different attributes of biodiversity and the processes that produced them. There is an opportunity, now, though for new national datasets and analytical methods to improve our descriptions of biodiversity and what is needed to conserve and manage them. This will require philosophical and technical development, and is one of the research areas proposed for the new NERP Marine Biodiversity Hub.

References

- Bray, J.R. & Curtis, J.T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27, 325–349.
- Breiman, L. 2001. Random Forests. *Machine Learning*, 45:5–32
- Dunstan, P. D. and Foster, D. D. (in press) RAD Biodiversity: Prediction of Rank Abundance Distributions from Deep Water Benthic Assemblages. *Ecography*, doi: 10.1111/j.1600-0587.2010.06552.x
- Ellis, N., Smith, S.J., Pitcher, C.R. (2010) Gradient forests: calculating importance gradients on physical predictor.
- Foster, S. D. and Dunstan, P. K (2010) The Analysis of Biodiversity Using Rank Abundance Distributions. *Biometrics* 66, 186–195. doi: 10.1111/j.1541-0420.2009.01263.x
- Guilhaumon, F., Gimenez, O., Gaston, K. J., Mouillot, D., (2008) Taxonomic and regional uncertainty in species-area relationships and the identification of richness hotspots. *Proceedings of the National Academy of Sciences* 105 (40), 15458–15463.
- Kruskal, J.B. (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29, 115–129.
- Legendre, P. & Anderson, M. J. (1999). Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs* 69, 1–24.
- Liaw, A. and M. Wiener. 2002. Classification and regression by randomforest. *R News*, 2:18–22.
- Parviainen, M., Marmion, M., Luoto, M., Thuiller, W., Heikkinen, R. K., (2009) Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biological Conservation* 142 (11), 25012509.

Pearman, P. B., Weber, D., (2007) Common species determine richness patterns in biodiversity indicator taxa. *Biological Conservation* 138 (1-2), 109-119.

Pitcher, C.R., Lawton, P., Ellis, N., Smith, S.J., Incze, L.S., Wei, C-L., Greenlaw, M.E., Wolff, N.H., Sameoto, J., Snelgrove, PVR (2010) The role of physical environmental variables in shaping patterns of biodiversity composition in seabed assemblages.

R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL

Strobl, C., A. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. Conditional variable importance for random forests. *BMC Bioinformatics*, 9:307-317.

Williams, A., Althaus, F., Dunstan, P. K., Poore, G. C., Bax, N. J., Kloser, R. J., McEnnulty,

F. R., (2010) Scales of habitat heterogeneity and megabenthos biodiversity on an extensive Australian continental margin (100-1100 m depths). *Marine Ecology* 31 (1), 222–236.

Appendix 1

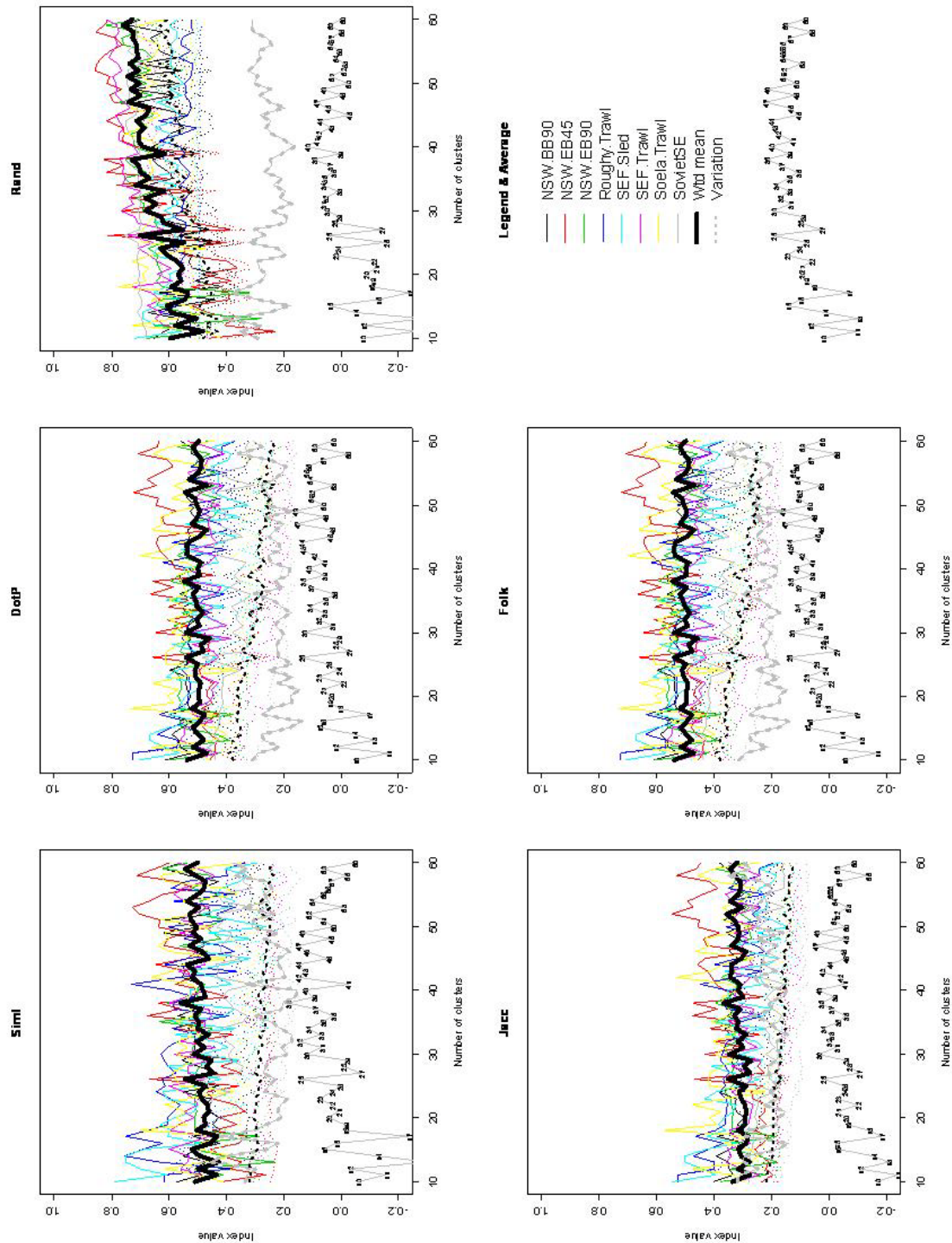


Figure A-1. Measures of similarity of each clustering of the SEMR with the cross-validated terminal node assignment of each constituent survey assessed using the **R** package **clv** (similarity index, dot product, Rand, Jaccard, Folk-Mallows).



Prediction Program

HARVEST PROGRAM 2:

Further development of market based instruments to support structural re-adjustment options

Methods and models developed by the Marine Biodiversity Hub to examine the potential of market-based incentives in the Eastern Tuna and Billfish fishery have been found to have potential in assisting the East Planning Team of SEWPaC to predict the likely consequences of marine reserve design on fishing activities and to predict the distribution of loggerhead turtles at sea from stranding data. Extension and validation of these methods was requested by Phil Boxall, leader of the East Planning team, to support their marine bioregional planning.

TASK: Fishers adaptation strategies

Application of existing model to scenarios provided by the East planning team - investigation of fishers' adaptation strategies to MPA designations in the East Planning Region and their implications for economic impacts on the fishing industry

We developed the project in consultation with SEWPaC's East Planning team as an exploration of potential industry costs from the MPA system that will be developed for the region. The method of estimating cost that the East Planning team is currently using in their reserve design analysis assumes that the foregone catches in the closed areas are representative of cost (i.e. historical catches are representative of future losses).

However, fishers often have a variety of adaptation options including switching locations, trip duration, and targeting strategies. In preliminary analysis we found that adaptation strategies can lead

to minimal realised costs, despite the closure of historically high value regions of a fishery.

We developed two models of fishery behaviour in the Off-reserve management program of the Marine Biodiversity CERF hub that could be used to examine how fishers respond to MPA declaration. While these models were originally developed to look at the effect of incentives on fisher behaviour (location choice and bycatch rates), they can be used for making predictions of responses to MPAs and resulting changes in catch, revenue, and cost.

Project objectives

- Evaluate the potential economic effects of MPA declaration on the Eastern Tuna and Billfish Fishery, the primary fishery operating in the East Planning Region
- Investigate the potential distributional effects of MPA declaration, addressing the relative

- magnitude of shifts in revenue and cost across the communities and fishing ports
- Describe how the distribution of fishing effort might change across the region with MPA declaration

The East Planning team provided 10 MPA scenarios they had developed using the MARXAN software; 5 designed with the goal of minimising cost (while meeting biodiversity objectives) and 5 designed without cost, based on biodiversity alone. We evaluated the response of the tropical tuna fleet, incorporating effects of vessel size and home port, to the 10 MPA scenarios. The evaluation was based on data from the fishery for two years, 2004 and 2007, which represent significantly different years in the fishery due to the availability of fish and the behaviour of the fishers. The focus of the analysis was on whether MPA scenarios result in similar costs when adaptation is considered to the costs estimated based on historic catches alone.

Key findings

In many cases, MARXAN scenarios that considered cost were predicted to have lower costs, even when fishers were able to adapt to the change in availability of fishing grounds. However, one of the reserve designs that considered only conservation values actually resulted in lower costs than all 6 comparable reserves that were designed to minimise cost. In all years, some of the designs that considered both cost and conservation assets were more expensive than those that consider assets alone. In one year, more than half of the designs that considered cost were more expensive than designs that considered conservation assets alone.

There also was significant variation among strategies that MARXAN predicted to be equivalent in their cost, with the range of profit reductions among the 6 strategies being 24% to 78% of the minimum reduction depending on which model was used and which year in the fishery was considered. There was also significant variation in the cost of a particular reserve design in differing years, resulting in the relative ranking of reserve designs by cost changing between years.

These variations are driven by the reaction of fishers to a given MPA design. The historic catches used as a proxy for fishery losses in MARXAN do not consider the responses of fishers, as the former are a static measure of cost. In addition to direct economic effects on the fishery, closure of areas also changes the spatial distribution of fishing effort and the targeting strategies employed by the fishers. We found that fishing effort shifted significantly depending on the spatial configuration of the reserve scenario, and in some cases vessels were also predicted to change target species. This can have significant implications for incidental capture of threatened or protected species such as seabirds and turtles as the catch rate of these species varies depending upon where fishers operate and what species they are targeting (which affects bait, time of day, and gear configuration).



MANAGEMENT IMPLICATIONS

The analysis of economic impacts suggests that the historic catches in a region may not be a good approximation for the actual costs imposed on fishers by a reserve system. Overall it is difficult to make an accurate prediction, as fisher's reactions, and thus costs, depend not only on the configuration of the reserve but also on other factors such as fish prices, fuel prices, and the distribution of fish in the fishery in a given year.

There are two major conclusions relevant for the NSRMPA process that can be drawn from this project.

First, it may be a useful project to evaluate reserve designs using economic models that are more realistic than a simple average value of past catch in a location. While this is likely to be very difficult during the early design process using MARXAN, due to computational limitations of finding optimal reserve designs while incorporating dynamic behaviour of fishers, it can be done post hoc. This would imply using MARXAN to identify potential reserves designs that

MARXAN considers to be equivalent, but subsequently evaluating them using a more complex model as we have done here to further screen for reserves that are likely to impose the least cost on industry for a given benefit for conservation.

Second, given the variability over time, between ports, between models, and among vessels in the costs imposed by a given MPA design, it will be difficult to accurately predict compensation a priori. If fishers are to be compensated for their losses due to MPA establishment, it will be important to use empirical data on costs and revenues prior to reserve establishment and after establishment to estimate losses. However, this approach has risks as the payment system can be manipulated if fishers operate in ways that generate less income than is possible in order to maximise payments. One approach to address this issue is to provide loans immediately upon MPA establishment to assist fishers in accommodating additional costs imposed, with repayment of the loans being discounted depending upon actual losses estimated some time after the MPAs are in place. If the audit to establish losses occurs at an unpredictable time somewhere within a relatively long window, eg using costs and revenues from a 6 month period some time in the 5 years following MPA establishment, it will be much more difficult for claimants to manipulate the payment system.

Outputs

A draft manuscript has been provided to the East Planning team with the full analysis from the project. This manuscript will be submitted to a scientific journal and be available through that venue.

TASK: Threatened species distribution

Refinement and validation of existing model to provide predictions of at-sea distributions of loggerhead turtles - using novel data sources to predict the distributions of threatened marine vertebrates

Outside the immediate coastal zone there is very little information on the distribution of marine vertebrates. Fisheries observers provide some information for species that are frequently encountered, such as seabirds. In some cases electronic tagging can be useful in providing data on areas of high usage by individuals, which is often used as a proxy for areas of high population density. However, the cost of electronic tags generally means that this technology is generally limited to studies with relatively few individuals. However, for many species of concern such as marine turtles neither of these approaches is satisfactory. Fisheries observers rarely see turtles, and the critical life stages are difficult to access to tag. Due to these two characteristics, the at-sea distribution of many marine turtles remains largely unknown.

