

Analysis of relationships between seabed species/assemblages and their physical environment using Random Forests statistical methods

C. Roland Pitcher ^{1*}, Nicholas Ellis ¹, Peter Lawton ^{2*},
Lewis S. Incze ⁴, Stephen J. Smith ²,
Thomas C. Shirley ³, Chih-Lin Wei ⁵, Gilbert T. Rowe ⁵,
Nicholas H. Wolff ⁴, Michelle E. Greenlaw ², Jessica A. Sameoto ²

¹ Commonwealth Scientific and Industrial Research Organisation, Australia

² Fisheries and Oceans Canada, Canada

³ Harte Research Institute, Texas A&M Univ.-Corpus Christi, USA

⁴ University of Southern Maine, USA

⁵ Texas A&M Univ.-Galveston, USA

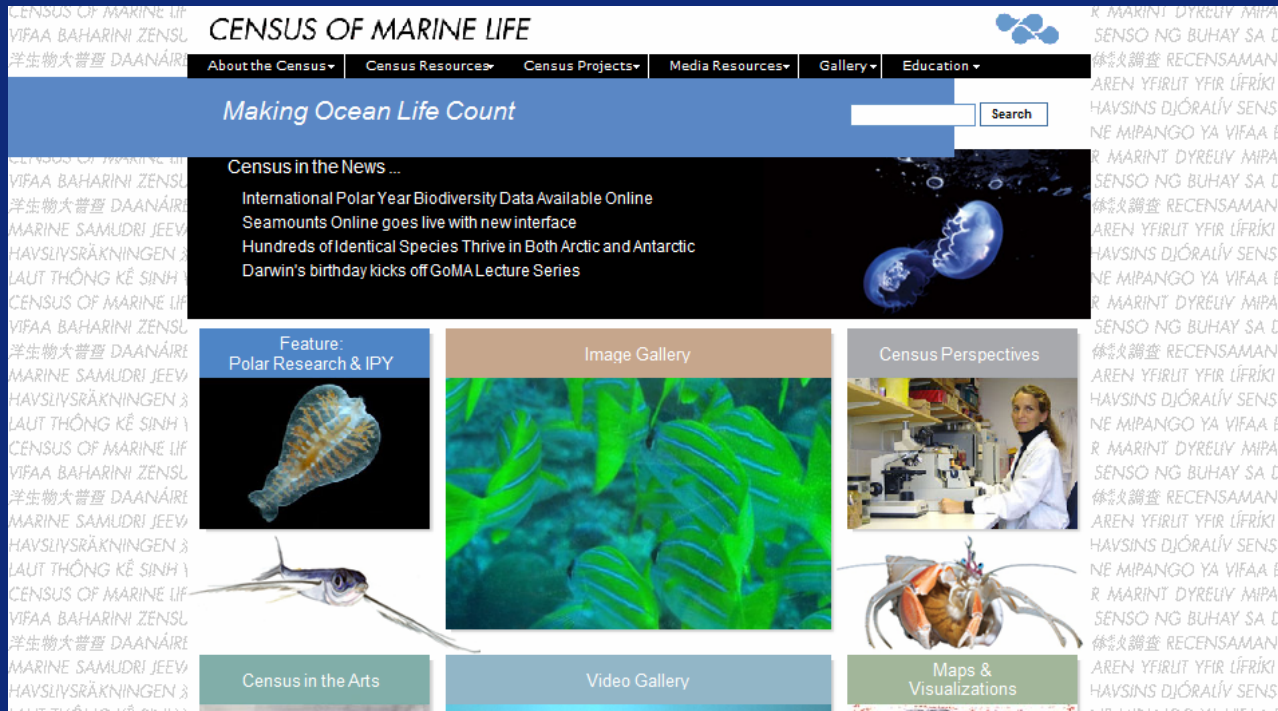
* Roland.Pitcher@csiro.au

* Lawtonp@mar.dfo-mpo.gc.ca

www.marinebiodiversity.ca

Census of Marine Life (CoML: www.coml.org) Synthesis Phase

- Integrate and synthesize the vast information gathered by CoML Projects and other research activities into common themes and overarching messages to ensure comprehensive content when final results are released in 2010
- 22 new cross-cutting synthesis projects were funded in 2008



The screenshot shows the homepage of the Census of Marine Life website. The header includes the site title "CENSUS OF MARINE LIFE" and a navigation menu with links for "About the Census", "Census Resources", "Census Projects", "Media Resources", "Gallery", and "Education". A search bar is located on the right side of the header.

The main content area features a large blue banner with the text "Making Ocean Life Count". Below this, there is a section titled "Census in the News ..." with a list of news items: "International Polar Year Biodiversity Data Available Online", "Seamounts Online goes live with new interface", "Hundreds of Identical Species Thrive in Both Arctic and Antarctic", and "Darwin's birthday kicks off GoMA Lecture Series".

Below the news section, there are several featured content blocks:

- Feature: Polar Research & IPY**: Includes an image of a jellyfish.
- Image Gallery**: Includes an image of green leaves.
- Census Perspectives**: Includes an image of a scientist in a lab coat.
- Census in the Arts**: Includes an image of a fish.
- Video Gallery**: Includes an image of a crab.
- Maps & Visualizations**: Includes an image of a crab.