As part of a project on Incentive based management in the Off-reserve management program of the Marine Biodiversity CERF Hub, we used a model of oceanographic drift to analyse the expected at-sea distribution of loggerhead turtles based on records of strandings along the east coast of Australia. Based on 1,200 strandings, this model provides the first synoptic picture of

the distribution of loggerhead turtles in the Coral and Tasman seas.

However, we developed the model to give a rough picture of loggerhead distributions that we could use in another project. As such, we did not attempt to statistically verify the model against other fisheries observer data, nor did we correct for the distribution of mortality sources at sea which will affect the estimates of relative density of turtles in different regions. The East Planning team had expressed interest in being able to use these data in their planning process, which would require the two shortcomings above to be addressed. Furthermore, improving the current knowledge of the at-sea distribution of the turtles will be widely useful within the Department for strategic assessments, assessing threats from other sources, providing supporting information for the marine turtle recovery plan, and interpreting variations in the patterns of strandings observed over time.

Project objectives

- Develop a model for predicting the at-sea distribution of marine turtles using an oceanographic model to predict the drifting paths of turtles found stranded on beaches
- Refine this analysis by correcting the stranding data for observation effort to obtain a standardised density of stranded turtles
- Incorporate potential sources of mortality at sea, in particular the distribution of longline and trawl fishing operations, to correct the predicted distribution for variation due to differential mortality across the region.

Approach

We obtained stranding data from the Queensland Department of Environment and Resource Management's Marine Wildlife Stranding Database on the stranding locations of loggerhead turtles. There were approximately 1,200 records spread from northern NSW to the NT border. For each stranding record we queried a CSIRO database to obtain a large sample of locations at sea that would lead to a drifting object washing ashore in the

distributions for each turtle across all of the turtles in the stranding database.

We corrected these stranding records for the probability a turtle in a given location would be observed by weighting the strandings by the inverse of the local human population density around the stranding site. To correct for differences in the likelihood of a turtle in a certain location dying and eventually washing up on a beach to generate a stranding record, we will normalise the at-sea distribution of the turtles by the inverse of the fishing effort in that region. The two types of fishing effort will be weighted differentially according to the observed number of turtles captured per unit of effort.

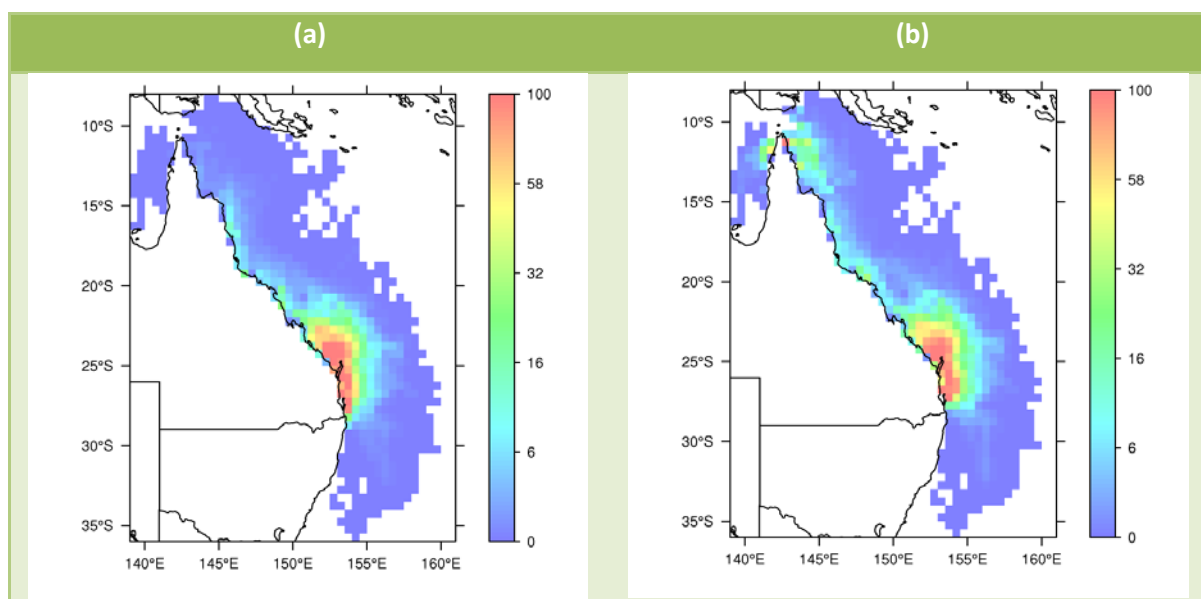


Figure 1. Potential source areas of stranded turtles estimated by ocean drift analysis (a) and weighted by the inverse of human population density (b)

stranding location. For each stranding we combined these locations, normalising them to generate a probability distribution for the source of the stranded turtle. We generated an overall at-sea distribution for loggerhead turtles stranding along the coastline by summing up the probability

Key findings

- We have implemented and tested the GIS and database procedures required to perform the drift analysis and correction for human population density. Next steps are to tune the weighting function used for the correction and to add the final correction for fishing effort.

- The primary effect of reweighting the distribution by population density is to increase the estimated distribution of turtles in the northern portions of the Queensland coast. This matches the general expectation, as there are turtle nesting and feeding grounds in this region. In addition, even corrected for observation probability the large concentration off the southern Queensland coast remains. This matches the general pattern that would be expected, as the largest nesting ground for loggerhead turtles is at Mon Repo on the southern Queensland coast.
- Due to the complexities involved in the analysis we were not able to incorporate the effect of mortality sources on the distribution. The analysis is technically feasible, it was just not possible to complete within the project.

MANAGEMENT IMPLICATIONS

The map provides the first at-sea distribution for loggerhead turtles in Australia. While there certainly are records from fishing vessels and a small amount of satellite tracking data, to date there has been no quantitative estimate of density at sea for any of the Australian species at a large scale.

The methodology developed in this project can be applied to other types of animal stranding data. A number of states collect strandings data, and there are efforts to develop a national database for strandings which would include all marine species, and a specialised one for marine mammals. The approach we demonstrate here could make that data much more useful for understanding the biology of threatened marine species in Australian waters.

TASK: Further develop market based instruments to support structural re-adjustment options

Expertise developed by the Marine Biodiversity Hub in developing alternative incentives for conservation management has previously been requested by SEWPaC to support development of options for structural adjustment in fisheries following the declaration of marine reserves.

Objective

The Marine Biodiversity Hub hosted a workshop on 4 November 2010 with the Department of Sustainability, Environment, Water, Population and Communities in Canberra to bring together policy-makers, managers and researchers in marine management and incentive-based measures. The goal of the workshop was to foster understanding about the future policy and program challenges in managing the marine environment, and to identify potential opportunities for market-based instruments to contribute to improved and cost-effective conservation outcomes.

Workshop attendees

Paul Garrett, Nathan Hannah, Nigel Routh, Steve Jackson, Nicole Middleton, Chris Murphy, Dave Johnson, Gareth Evans, Stefan Caddy-Retallic, Claudia Cooney (SEWPaC), Nic Bax, Chris Wilcox, James Innes, Michaela Guest (CERF).
Invited experts: Tony Smith (CSIRO), Keith Sainsbury (AFMA, Marine Stewardship Council), Bill Langford (RMIT), Sarah Jennings (UTAS).

Workshop sessions

Session 1: Marine Biodiversity Hub Overview. Nic Bax

- Research completed by the CERF Marine Biodiversity Hub to date includes the prediction of biodiversity. Predictions of biodiversity are probabilistic and include uncertainty of prediction.
- These predictions were used in marine bioregional planning.
- Probabilistic predictions provide the potential for a greater variety of management options to achieve declared goals, including integrated on- and off-reserve management.
- Market-based instruments have been examined as one of the options for managing marine biodiversity and advice was provided at SEWPaC request on their potential to predict effort dislocation and provide alternatives to structural adjustment for marine reserves.
- There are future opportunities to explore the application of MBIs in marine conservation through an integrated approach to marine biodiversity management, and improved management of listed marine species. These are two programs within the NERP bid.

Session 2: What is an incentive based measure? Chris Wilcox

- Globally, current approaches to managing biodiversity have not been successful in managing biodiversity decline.
- Market-based instruments offer an alternative to direct regulation of environmental impacts.
- MBIs incorporate the damage into the cost of production. i.e. the proponent pays the social/environmental cost of the activity.

- MBIs create an incentive for business to innovate and reduce impacts, can improve efficiency and cost effectiveness, and shift the burden of proof to private parties, focus on outcomes consistent with the EPBC Act.
- Different MBIs have different information requirements.

Issues/Questions

- The application of MBIs requires meaningful, quantifiable targets for the management of biodiversity.
- The application of MBIs need to consider the complete costs born by the regulator and the administrator including those additional costs that may be incurred during the transitional period.

Session 3: SEWPaC context, objectives and challenges for the application of MBIs. Paul Garrett

- High-level drivers: EPBC Act, MNES, SEWPaC Strategic Plan
- Relevant programs: Environmental assessment and approvals, MBP, Species conservation, Strategic assessment of fisheries, MPA management;
- MBIs considered part of the solution to halt the decline of biodiversity ;
- Challenge with how to apply MBIs and the cost of administering them.
- Challenge with gaining acceptability for the use of MBIs with SEWPaC stakeholders- concern of legitimacy and effectiveness.

Issues/Questions

- Challenge with understanding the applicability of MBIs for small, data poor fisheries such as state managed fisheries.
- SEWPaC need to establish clear objectives for successful use of MBIs.

Session 4: Managing marine fisheries: opportunities for the use of MBIs to achieve environmental and fisheries outcomes. Prof Keith Sainsbury

- MBIs can be powerful tools that must be used judiciously. MBIs are not the silver bullet but one of a suite of tools needed to achieve desired outcomes.
- Challenges associated with the application of MBIs include – technical difficulties in predicting/measuring effect of MBI; interactions between multiple scales and sectors subject to different MBIs; potential costs of transition, transaction and compliance of MBIs; changed distribution of costs and benefits; potential for politicisation and monopolies.
- Most methods need a combination of legislative, policy, and regulation
- Need to improve coherence of MBIs and align them with ESD goals of relevant sectors and develop autonomous rationalisation pathways across sectors, eg cross sectoral tradability of access/impact allocations.
- There are a wide range of MBIs and options for financing MBIs that have been used successfully and that may be transferred to other sectors/scenarios.

Issues/Questions

- MBIs may have role as temporary tool for changing a fishery's environmental impact.
- Need to bring together scientists, managers, economists and business.

Session 5: If offsets don't deliver no net loss, what parts of the policy are failing? Bill Langford

- It is often difficult to evaluate the effectiveness of policy/management decisions due to the sets of actions with sequential dependence, uncertainty in all steps, and the idiosyncrasies of local context.

- In the case of offsets policy, there is a sequential process (screen, assess, choose, restore, manage, protect) by which the policy is developed that can be modelled using uncertainty estimates to identify the components of the process likely to contribute to policy failure if not implemented properly.
- There are clear policy implications including resource allocation. In the example provided, protection was what really mattered, screening and assessment not worth refining, and restoration more important than screening and assessment.
- The applicability of offsets needs to be examined for each specific situation.
- Modelling likely policy outcomes can be helpful in managing and understanding implications of uncertainty in a management context.

Issues/Questions

- SEWPaC expressed interest in being able to model uncertainty associated with policy decisions in addition to the uncertainty associated with species and ecological interactions from a fisheries perspective. The Hub indicated that there had been much work in this area of management strategy evaluation and there is capacity in the Hub under the integrated management project of the NERP proposal to further develop this.

Session 6: Worked examples and discussion

Seabird-long-line example

- Chris Wilcox outlined the example of using MBIs to manage by-catch mortality of protected seabirds in long-line fisheries and noted the reluctance of SEWPaC to adopt an approach that essentially placed a quota on seabird mortality within the fishery.
- SEWPaC indicated the legal constraints provided by the relevant sections of the EPBC Act to approve injuring or killing of a protected species.
- SEWPaC indicate that language around the use of MBIs to achieve conservation outcomes must be consistent with that used in the EPBC Act if MBIs are to get Departmental and stakeholder support for their use.
- SEWPaC indicate that in the context of the seabird mortality associated with long-lining, this may be viewed as an interim management measure that sets a limit on incidental take of seabirds with a goal of zero incidental take of seabirds over time.
- SEWPaC indicate that there is a role for the Hub in helping to demonstrate/provide evidence of the benefits of MBIs to achieve conservation outcomes but note the importance of ensuring stakeholder support including fisheries and eNGOs.
- SEWPaC also note the challenge of developing good management triggers without good population data – CERF Hub note that this is part of integrated management project under the NERP and there are a range of approaches for dealing with uncertainty useful to the application of MBIs to achieve conservation outcomes.
- SEWPaC expressed interest in further discussion with the Hub to understand what aspects of recovery planning are failing and to identify specific opportunities where the use of MBIs may help to reach better conservation outcomes for protected species. Species mentioned were seabirds, seals, turtles.

Marine turtles and oil & gas development example

- Question characterised by impacts of unknown size on poorly estimated population
- Deal with uncertainty by extending time period over which management applies and using performance bonds as alternative to offsets.

Issues/Questions

- Can be difficult to prove that company holding is bond is responsible for general population decline over extended period.
- Performance bonds and offsets typically come in when sunk costs are already high and management options have been reduced.
- Need prior estimates of environmental costs of alternative development options at an early stage in planning when alternatives are possible (eg. Gorgon development could have been on mainland).
- Need objectives and targets.

Use of MBIs in marine protected areas

- SEWPaC outlined current work program to move from management of a small number of iconic reserves to a large number of reserves that constitute a network. Identify challenge of managing the network – cannot simply scale up management approaches. Budgetary constraints. Note: The EPBC Act excludes all activities in a reserve and the management plan provides the approval for activities and the conditions under which they can occur.
- SEWPaC note for many reserves we know little about the environment that we are protecting. How do you know which MBI to choose, and how do you decide if it is working?
- SEWPaC interested to understand how you can structure MBIs to minimise

impacts of activities in reserves that is not inconsistent with the IUCN zoning of reserves and legal requirements of management plans under the EPBC Act.

- SEWPaC note the closer you are to an asset/conservation value, the higher the level of protection required. Can MBIs be stratified to reflect the probability of interaction with an asset such as a seal colony? How would these compare to alternative zoning?

Issues/Questions

- SEWPaC and Hub to meet to further discuss potential scenarios in reserves to which MBIs may be applied and understand the legal context that may facilitate or constrain the use of MBIs in MPAs.
- SEWPaC - Can MBIs reduce cost burden of managing MPAs?
- Retaining some fishing in MPAs provides monitoring capacity.
- On- and off-reserve a continuum and advantages to considering together.

General comments/questions

- SEWPaC indicate there is a need for the Department to determine what we want the Department to look like often longer time horizons (eg 30 years). It would then be possible to work back from that time horizon and consider what activities would facilitate/hinder reaching that goal. SEWPaC can then consider how you use MBIs to reach the long-term vision for the environment.
- The Hub indicates interest to further discuss development of a decision-making tool that may help envisage different future scenarios and potential management options.
- SEWPaC and the Hub express interest in understanding the role of MBIs in combination with co-management arrangements – incentivising proponents to manage impacts themselves and understanding different co-management structures. Sarah Jennings (UTAS) has done work on these issues with understanding fisher behaviour (“experimental economics”).
- Need to take suite of concrete MBI options (eg. for MPA management) to legal folk to test under Section 15 of the Act.

- SEWPaC identify budgetary constraints on implementing TAPs and managing the Commonwealth marine reserve estate - keen to know how MBIs can help. Welcome the identification of 3 TAP species on which the Hub can further develop examples of the costs/benefits of MBIs to achieving cost-effective conservation outcomes.

Next Steps

- The CERF Hub to work with SEWPaC staff to further define issues where this workshop agreed the application of MBIs may provide improved conservation outcomes. These areas relate to:
 - protected species conservation (Species Conservation Section and Sustainable Fisheries Section, MD)
 - fisheries and protected species interactions (Sustainable Fisheries Section)
 - future scenario development and decision-making (Environment and Assessment Branch, AWD), and
 - management of marine protected areas through on and off-reserve approaches (MPA Futures Section).

Agenda, notes and presentations from this workshop are available on the CERF Marine Biodiversity Hub website: <http://www.marinehub.org/workshops-2010>

EXTENSION PROGRAM 1:

Extending national data holdings in preparation for future marine biodiversity research to support implementation of marine bioregional plans

This project took advantage of research and data developed during the Marine Biodiversity Hub by hub partners and collaborators to prepare for future marine biodiversity mapping around Australia to support the implementation of marine bioregional plans, provide new national data sets to support the work of Heritage, and to prepare for a possible update of IMCRA 4.0. It developed new regional and national datasets for polychaetes, ophiuroids and sponges, extended analyses of samples from Carnarvon shelf to four additional phyla, developed new national data layers for physical surrogates especially satellite-derived which have special relevance to a pelagic habitat mapping. In addition, this program identified gaps in the coverage of marine biodiversity information (and physical surrogates) in Australia to guide future research. This program supported current CERF researchers, and engaged additional experts from the partners, and taxonomic experts from museums who were not part of the Hub.

TASK: Harvesting of polychaete diversity data

Develop North Australia polychaete dataset from Hub and other data sources and based on a consistent taxonomy. (Eunicidae, Nereididae, Phyllodocidae, Polynoidae, perhaps Spionidae)

CERF Marine Hub sampling and analysis from Voyages of Discovery on the Western Australian continental margin addressed local aims of that program. This extension task aimed to harvest new data from relevant collections and establish new data sets representing wider geographic and bathymetric ranges from which to predict patterns of marine biodiversity distribution for Australia. These data will improve our understanding of the veracity of existing data used in

marine bioregional planning to predict patterns of marine biodiversity.

Project objectives

- Identification of selected polychaetes collected by Geosciences Australia projects (Surrogacy Program) at Carnarvon, WA and Lord Howe Island and Jervis Bay, NSW.
- Fill gaps in national data sets by identification of existing (but previously unidentified) target polychaete taxa from northern and north-eastern Australia.

Outcomes

- Ability to extrapolate from Surrogacy Program projects to the Prediction Program by correlating identifications using a shared set of taxa.
- Wider context for addressing questions of endemism from discoveries on Western Australia Voyage of Discovery cruises.
- Expansion of national biodiversity datasets available for testing bioregional hypotheses.

Approach

Target taxa included: Eunicidae, Nereididae, Phyllodocidae, Polynoidae and Spionidae. These are polychaete taxa for which expertise was available and a stable taxonomy for Australian species was available. Existing data from Australian natural history collections have been previously assembled into a single relational database; queries to that database identified gaps in distribution data for the above taxa for northern and north-eastern Australia. Material made available for study included: collections made by Geosciences Australia projects at Carnarvon, WA and Lord Howe Island and Jervis Bay, NSW; shallow-water collections from Australian Museum, Sydney, Queensland Museum, Brisbane, Museum and Art Gallery of the Northern Territory, Darwin; and CReefs collections from Lizard and Heron Islands, Qld and Ningaloo,

WA. Material from these collections and taxa was obtained and identified using standard taxonomic tools (Wilson, et al., 2003) with new taxa added as they were encountered. The bulk of the material represents a bathymetric range of <100 m. Data were entered into a local relational database for quality control and import to relevant institutional databases. Data delivery will ultimately occur through the **Atlas of Living Australia (ALA) Spatial Portal** (directly at <http://www.ala.org.au/explore/species-maps/> or via the appropriate metadata link at <http://www.marinehub.org/documents>) which accesses federated datasets from all relevant Australian natural history collections. Since ALA data depend on workflows at State museums, in the interim data will be available through **Museum Victoria (MV) Collections Online** - at the following link, enter Collection Name = "CERF" and enter taxon name for specific searches: <http://collections.museumvictoria.com.au/search.php?type=NS>) as collections data (Museum Victoria specimens including all Geoscience Australia material) or "virtual records" (specimens belonging to other museums). Data and metadata will also be accessible through the **CERF Marine Hub website**, eg <http://www.marinehub.org/macrobenthic-biodiversity-data-western-australian-continental-margin> but primary data source for these data will be museum databases accessible via ALA. Data for non-target polychaete taxa will be entered and available at Family-rank identifications for gap analyses and to facilitate future data harvesting opportunities.

Key findings

- Taxonomic results *per se* are not an objective of this project, nevertheless access to a standard taxonomy is part of the metadata for these results and is also vital for any subsequent further development of the biological data. For the target taxa, >100 species not present in the current identification tool (Wilson, et al., 2003) were discovered during the CERF Marine Hub work and the present data-harvesting extension task. Options are now being explored for web-based

distribution of a new version of the identification tool (to be supported outside CERF and CERF-extension funds).

- 'Biodiversity results' (ie new distributional data) are the principal output of this project. Interim results are now available online through Museum Victoria (MV) Collections Online at the above link, both for polychaete and for other taxa. Those links are dynamic queries which will generate increasingly large data sets as data entry continues (in particular, northern and north-eastern Australian gaps will increasingly be filled as "virtual records" (specimens belonging to other museums) are imported from March-June 2011.
- The above dynamic links retrieve both CERF-related and previously available data from the MV collections. Approximately 1200 records have been added through CERF extension funding with some data still to import as of 11 Feb 2011; work continues until June 2011 so at least several hundred additional records and an increasingly useful national coverage will result.

Reference

Wilson, R.S., Hutchings, P.A. & Glasby, C.J. (2003) *Polychaetes - An interactive identification guide [CD-ROM]*. CSIRO Publishing, Melbourne, pp.

MANAGEMENT AND RESEARCH OPPORTUNITIES

Polychaete worms are typically the most abundant taxon reported from quantitative ecological studies on soft sediment communities in inshore waters (where human impacts are greatest). However on a national scale, data representing this dominant taxon have not been available to assist marine planning. The CERF Marine Hub and extension funding has aimed to make polychaete data available and to test the ability of existing data, such as those used to generate IMCRA, to predict patterns in the fauna as a whole. The utility of these data and their value to SEWPaC and State government agencies will now be testable on a national scale as a result of the CERF and CERF extension funding.

TASK: Paleogenetic analyses

Report on utility of approach to develop hotspots of speciation and extinction by extending paleogenetics work on cryptic species identified by the Marine Biodiversity Hub.

Diversification rates of biota (i.e. extinction – speciation dynamics in space and time) are influenced by environmental conditions and the geographical complexity of the ecosystems. At the species level, the evolution of morphological characters, aligned with patterns of molecular phylogeography, represent a robust instrument to understand the diversity encountered within and among populations, approximate the timing of diversification, and associate the levels and patterns of diversity with environmental conditions and geographical profiles. This information is useful to predict the potential resilience of marine biodiversity to future change and improve decision-making in the protection and ecological sustainable use of marine systems.

Project objectives

- To understand how species evolutionary history is shaped over time, over large geographical scales as well as locally, and;

- To identify hot spots of: a) elevated diversity, useful sources of species-specific novel genetic variants and b) fragile regions exhibiting low rates of biological diversity, eventually experiencing extinction events.

Approach

The tropical to subtropical squat lobster *Agononida incerta* Henderson, 1888 (Galatheididae) is a widely distributed taxon in the Indo-West Pacific found from East Africa to Japan and Australia. Geographically disjunct populations of this species may represent biologically and genetically distinct cryptic species. Given a time-calibrated phylogeny, this species represents a good model to detect significant fluctuations in diversification among genetically distinct lineages and to locate major shifts of speciation/extinction in the light of environmental events and when they occurred.

Key findings

Preliminary mitochondrial phylogenies inferred from selected specimens of the species collected from New Caledonia, Taiwan and several sites from North Western Australia indicated that *Agononida incerta* represents a cryptic species complex and consists of up to six genetically distinct lineages some of which can be recognised at the morphological level. So far, additional material (total of 65 specimens) has been obtained from overseas collections and Zoological Museums, described at the morphological level and transported to AIMS for genetic analysis.

In progress

Morphological delineation of the new species and molecular phylogenies are in progress on the global dataset. This is to identify robust taxonomical units and/or geographically associated populations at a larger scale and infer their genealogical history. Statistical analyses will estimate the intra-clade variability, the age of single clades and the changes in the distribution of nodes and internal branch lengths within the topology. This will provide the elements to evaluate fluctuations in extinction – speciation processes within species and populations.

INTERPRETATIONS AND MANAGEMENT

Key estimates will focus on shifts in diversification rates as the result of variations in population size and genetic variants. Increased rates are interpreted as speciation processes related to biodiversity hot spots. Decreased rates are associated with extinction risks due to low levels of genetic variability within species, typically encountered in small, unconnected and fragile populations susceptible to biological invasions and fragmentation. The latter is particularly relevant in cases of known areas of ancestral radiation experiencing ongoing extinction events. This can be used to predict theoretically fragile, rather than resilient, ecosystems potentially requiring additional management measures. This will support SEWPaC's commitment to understanding the status and value of species or ecosystems, their vulnerability to climate change, and potentially the feasibility of adaptive responses under the Marine National Climate Change Adaptation Research Plan.

TASK: Comparative investigation of phylo-geographic patterns of connectivity

Extend genetic analyses of regional ophiuroid populations analyses to tropical areas, including the Coral Sea to delineate populations and identify levels of connectivity

The CERF hub investigated patterns of connectivity between populations of deep-water seabed animals across the Southern Australian region. The aim was to see if larval dispersal was directional, ie flowed from regions that acted as 'sources' of larvae to other regions that largely received them as a 'sink'. Understanding the direction and scale of connectivity between populations is important for marine park planning, climate change adaptation and restoration of habitat following human impacts (eg oil spills, mining). The previous CERF projects used mitochondrial and microsatellite DNA sequences to calculate larval connectivity between populations. However, obtaining microsatellite data is expensive (and thus obtained for relatively few species) but population genetic research requires a number of independent molecular markers to resolve connectivity patterns.

Project objectives

- Investigate patterns of source-sink connectivity between populations of tropical seabed species from the Australian region. Species of ophiuroids were used due to availability of

appropriately-preserved tissue samples.

- Trial the cost effectiveness of using new genomic technologies to identify informative molecular sequences that can be used for population genetics of offshore animals.

Outcomes

- The identification of patterns of gene flow for deep-water species across the tropical regions of Australia.
- The identification of appropriate molecular markers for deep-sea population genetic studies.

Approach

- Tissues were sourced from two species of ophiuroids from Museum Victoria, the NZ National Institute of Water and Atmospheric Research (NIWA), and the Muséum National d'Histoire Naturelle, Paris (MNHN).
- DNA was successfully extracted from the majority of samples.
- Partial genomes of three specimens of each species were sequenced using Illumina genomic technology. In all, 26 million base pairs of data were obtained. This data was aligned for both species using Velvet software.
- A series of molecular sequences were identified that appeared to be informative for population genetic studies. Primers were developed for four markers for each species.
- The genomic 454 technology is being used to sequence the four new markers for all extracted samples (in progress). The results will be assembled into appropriate formats for population genetics using bioinformatic software.



- The 'barcode of life' gene (COI) was sequenced for all individuals as a comparison.
- The data will be analysed to identify population structure over large spatial scales and directional (source-sink) dispersal of propagules.

Key findings to date

- Partial genomes can be successfully obtained from ethanol-preserved specimens collected up to 10 years ago by national research agencies and deposited into museum collections. The large investment made to collect existing museum specimens can be leveraged to research patterns of population structure for deep-sea animals.
- Novel molecular markers useful for population genetics can be identified from these genomes and sequenced in bulk using modern genomic technology.

MANAGEMENT AND RESEARCH OPPORTUNITIES

Investigating patterns of connectivity for deep-sea animals is problematic due to the expense and difficulty of collecting adequate material. Consequently, the use of previously collected material deposited in museums is an important scientific resource. The emergence of new genomic technologies facilitates the extraction of useful genetic data from these specimens with entire studies completed in relatively few procedures. Establishing source-sink relationships between populations of seabed animals will assist in the understanding how ecosystems will respond to climate change and recover from other human impacts such as oil spills and resource extraction.

Develop a sponge database that is consistent at the OTU (Operational Taxonomic Unit) level for sponge collections in Northwest Australia to support detailed biodiversity mapping in this area

al., 2009). However, sponges are difficult to identify without detailed examination of sponge tissues. Moreover, sponge taxonomy is incomplete and currently in a state of review. In 2008, the CERF Surrogacy program sampled 110 stations for Carnarvon shelf epifauna, identifying this region as a hotspot in sponge biodiversity (754 specimens, est. 261 species/OTU's (Operational Taxonomic Unit), 112 genera). This survey contributed substantially to the museum's sponge collection from northwest Australia. However, an OTU recognised in one dataset may differ from that in another and consequently, there was a



Figure 2: An example of the comprehensive new sponge catalogue for northwest Australia.

need to standardise OTU definitions across different data sources and biogeographic locations.

- The purpose of this pilot project was to develop a sponge database that is consistent at the OTU (Operational Taxonomic Unit) level for sponge collections currently available from northwest Australia.

Approach

Sponge taxonomists from the Museum of Western Australia (Dr Jane Fromont and Mr Oliver Gomez) and the Museum and Art Gallery, Northern Territory (Dr Belinda Alvarez de Glasby) compared sponge OTU's from the Carnarvon Shelf CERF Surrogacy program (754 specimens; 261 OTU's; 112 genera) with existing museum sponge specimens from northwest Australia (531 OTU's, 134 genera) to provide standardised sponge OTU's for the northwest. Sponge OTU's were catalogued into a comprehensive new sponge identification database.

Key findings

This pilot study examined >1100 specimens comparing 206 OTU's (79% of CERF OTU's). Priority was given to numerically dominant sponge orders (eg Astrophorida, Spirophorida, Homosclerophorida, Haplosclerida). This comparison recognised 125 genera, of which 51 (41%) were present in both the CERF and existing museum collections, while 74 genera (59%) were identified as new records to the region (Schönberg & Fromont, *in review*). Of the three numerically dominant sponge orders, 178 specimens were matched (160 Astrophorida, 15 Spirophorida and 3 Homosclerophorida), while 54 matches were recorded from the less abundant Haplosclerida. Astrophorida were reduced from 23 to 18 OTU's and 10 to 9 genera, while Spirophorida were reduced

from 5 to 3 OTU's and 8 to 4 genera. Based on these comparisons, Schönberg & Fromont (*in review*) have now predicted 300 species/OTU's for the larger Ningaloo bioregion. Sponge species/OTU's from the CERF and museum collections are now catalogued by visual appearance, colour, and tissue and spicule morphology within a new comprehensive sponge identification database (eg Figure 2).

References

Schlacher, T.A., Williams, A., Althaus, F., Schlacher-Hoenlinger, M.A., (2009). High-resolution seabed imagery as a tool for biodiversity conservation planning on continental margins. *Marine Ecology*, 31: 200–221

Schönberg, C.L., Fromont, J. (*in review*). Sponge gardens of Ningaloo Reef (Carnarvon Shelf, Western Australia) are biodiversity hotspots. *Hydrobiologia*.

NEW KNOWLEDGE AND OPPORTUNITIES

These data will provide a powerful tool for mapping sponge diversity, and assessing the utility of sponge diversity as a surrogate for seabed biodiversity in northwest Australia. Improved biodiversity maps for the northwest will increase confidence in managing the impacts of planned oil and gas development, implementation of marine bioregional plans, and identification of areas of interest for Heritage listing. Subsequent biodiversity predictions of sponges arising from the revision of OTUs of sponges provided here will also improve on those currently available. These data will contribute to a future nationally-standardised sponge dataset.

TASK: Extension of biological analysis for Carnarvon Shelf

Extend Carnarvon Shelf survey analyses to include four additional animal phyla (ascidians, crustaceans, echinoderms, and molluscs).

The ability to manage and conserve marine biodiversity depends on knowledge about species richness, and about the spatial patterns of distribution and abundance of individual species and/or higher taxonomic groupings. In the Surrogates Project of the CERF Marine Biodiversity Hub, plants and animals were collected from the seabed at >110 stations across the Carnarvon Shelf, offshore from Ningaloo Reef. Initial examination of the sampled invertebrates showed a highly diverse marine fauna that justified more detailed species-level identifications than could be achieved at the time of collection. This decomposition of some higher groups (eg Family, Phylum, etc.) will enable the project team to search for environmental proxies (surrogate variables) that can be used to predict the distribution and abundance of individual species. Spatial modelling using robust surrogates yields more accurate maps of species distribution and abundance than simple interpolation of sampled abundances over sparsely sampled domains. These same surrogates may potentially be used in different places to predict the local biology without any direct sampling of the new area. An indirect benefit of this task is that the taxonomy of the CERF collections has been

standardised with other collections held by different repositories (eg museums, federal collections) and has expanded the information base for the marine bioregion that includes Ningaloo Reef.

Project objective

- The purpose of this CERF Transition Task was to increase the resolution of spatial modelling over the Carnarvon Shelf by additional taxonomy on four abundant phyla (ascidians, crustaceans, echinoderms, and molluscs) that had many species unresolved at the time of collection.

Approach

Species from the four phyla were identified by taxonomic specialists at the Western Australian Museum (WAM) or by Curators at other Museums as directed by Dr Jane Fromont (WAM). Due to a lack of ascidian experts within Australian museums, ascidians were substituted with bryozoans; which was another abundant phylum in the collections. Crustaceans were identified by Andrew Hosie and Lee Betteridge (WAM); echinoderms by Loiset Marsh (WAM); molluscs by Corey Whisson and Jenelle Ritchie (WAM), and bryozoans by Kevin Tilbrook (Museum of Tropical Queensland). All identifications have been integrated with the Western Australian Museum species archive and identified as a data set arising from the CERF Surrogates Program.

Key findings

Crustaceans

A total of 270 specimens, all belonging to the Infraorder Brachyura (true crabs), was identified as belonging to 111 morphospecies from 16 families. Several entirely new species were



Figure 1: *Portunus cf. orbicularis* (Richters, 1880), a new Australian record. Previously known only from the western Indian Ocean.

recorded, while other species were new records for Australia (see below). The nomenclature of the new species is currently being determined with a view to publication in the scientific literature. The remaining specimens, including false crabs (eg hermit crabs), prawns, shrimps, lobster and non-decapod crustaceans (eg isopods) are yet to be examined by specialists.

Molluscs

A total of 291 specimens was identified as belonging to 118 mollusc species. The collections contained 49 bivalve species representing 19 families, 48 gastropod species representing 26 families, three cephalopod species representing two families; and two scaphopod species. Two new predatory snail species were recorded from the family Muricidae.

Echinoderms

A total of 50 previously unidentified echinoderms was examined including holothurians (sea cucumbers), Gorgonocephalidae (basket stars), and crinoids (feather stars). In addition, all asteroids (sea stars) were cross-checked and validated. From this, 12 holothurian species were identified from five families along with two basket stars. Twenty-two asteroid species were validated. Crinoid identifications are yet to be finalised.

Bryozoans

A total of 63 specimens was identified as belonging to 30 species from 18 families.

NEW KNOWLEDGE AND OPPORTUNITIES

These small collections included several new taxonomic discoveries for the Carnarvon shelf and the broader North-West region, which suggests that knowledge of the diverse invertebrate fauna of this region remains incomplete. The new data adds to knowledge of the biodiversity from this marine domain and will be used in future research to test the utility of biophysical surrogates for predicting spatial patterns in marine biodiversity.

TASK: Gap analysis

Report on gaps in datasets provided to SEWPaC and marine scientists in Australia - identify gaps in regional and national physical and biological datasets that are used to support implementation of marine bioregional plans

A. Gap identification for marine biodiversity

Marine bioregional management requires maps of biodiversity composition patterns over large regional scales. Advanced methods for predicting biodiversity from environmental surrogates have been developed by the Marine Hub to achieve large regional scale coverage. However, the results can only be as good as their foundation in quality survey data, for both biological and physical data; where quality is both information content of samples and sample density. Based on outputs already delivered by the Hub, this task assessed gaps in biodiversity knowledge, by area, environment and major biological groups. A complementary task has assessed spatial gaps for national bathymetric and sediment data layers. This information will make a valuable contribution to identifying future research required to characterise biodiversity and inform the management of the Commonwealth marine environment particularly as it relates to marine reserves.

Project objective

- To identify and prioritise biological data gaps by area, environment and major biological groups.

Approach

The gap assessment was based on biological survey data collated by the Marine Hub, and outputs from Gradient Forests analyses of biodiversity composition for each marine region, completed by the Hub to deliver seabed assemblages maps to inform SEWPaC's program to establish the national representative system of marine protected areas (provided as [Product Descriptions](#) on the hub's

website). Those analyses used biological information to transform the available environmental layers from physical scales to a common biological scale for each layer. The transformed multi-dimensional space represented a biological space where greater distances between points in the space are expected to have greater differences in biodiversity composition. For each region, models were developed for the relationships between compositional differences among survey sites and distances in both Gradient Forest biological space and geographic space. Mapping the modelled compositional differences provides information on biological data gaps that is much more than a simple map of sample density, but uses surrogate relationships as a function indicating how quickly biodiversity changes along environmental gradients, and also takes spatial relationships into account. Maps of sampled site locations were also produced.

Key Findings

Maps of sampled sites on following pages (Figures 1 and 2, left) show obvious gaps in surveyed locations. For tropical coral reefs, only a few selected reefs in the Great Barrier Reef, Torres Strait and North-West Shelf have been sampled, with large gaps in these regions and little or no data for the North Marine Region and Coral Sea. Similarly for temperate reefs, there are gaps in sampled areas and large tracts of the southern coastline are unknown. The compositional differences for these data types could not be modelled as there are not complete data layers indicating the location of either tropical or temperate reefs across their respective regions.

For the continental shelf and upper slope, fish trawl surveys provide the greatest spatial coverage (Figure 2) but narrowest biodiversity coverage (fewer species of larger fishes). Even so, there are significant gaps for most of Queensland and in South-West WA. The broadest dataset is from the Soviet fleet (Figure 2A, purple points), which dates back to the 1960s and 1970s but has numerous identification issues, leaving few species that can be used with confidence. Prawn trawl surveys are more diverse (numerous species of smaller fishes and some mobile invertebrates) but restricted to the NE shelf from ~Sydney to Torres Strait and into the Gulf of Carpentaria. More than half of the

Australian continental shelf has no data for these types of species. Benthic sled surveys tend to be the most diverse (many sessile and mobile invertebrates) but are restricted to the GBR, Torres Strait and Gulf of Carpentaria with sparse sampling in the SE and on the western upper slope. Large tracts of shelf in eastern, southern and north/northwest Australia have not been sampled.

The modelled compositional differences (Figure 2, right) reflect the distribution of sampled sites, but also estimate how knowledge falls away with increasing geographic distance and changes in environment. Green shades indicate areas where any new samples are likely to be as similar to existing nearby samples as any pair of existing neighbouring samples (a possible exception being areas that are substantively based on Soviet data). Yellow shades indicate areas where composition is likely to be moderately dissimilar to existing nearby samples, with possibility of new species discovery. Orange shades indicate areas where composition is likely to be substantially dissimilar to existing samples, with new species discovery likely. Red shades indicate areas where composition is likely to be completely dissimilar to existing samples, with new species discovery highly likely. The analyses of compositional differences also demonstrate that neighbouring benthic sled samples tend to be much less similar to each other than neighbouring fish trawls. Prawn trawls were intermediate, but closer to fish trawls. This is because sled samples have greater variation in species diversity than trawl samples. For example, in the Great Barrier Reef, individual sleds and prawn trawls tend to have similar average richness (number of species), but each sled tended to add more new species and sled samples overall comprised many more species. As a consequence, ~450 prawn trawls in the GBR more completely described their biota (more green areas in Figure 2) than the sled with ~1200 samples in the same area. The red shades that characterise deeper waters around Australia, indicate the lack of any samples in most of these deeper waters. Any new samples would provide new information.

A: Tropical coral reefs surveys

B: Temperate rocky reefs surveys

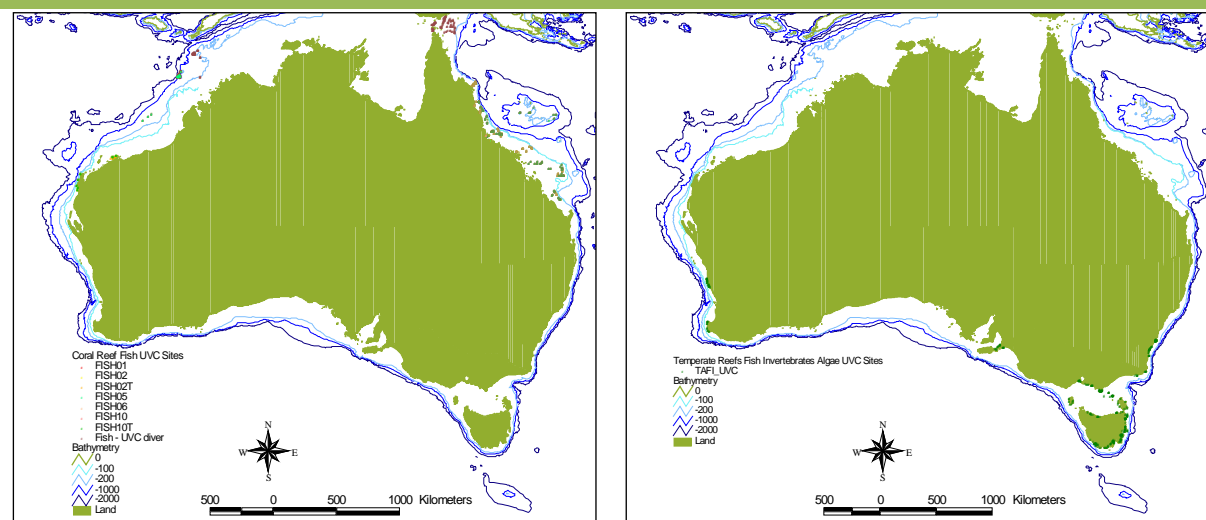


Figure.1. Maps of sampled sites for diver visual surveys of fishes and mega-invertebrates on tropical coral reefs and shallow temperate rocky reefs.

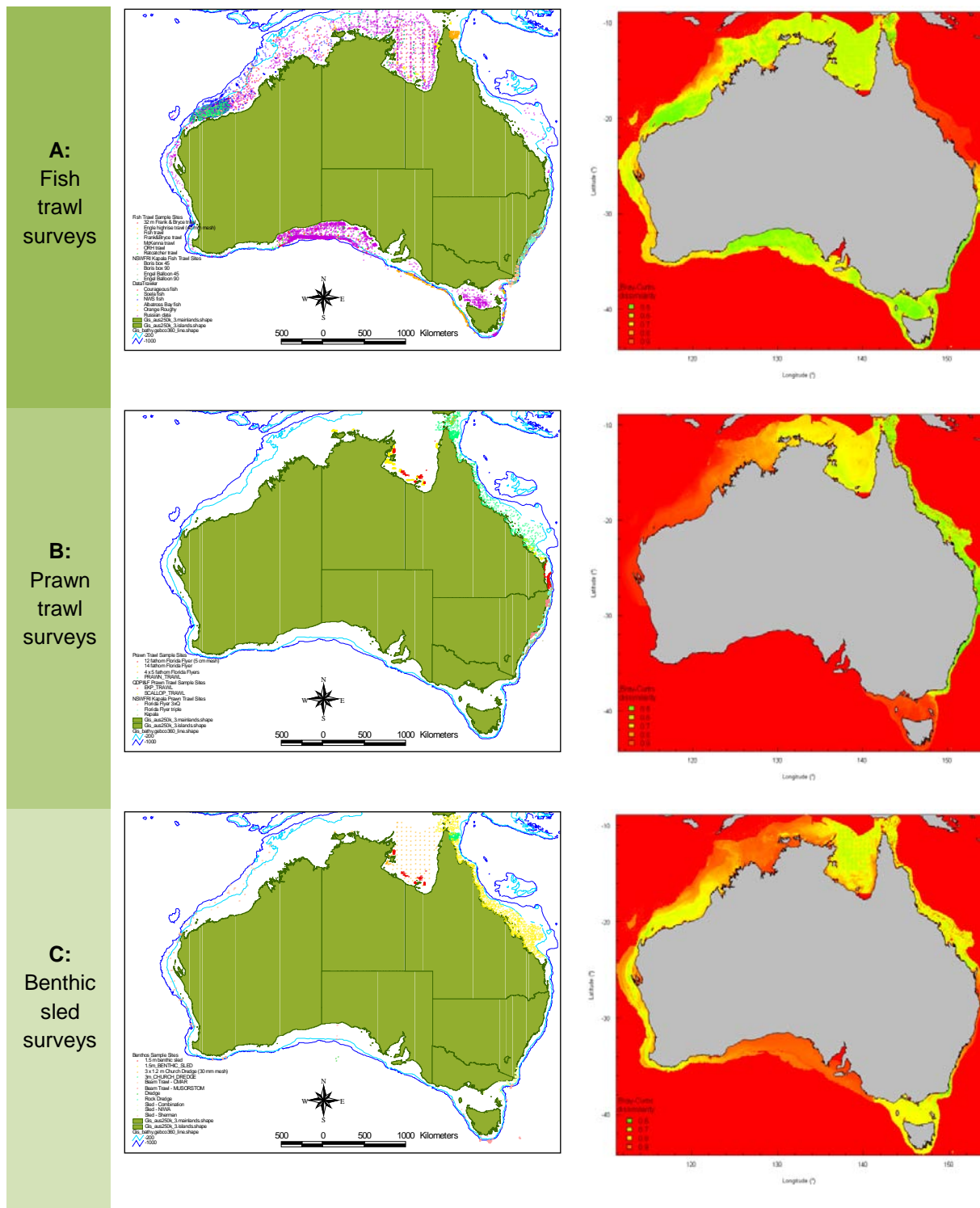
OPPORTUNITIES FOR MANAGEMENT AGENCIES AND FOR FURTHER WORK

The maps in Figure 2 indicate the types of data upon which shelf/upper slope biodiversity maps within each marine planning region are based. This is a scale below provincial bioregions that are based on entirely different data. The South-East, South-West and North-West Marine Regions are substantively based on fish trawl data; in particular the South-West Marine Region is substantively based on unreliable Soviet data. The East Marine Region is based on fish and prawn trawl data, but with

few invertebrates. The North Marine Region includes all three data types, though prawn trawl and sled are not complete.

These maps show that the deeper slope and seabed beyond is essentially unknown. Even on the slope there are large areas that are essentially unknown for one or more types of biota and, unfortunately, extensive analyses of biological surrogacy (as referred to in the Hub's [final report](#)) demonstrate that patterns of composition in one of these types of biota cannot be used as a surrogate for patterns in another at sub-provincial scales.

These identified gaps are areas of greatest uncertainty in national predictions and hence are priorities for future field programs that aim to fill significant gaps in our current knowledge of large-scale biodiversity; especially where there are also implications for management for sustainability & conservation.



B. Gap identification for sediment and bathymetry

An essential step towards improving the performance of models of seabed biodiversity is to identify gaps in the spatial distribution of the seabed environmental data that are employed as predictive variables. This surrogates approach to biodiversity mapping is required due to the limited distribution of biological data within the vast and diverse Australian Exclusive Economic Zone.

Project objective

- Identify significant gaps in bathymetric and seabed sediment data held at Geoscience Australia. These data and their derivatives represent useful surrogates of marine biodiversity.

Approach

Sediment data

Geoscience Australia's seabed sediment database, MARS, currently contains records for approximately 13,000 surficial seabed sediment samples collected by a range of Australian and international marine science organisations and private industry, and new records are regularly added. The majority of samples was collected by dredges, grabs or cores. Standard data for sediment samples include grainsize (sieve fractions or near-continuous grain-size population distributions derived from laser analysis), and/or percent calcium carbonate.

To identify the spatial distribution of the sediment samples, Arc-GIS shapefiles of the carbonate and grainsize data were generated. These point data were aggregated

on a 5 km resolution grid to create a polygon shapefile for the AEEZ.

Bathymetry data

GA currently holds a range of bathymetry data types for the AEEZ and surrounding areas collected by a range of Australian and international marine survey organisations. Some data types, such as Laser Airborne Depth Sounder (LADS) and satellite-derived bathymetry, are not included as these are yet to be appropriately quality controlled and formatted for input to a common database.

Arc-GIS polygon shapefiles were created that incorporate the area covered by all multibeam data held by Geoscience Australia to depict the coverage of the best quality bathymetry data. The tracklines of these data total 8,515,789 km. The individual multibeam data shape files were clipped to the AEEZ area and overlapping polygons merged to allow calculation of the total area for which multibeam data are available. Secondly, all bathymetry data that were incorporated into the national 250 m bathymetry grid were merged to calculate the density of depth soundings throughout the AEEZ.

Key findings

Sediments

More than 85% of samples in MARS were collected in water depths <200m (i.e. on the continental shelf; Figure 1). Only 4% occur in water depths >2,000m, although this comprises >55% of the AEEZ area (Figure 1). Sediment data were found to attain densities of greater than one per 25 km² grid cell in only

approximately 36,300 km² (3%) of the AEEZ. Sample densities are highest (regularly exceeding 1 sample per 25 km²) in the north east (GBR - Cape York) and south east of the continental shelf and in the Gulf of Carpentaria (Figure 1).

The distribution of samples is influenced by the relatively small areas targeted in most marine surveys, and the large effort and cost required to access deepwater and offshore areas. Additional data for the continental shelf and deep water areas could be derived from seabed samples collected by other marine agencies and universities. In particular, relatively few samples are from State waters (Figure 1b) as these areas have not been a priority for Commonwealth Government surveys or procurement for Commonwealth marine planning research. However, it is likely that samples if not data for these areas exists in relevant State Government agencies and universities.

Bathymetry

Multibeam sonar data has sounding densities orders of magnitude greater than single-beam echo sounder data, as shown by comparing Figure 2 (multibeam sonar data coverage) and Figure 3 (data density for all bathymetry data). Even in the deeper areas of the AEEZ (eg 4000 m depth), multibeam data can support bathymetry grids of at least 200 m resolution. However, currently multibeam bathymetry data cover

approximately 1,375,400 km² or 20% of the AEEZ (excluding external territories; Figure 2). Broad areas in which there is predominantly a very low density of depth soundings include the Northern and Northwest marine regions, the Great Australian Bight, Bass Strait and the central region of the Great Barrier Reef (Figure 3). Useful additional bathymetry data coverage could be obtained by adding LADS data collected by the Australian Hydrographic Service (AHS) and industry, and satellite-derived bathymetry for a range of shallow-water areas. Also, additional multibeam data for the AEEZ may be obtained from overseas marine survey organisations that have collected data during surveys in or transits through Australian waters.

To enable more robust gridding of Australia's bathymetry data and to facilitate incorporation of LADS and satellite-derived data into a combined dataset, all data need to be checked for quality then uploaded to a specialised marine survey database (eg CARIS Bathymetry Database). This type of database would also enable confidence parameters to be generated for data extracted from the database, as well as for products derived from the data, such as digital relief models.

References

- Harris, P., Heap, A., Passlow, V., Sbaffi, L. Fellows, M., Porter-Smith, R., Buchanan, C., Daniell, J. 2005. Geomorphic Features of the Continental Margin of Australia. Geoscience Australia, Record 2003/30, Geoscience Australia, Canberra.
- Li, J., Potter, A., Huang, Z., Daniell, J.J., Heap, A.D., 2010. Predicting seabed mud content across the Australian margin: comparison of statistical and mathematical techniques using a simulation experiment. Geoscience Australia Record 011/2010.



MANAGEMENT AND RESEARCH OPPORTUNITIES

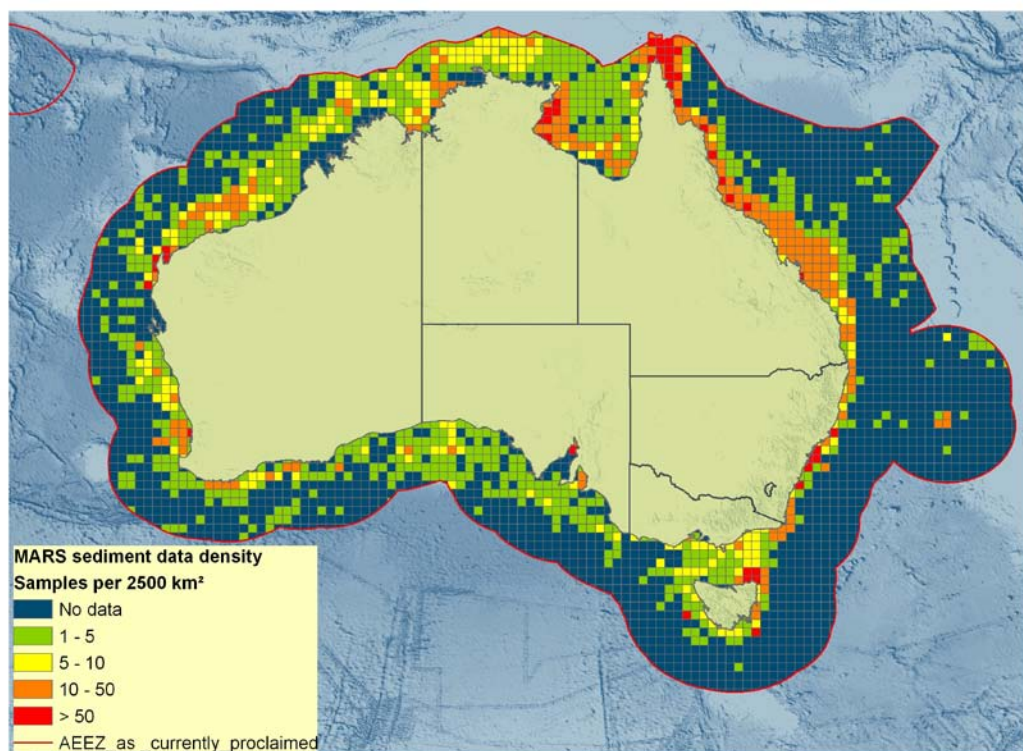
Sediments

It is clear that additional sediment samples are required in most areas of the AEEZ. More advanced spatial interpolation methods, which are used to generate grids from the sediment point data, can provide more accurate representations of sediment distribution and therefore enhance the value of the existing data (eg Li et al., 2010). Sediment grids are employed in biodiversity models and adopting better interpolation methods may therefore improve biodiversity predictions. Also, mapping the density of samples relative to seabed complexity may provide a better estimate of the density of samples required to accurately represent the distribution of sediment types in different areas. For example, at the broadest scale, seabed complexity is generally much higher on the continental shelf and slope than in abyssal areas. Seabed geomorphic features, an existing dataset, capture aspects of seabed structural complexity and may be useful for this purpose (Harris et al., 2005).

Bathymetry

The population of a more useful specialised marine bathymetry database with national coverage is an important objective for enhancing this dataset and the useful seabed parameters that are derived from the data. A staged approach to this issue is a realistic, cost-effective option. For example, on a region by region basis, once data is checked for quality, all existing types of bathymetry data for the region could be uploaded to the database. This could usefully start with NERP priority areas such as Northwest and Northern Australia.

A



B

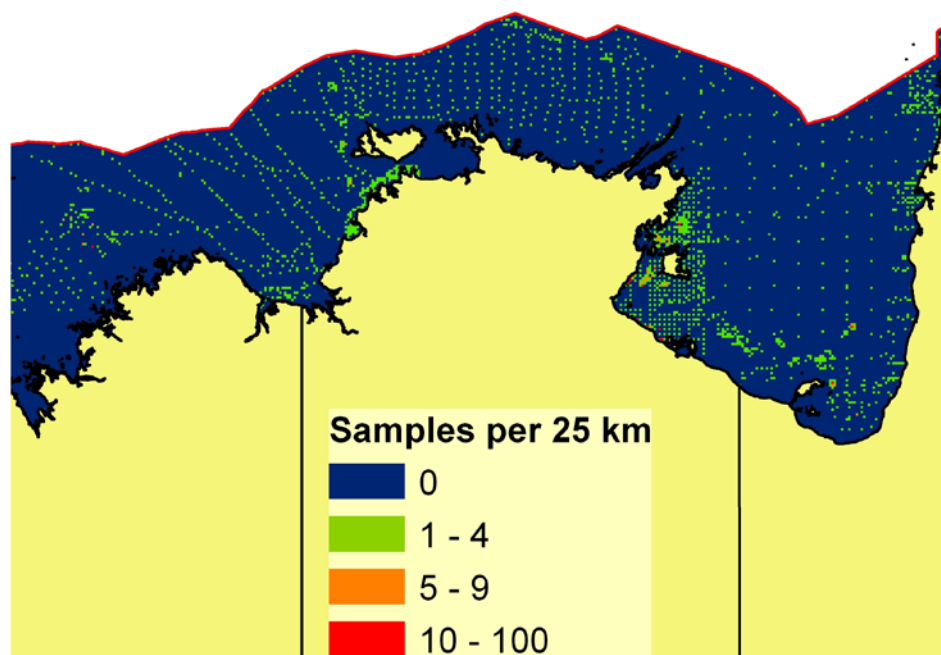


Figure 1:

A: Sediment data density within the Australian Exclusive Economic Zone ranges from up to 250 samples to < 1 sample per 2500 km².

B: Sediment sample density per 5 km grid cell (25 km²) for the Northern and part of North-West Marine Regions.

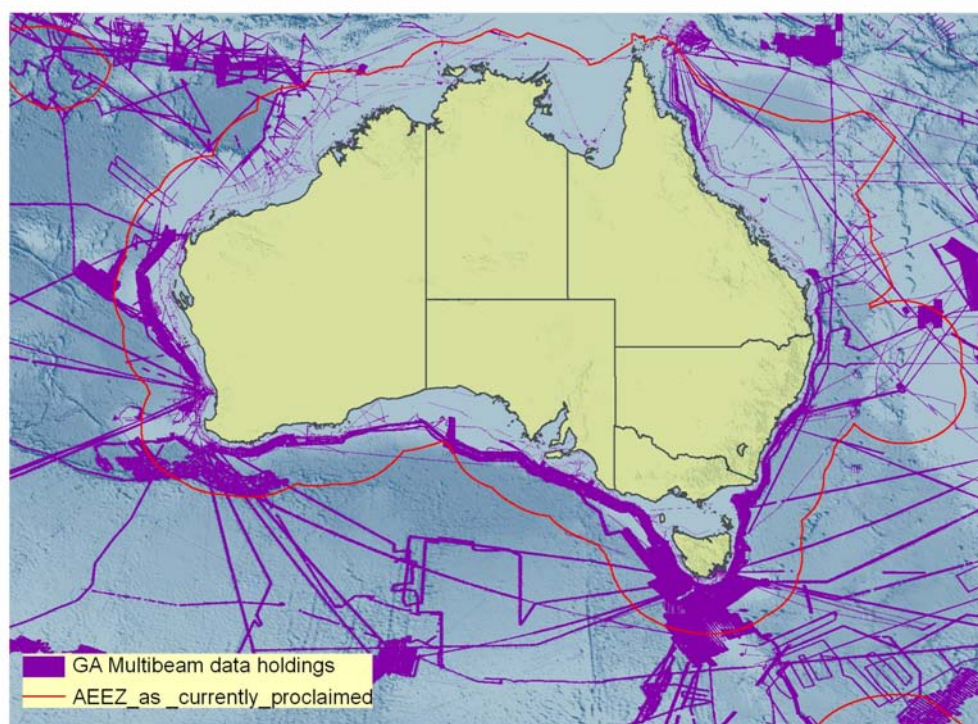


Figure 2: The coverage of multibeam sonar bathymetry data held by Geoscience Australia.

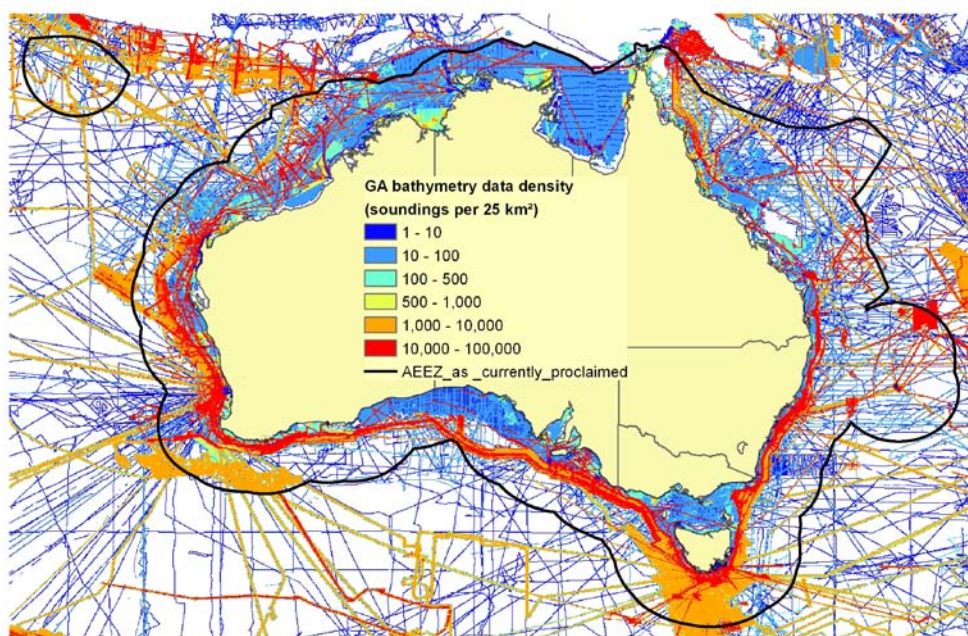


Figure 3: Sounding density per 5 km grid cell (25 km²) for all data that are included in the national 250 m bathymetry grid.

TASK: New biological and physical data layers

Incorporate new data layers into Hub publically available data directory for future biodiversity mapping - identify and collate new available surrogate layers for use in predicting national biodiversity, in particular new satellite and derived layers

In the Marine Biodiversity Hub, many environmental data layers were found to have utility as surrogates or predictors for marine biodiversity. Several new predictor layers, in addition to those already captured by the Marine Biodiversity Hub, have since become available and have been acquired by this CERF Transition task, along with several updates of existing layers.

Project objectives

- Scope availability of additional surrogate data layers,
- Acquire the readily available data layers and make them available for supporting future SEWPaC needs.

Approach

Nine new environmental datasets were acquired, and 4 existing datasets were updated with more recent years of data. The sources of new and updated data include: the Ocean Productivity group at Oregon State University, the NOAA National Geophysical Data Center (NGDC), the International Comprehensive Ocean-Atmosphere Data Set (ICOADS), and Geoscience Australia. The datasets were re-sampled to the existing Marine Biodiversity Hub

0.01° resolution spatial grid for Australia's EEZ of more than 8 million cells. Many of the new and updated layers are in the form of monthly means over 8 to 20 years, representing nearly a billion data records in total. For temporal datasets, we calculated long-run Jan-Dec month and annual statistics (min, max, mean, median, std dev. and seasonal range (s.rng)).

- The principal product of the Ocean Productivity group is an estimate of global ocean Net Primary Production by the new Vertically Generalized Production Model (VGPM) ([Behrenfeld and Falkowski 1997](#)). This is calculated using the Sea Surface Temperature (SST), Chlorophyll (CHL) and Photosynthetic Available Radiation (PAR) derived from the MODIS and SeaWiFS satellites. We also calculated new estimates of Export Particulate Organic Carbon flux (EPOC) to the seafloor using the exponential decay model ($EPOC_{depth} = 3.523 \times NPP \times depth^{-0.734}$, Pace et al 1987) and Benthic Irradiance ($BIR = PAR \times e^{(K_{490} * Depth)}$). We have also acquired 20 years of monthly ICOADS data for rainfall, cloud cover, surface winds from NOAA/NASA, which can be exploited further under NERP.
- We also updated the Marine Biodiversity Hub's Bathymetry layers with the most recent 2009 Geoscience Australia 9" Bathymetry DEM and extended its extent with data from NOAA/NGDC ETOPO1 1-arc-minute global DEM.
- Maps of the Australian EEZ for the processed data layers were prepared.

Key findings

- Eleven new ⁽ⁿ⁾ or existing datasets were acquired or updated ^(u) (see table 1 on page 55) and stored in a secure database. Each of these datasets has two, 3, 6 or more associated statistics as data layers for the Australian EEZ. In this short report, there is space for maps of only six of the new or updated data layers that have been acquired or derived (see Figure 1 on page 56). These additional and updated layers provide enhanced capability for modelling biodiversity distribution and abundance for planning and management purposes, including the possibility of predictive monitoring.
- Compared with the datasets collated by the previous Marine Biodiversity Hub, the longer data time series and additional datasets expanded the data storage and data handling challenges by two orders of magnitude.

References

- Behrenfeld, MJ, Falkowski, PG (1997) A consumer's guide to phytoplankton primary productivity model. *Limnology and Oceanography*, 42:1479-1491
- Pace, M. L., G. A. Knauer, D. M. Karl, J. H. Martin. 1987. Primary production, new production and vertical flux in the eastern Pacific Ocean. *Nature* 325:803-804.

MANAGEMENT AND RESEARCH OPPORTUNITIES

The new and updated layers acquired and derived by this Transition Task have added significantly to the current data holdings of Marine Biodiversity Hub, and can now be used to support implementation of Marine Bioregional Plans, ERIN species prediction, ALA prediction, and other SEWPaC marine planning needs. They will also support several new NERP Marine Hub projects - implementing RMPs, monitoring, and extending national biodiversity maps. Also, some additional layers have been identified (eg shelf sediment stability/mobility; ocean stratification index) that if available can be acquired by the new NERP Marine Hub. Further, while this task has calculated long term statistics for NERP use, now, given the increasing length of the time series there are opportunities for using higher temporal resolution aspects of the data for surrogacy and prediction purposes. These opportunities will be facilitated in the near future by the commissioning of a new CSIRO Marine and Atmospheric Research Queensland database server, which will expand storage and data manipulation capabilities for these larger, longer time series, climate sets.

Table 1: Eleven new ⁽ⁿ⁾ or existing datasets acquired or updated ^(u)

Data layer	Frequency	Date range	Raw resolution	Key statistics
^u Sea Surface Temperature (SST)	monthly	2002 Jul — 2010 Jun	0.083 deg	ann.mean, s.rng
^u Chlorophyll A (CHL)	monthly	2002 Jul — 2010 Jun	0.083 deg	ann.mean, s.rng
ⁿ Net Primary Productivity (VGPM)	monthly	2002 Jul — 2010 Jun	0.083 deg	ann.mean, s.rng
ⁿ Export Particulate Organic Carbon (EPOC)	monthly	2002 Jul — 2010 Jun	0.083 deg	ann.mean, s.rng
^u Benthic irradiance (BIR)	monthly	2002 Jul — 2010 Jun	0.083 deg	ann.mean, s.rng
ⁿ Photosynthetic Available Radiation (PAR)	monthly	2002 Jul — 2010 Jun	0.083 deg	ann.mean, s.rng
ⁿ NASA – TRMM Monthly Rainfall – 3B43	monthly	1990 Jan — 2010 Dec	0.25 deg	ann.mean, s.rng
^u ICOADS – SST	monthly	1990 Jan — 2010 Dec	1.0 deg	monthly means
ⁿ ICOADS – Sea Level Air Temp	monthly	1990 Jan — 2010 Dec	1.0 deg	monthly means
ⁿ ICOADS – Sea Level Pressure	monthly	1990 Jan — 2010 Dec	1.0 deg	monthly means
ⁿ ICOADS – Scalar Wind	monthly	1990 Jan — 2010 Dec	1.0 deg	monthly means
ⁿ ICOADS – Cloudiness	monthly	1990 Jan — 2010 Dec	1.0 deg	monthly means
ⁿ ICOADS – Humidity	monthly	1990 Jan — 2010 Dec	1.0 deg	monthly means
^u Bathymetry	n/a	up to 2009	variable	depth,slope,aspect

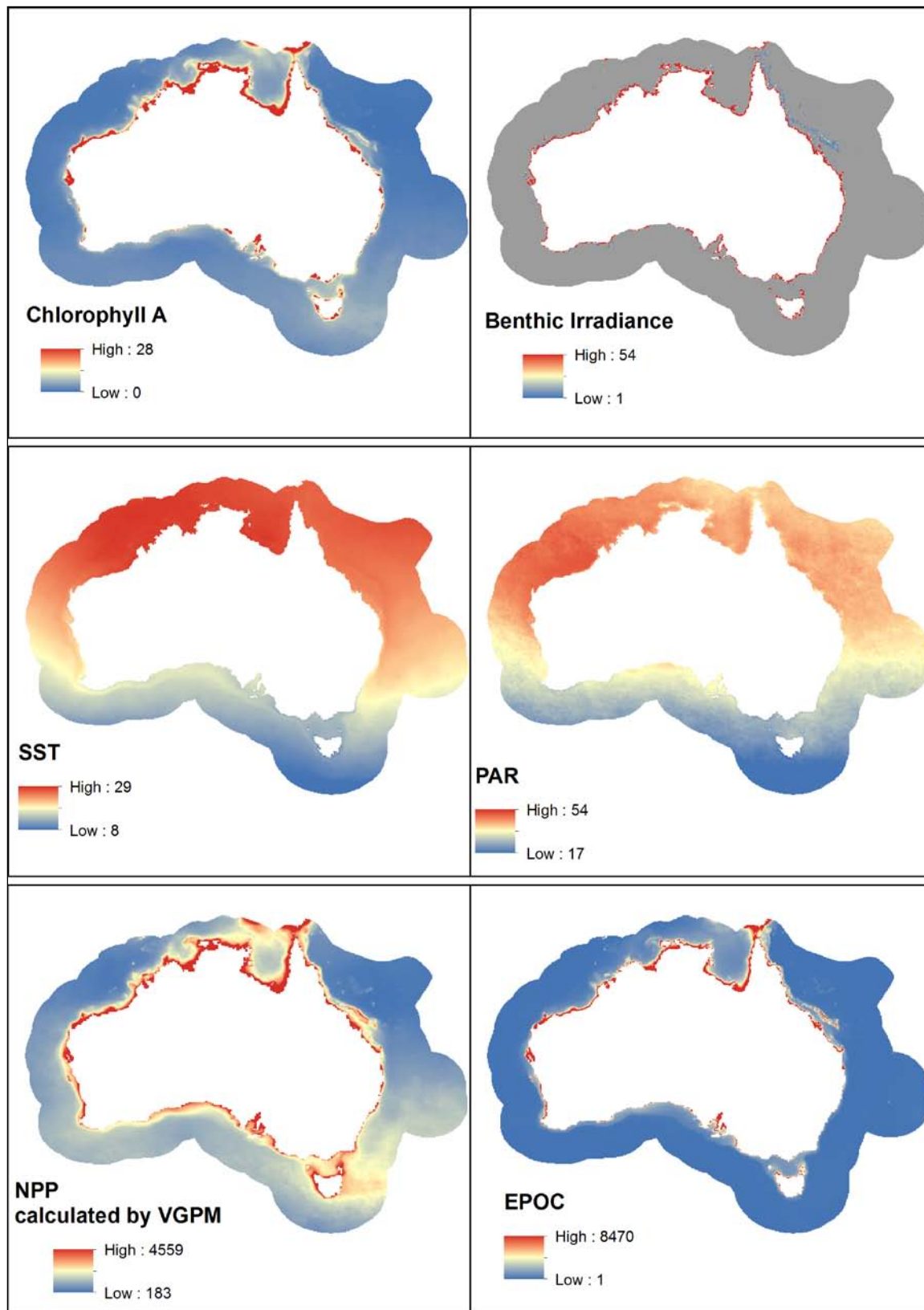


Figure 1: Maps of six of the new or updated data layers that have been acquired or derived

EXTENSION PROGRAM 2:

Extend statistical methods developed for marine biodiversity mapping to support further application to meet SEWPaC needs

Several novel methods for analysing biodiversity information were developed by the Marine Biodiversity Hub for mapping biodiversity at a national scale. One of these methods shows promise in other areas of biodiversity management (analysis of threatened species distributions), but this requires extending the method (so that it can accommodate abundance data as well as count data). Further development of two other methods supports improved use of varied data (eg. surveys with different gears), and provides estimates of uncertainty for the predictions provided to SEWPaC to support marine bioregional plans. Generalisation of these methods supports national mapping of biodiversity, implementation of marine bioregional plans and may assist in predicting the ranges of threatened species.

TASK: Uncertainty for biodiversity prediction

Develop a method for representing uncertainty for the predictions of compositional biodiversity patterns from the modified random forest approach used to generate national biodiversity maps in support of marine bioregional planning

The gradient forest method has been one of the key approaches used to provide the national coverage of regional biodiversity maps that contributed to SEWPaC's identification of potential areas for inclusion into the national representative system of marine protected areas. This is a novel modification to the random forests method described on page 8, however, like all multivariate biodiversity composition methods, the random forest method has been lacking an accompanying map of uncertainty in the predictions. This section shows how uncertainty can be quantified and mapped, and it is

demonstrated with an example from the Southeast Marine Region. This extension will enable managers to understand the robustness of the predicted compositional patterns and facilitate greater confidence in the use of biodiversity maps to inform the conservation and management of marine biodiversity.

Project objectives

- To find an objective measure of uncertainty for the predicted distribution of biodiversity and to represent this uncertainty geographically on a map.

Approach

The gradient forest approach to biodiversity composition on page 10 is a method that transforms environmental variables or surrogates so that they

represent how species respond to them (biological space). In biological space, sites close together are expected to have similar biodiversity composition whereas sites further apart are expected to have more different biodiversity composition. We wish to assess the uncertainty of this prediction by somehow comparing it with the 'true' composition. Unlike species abundances, which are measured by counts or biomass, there is no direct measure of composition. One approach is to take the commonly used Bray-Curtis dissimilarity to be the 'benchmark', and to compare the Bray-Curtis dissimilarity of pairs of sites with their distance in biological space. The lack of fit of a monotone regression between these two dissimilarities would provide an overall measure of the uncertainty in biological space. However, being based on *pairs* of sites, this approach would not readily allow one to assess the uncertainty at a *single* site, still less the uncertainty of predictions at all un-sampled locations. To overcome this limitation the Bray-Curtis dissimilarities must be converted into a form that allows direct comparison at individual sites. This is done by ordination of the sites using non-metric multi-dimensional scaling on the dissimilarities. The result of the ordination is a space where more similar sites are close together and dissimilar sites are further apart; this space is therefore comparable to the biological space generated by the gradient forest

approach. However, the ordination space has an arbitrary orientation and so must be rotated (using a method called Procrustes rotation) so that the sites align as closely as possible with those in biological space. The absolute distance between the sites in biological space and the rotated ordination space is called the Procrustes residual; this represents the uncertainty in biological space, and these residuals can be plotted on a map. To find the uncertainty at a new un-sampled location, we assume that this is similar to the uncertainty at sites that are close in biological space. We therefore smooth the uncertainty in biological space using a Gaussian kernel. The amount of smoothing is dictated by the distribution of sites in biological space (see Figure 1 on page 60). The kernel should not be too smooth (leading to lack of detail), nor too localised (leading to a noisy smoothing due to too few contributing sites). For the SEMR, a width of 0.005 was chosen as a compromise, to prevent isolated sites having too much influence (eg the red site in the *SovietSE* survey). The smoothed values can then be plotted on a map (see Figure 2 on page 60).

Key findings

The method has been applied to the Southeast Marine Region. The uncertainty is least on the east coast, in the middle of Bass Strait and some parts of the shelf south and east of Kangaroo Island. The uncertainty is greatest off parts of the coast of western Victoria, the west coast of Tasmania and the eastern Bass Strait. The uncertainty tends to be higher in poorly sampled areas, with the exception of areas around Kangaroo Island which have environments that are similar to well sampled areas further east. Although the smoothing is carried out in biological space, it is generally the case that the uncertainty is similar to that of sites that are nearby geographically. Because the uncertainty is estimated by smoothing over nearby sites, there is a degree of uncertainty in the uncertainty itself. This is better appreciated from the maps in the gap analysis areas (see page 44), where the modelled dissimilarity is large are also areas where the uncertainty is less well estimated.

MANAGEMENT AND RESEARCH OPPORTUNITIES

As far as we are aware, this is a novel method, and represents the first time uncertainty in biological composition has been mapped anywhere in the world. The method is a valuable tool that has been successfully applied to the South-East Marine Region and will be useful to understand the uncertainty in biodiversity composition in other regions around Australia to inform the management of marine biodiversity. The uncertainty information supplements that from the gap analysis (see page 44).

Where the gap analysis indicates which areas should be candidates for further sampling, the uncertainty measure can be used to prioritise the sampling within such areas. However, even within well-sampled regions there are areas with large uncertainty, which may correspond with high beta diversity. It will be more complicated to achieve representative management in areas with high beta diversity, due to the rapidly changing biological composition over relatively short distances. Such results can help to guide future research to improve the characterisation of biological composition.

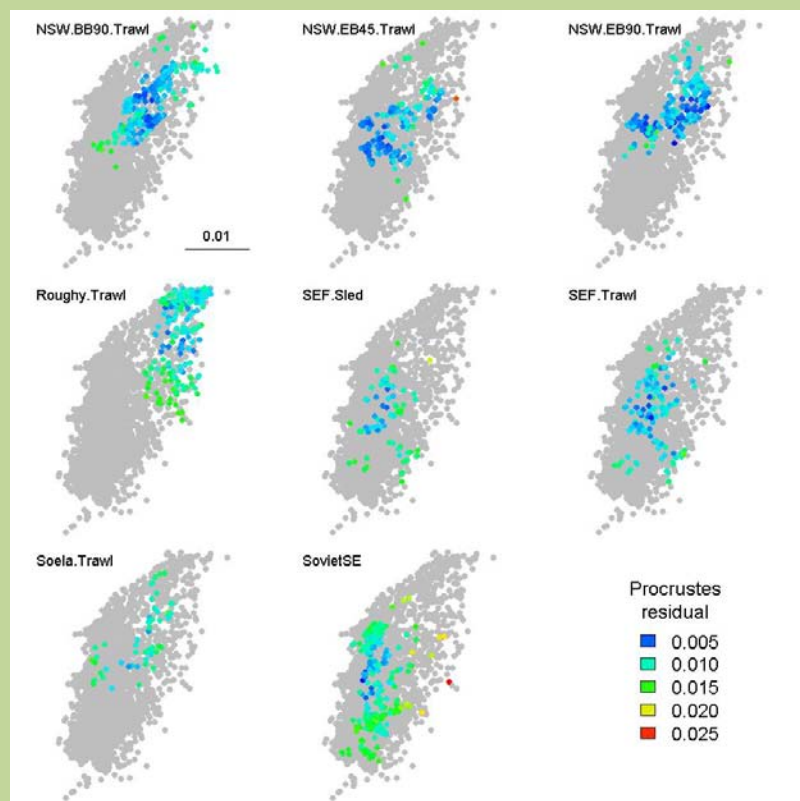


Figure 1. Biplots of first 2 principal components in biological space, explaining 89% of variation of cells in the SEMR (grey). Overlaid is the location of the survey sites separately for each of the 8 surveys. The colour denotes the distance in biological space between the sites and the MDS ordination of inter-site Bray-Curtis dissimilarity after Procrustes rotation. A distance of 0.01 is shown by the line.

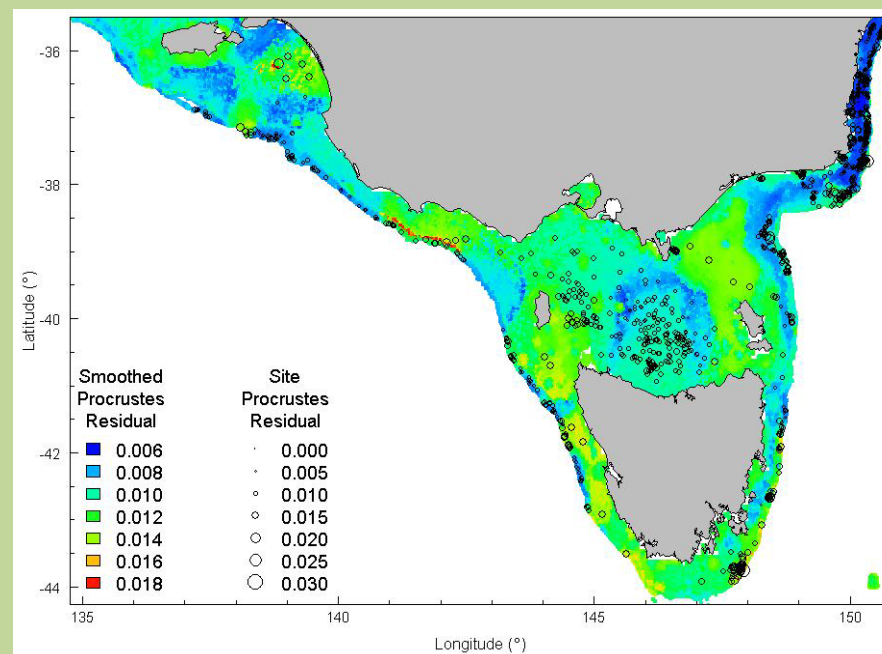


Figure 2. Map of the smoothed Procrustes residual (colour) for the SEMR, with sites from all 8 surveys overlaid (circles). The smoothing of the site residual is performed in biological space (see Figure 1) with a Gaussian kernel of width 0.005.

TASK: Extension of species archetypes method

Produce R code and report on example analysis and critical comparison of groups from a variety of sources - extend the modelling framework that underpins the creation of species-archetype groups to produce species-archetypes from abundance and biomass data types, and improve variable selection methods, with potential application to threatened species mapping

Project objectives

- During the CERF Marine Biodiversity Hub, a novel statistical method to analyse many species' presence/absence data simultaneously, was developed. This methodology allowed grouping similar species together, thus enabling one species to 'borrow strength' from other species that respond similarly to the measured environmental gradients.
- These unobserved species groups were labelled 'species archetypes' and reflect a series of archetypal ways that a species may react according to a number of different environmental gradients.
- Other types of data (species abundance and biomass) may provide much more information about the distribution of individual species, and hence the species-archetypes. This project

addressed this problem by developing and implementing extensions of the underlying statistical methodology.

Key findings

Models were developed to appropriately handle two biodiversity data types, abundance and biomass. Abundance data are counts of the number of individuals of each species in a sample. Biomass data consists of the total weight of all individuals in a sample. Models using simulated data are able to accurately estimate the relationships between species abundance and biomass and environmental gradients, grouping species with similar responses together. Comparisons with the results from presence/absence species archetypes suggests that modelling abundance data tends to produce fewer groups. An R-package that implements the methods has been produced.

Approach

The data were modelled using mixtures of generalised linear models, as for species-archetypes based on presence/absence data. For abundance data we used a negative binomial model that allows for counts and increasing variance as counts increases. Mixtures were over environmental gradients only, each species was allowed a separate model intercept (the overall mean number). This was done as two species might have an identical response to the environment, but have very different average abundances.

For example, one species might be rare, the other common, but both increase their abundance with increasing temperature. For biomass data, a flexible model based on the tweedie distribution was used. This model describes continuous biomass data and allows for samples where biomass is zero. Each species also has a separate intercept for this model.

RESEARCH OPPORTUNITIES

The methods will be applied to a range of data sets from around Australia, both marine and terrestrial. We will continue development, improving covariate selection and identification of group number. We have identified collaborators who will assist us to implement this form of modelling for presence only data. The performance of the methodology will be compared with single species models and other forms of multivariate ecological models.

TASK: Extension of hybrid GLM/TREE approach

Production of annotated R code for statistical community - extend the new hybrid GLM/Tree approach to support use of varied datasets in biodiversity mapping and other analyses

Introduction

In biodiversity mapping and species distribution modelling, parametric models such as generalised linear models have the advantage of allowing a smooth, continuous dependency of the response on the underlying predictors. Tree models have a similar purpose, but are essentially non-parametric. They model the dependency as abrupt changes at automatically chosen split points on the predictors, partitioning the predictor space into discrete regions within each of which the mean response is constant. Tree models have the practical advantage of automating the selection of predictor variables and including detection of various forms of interaction between predictors to emerge as part of the model fitting process. As well as some predictive capacity, they can have a useful *exploratory* function, as tree models are easy to interpret.

Many situations exist in which a biological response is known in advance to have a smooth dependence on some variables, such as seasonal or diurnal fluctuations, which can be expressed in some parametric form, but with respect to other predictors, such as environmental variables, the form of the dependency can be

adequately captured by tree model techniques. The combined use of these methods will provide greater rigour and flexibility than a single technique in modelling complex relationships such as those of biodiversity mapping and species distribution modelling needed by SEWPaC to provide confidence in decision-making for managing the marine environment.

Project objectives

Develop techniques for combining generalised linear models with tree models, leading to models that have a smooth parametric form with respect to some specified predictor variables, but a tree structure with respect to others. The models will perform like generalised linear models, but will include a factor term resulting from a tree construction operation with respect to the specified tree predictors.

Consider three main model cases:

- Continuous responses, such as log biomasses where the underlying GLM will be an ordinary linear model.
- Binary responses, such as presence/absence where the underlying GLM will be a logistic regression model.
- Frequency responses, such as species counts where the underlying GLM will be a log-linear Poisson model.

Provide tools for constructing, investigating and manipulating such models in the R language, now a standard platform for many biodiversity mapping and species modelling problems.

Approach

We develop new models and tools by extending and modifying the strategy outlined in Chen *et al*, [1]. The steps in constructing a hybrid glm/tree model are as follows: (1) the parametric component is fitted first using standard linear or generalised linear model techniques; (2) given the previous estimate of the parametric component, a tree model is fitted using tree model techniques with the parametric component as an offset; (3) the parametric model is then extended to use the tree component as an additional factor in the model, and re-fitted; (4) the parametric component is then updated by extracting the linear predictor from the current model and subtracting the tree model (factor) component. This then becomes the new offset for the next tree model fitting. (5) Steps 2-4 are iterated until the model stabilises. The final model has the form of a generalised linear model, but predictions from it will require tree predictions of the tree factor. Formal statistical inference at this stage is difficult due to the well-known difficulties in this context with tree models. The methods might primarily be considered exploratory and suggestive of patterns and dependencies in the data that might be investigated further with other methods.

Key findings and outcomes

The algorithm we use is a modified version of that used in Chen *et al*. [1] that typically stabilises much faster. For the continuous and binary data cases we have developed a suite of modelling tools including fitting, displaying and manipulating models in standard ways. For the count data case we have a working strategy with an illustration, but at this stage the tools are incomplete. The main challenge was the binary case. This required new algorithms to be devised for fitting binary classification tree models with offsets, and programming them. At this stage we have three prototype tools, namely: (1) a simplistic version based on an illustrative suggestion due to T. Therneau *et al* (Pers. comm.), which works correctly, but is too slow to be of much practical use; (2) a new faster version implemented mostly in R code that develops a suggestion of G. Robinson and L. Ryan (CSIRO/Pers. comm), which shows promise; (3) an even faster version of our own, again in R code, but limited to the special case of unweighted binary regression with the logistic link. This is the important case for most ecological applications, and though fast enough to allow some useful applications is still too slow to provide a complete practical method. A fully functional solution could probably be based on prototype 2, but it will need to be re-coded in a faster language such as C and linked to R.

Reference

Jinbo Chen, Kai Yu, Ann Hsing, and Terry M. Therneau, "A Partially Linear Tree-based Regression Model for Assessing Complex Joint Gene–gene and Gene–environment Effects", *Genetic Epidemiology* 31: 238–251 (2007)

MANAGEMENT AND RESEARCH OPPORTUNITIES

Hybrid GLM/Tree models have important advantages over existing methods by using the complementary strengths of each method to offset the weaknesses of the other. They have the potential to be useful exploratory tools and a potential predictive method for biodiversity mapping and species distribution modelling, as well as applications in a wide variety of other fields.

Example:

In the GBR mapping project the lagoon region was sampled separately using trawl and epibenthic sled.

*To illustrate the technology we chose one species, *Trachypenaeus anchoralis*, a prawn, with a reasonable capture rate in both devices. Hybrid linear/tree models were fitted to the log-biomass captured with an assumed smooth component in depth and a tree component built from a suite of environmental predictors.*

Figure 1 below shows estimated smooth components from the trawl (left) and sled (right) sampling, after allowing for other influences by the tree component. The profiles are similar, indicating a local maximum at about 15m in both cases. The central line is the estimated mean profile, with pointwise confidence regions indicated by broken lines. The points are the partial residuals after elimination of the tree components.

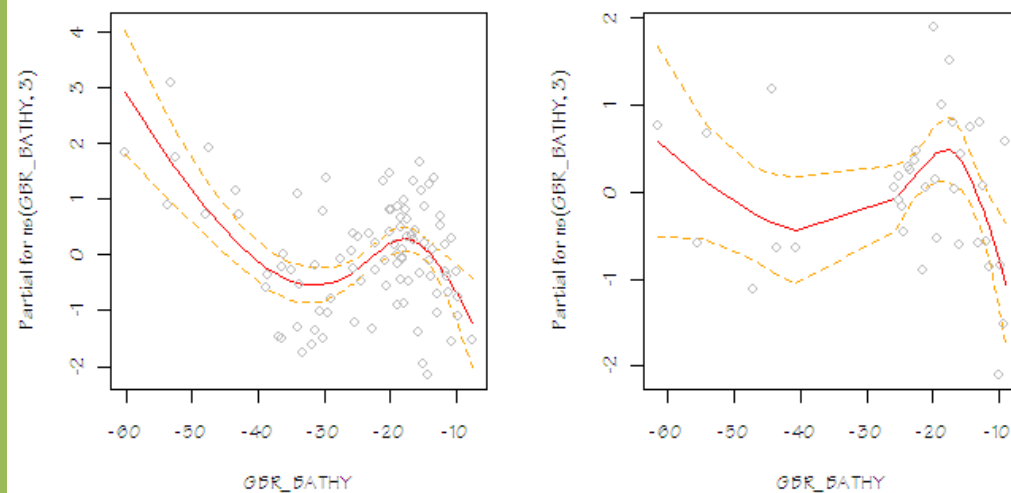


Figure 1: Estimated smooth components from the trawl (left) and sled (right) sampling, after allowing for other influences by the tree component.



Australian Government
Department of Sustainability, Environment,
Water, Population and Communities



Australian Government
Geoscience Australia



Australian Government



AUSTRALIAN INSTITUTE
OF MARINE SCIENCE



The Marine Biodiversity Hub is funded through the Commonwealth Environment Research Facilities Program (CERF), administered through the Australian Government's Department of Sustainability, Environment, Water, Population and Communities. The key aim of CERF is to provide sound advice to inform environmental public policy objectives and to better the management of Australia's unique environment. (Our stakeholder partners are: AFMA, APPEA, CFA, DAFF, SEWPaC, the Sustainable Tourism CRC, and WWF Australia)