Proposed Objectives

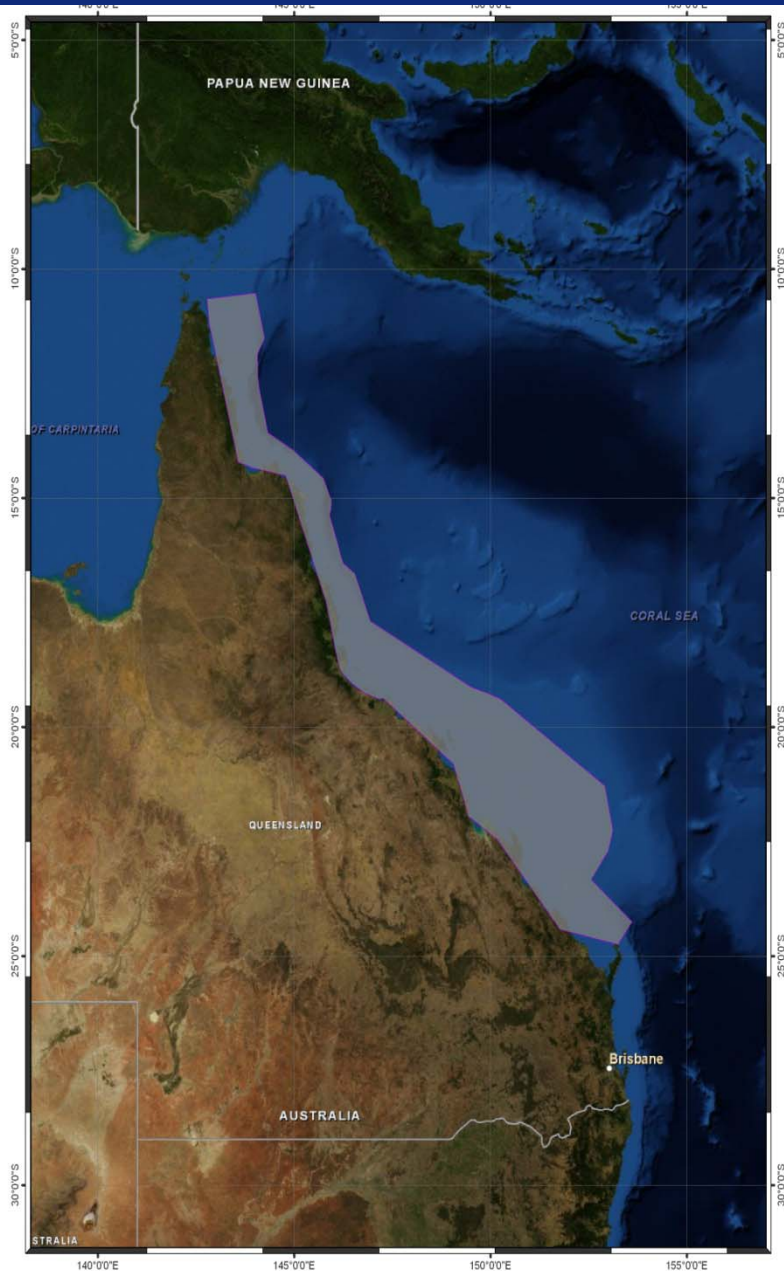
For defined seabed “biological patterns” (individual species, multi-species assemblages, and some diversity attributes) sampled over mesoscales:

- Examine extent to which physical surrogates may explain biological patterns;
- Rank importance of physical variables for structuring biological patterns;
- Examine common biological responses to physical gradients;
- Identify critical values for physical variables corresponding to 'threshold' changes

Overall Scope

- Short-term synthesis project:
 - Oct 2008: First workshop (Halifax, NS, Canada)
 - May 2009: Second workshop (Saint Andrews, NB, Canada)
 - Synthesis paper to be completed by Dec 31, 2009
 - Other papers (technical, within region, detailed) to follow
- Component programs:
 - GBR: Great Barrier Reef Seabed Diversity Program
 - GoMA: Gulf of Maine Area Program
 - GoMEX: Gulf of Mexico, Past Present and Future
 - DGoMB: Deep Gulf of Mexico Benthos Program (an affiliated project within COMARGE: Continental Margin Ecosystems)
- Similarities and differences:
 - Mesoscale programs, biological data encompassing range of physical environmental conditions, data availability, multiple sampling approaches, recent and/or historical studies

Scope: Great Barrier Reef



Comprehensive benthic surveys from 2003-2005

- Fisheries and conservation management drivers
- epibenthic sled: 1,190 sites; 4,723 species from >15 phyla
- prawn trawl: 457 sites; 3,510 species from >12 phyla
- Total >5,344 species
- Physical variables: 28 oceanographic factors
- Development of modified random forest methodology



Australian Institute
of Marine Science



CSIRO



Department of
**Primary Industries
and Fisheries**

Queensland Government



Queensland **Museum**
Queensland Government



Australian Government

Fisheries Research and
Development Corporation

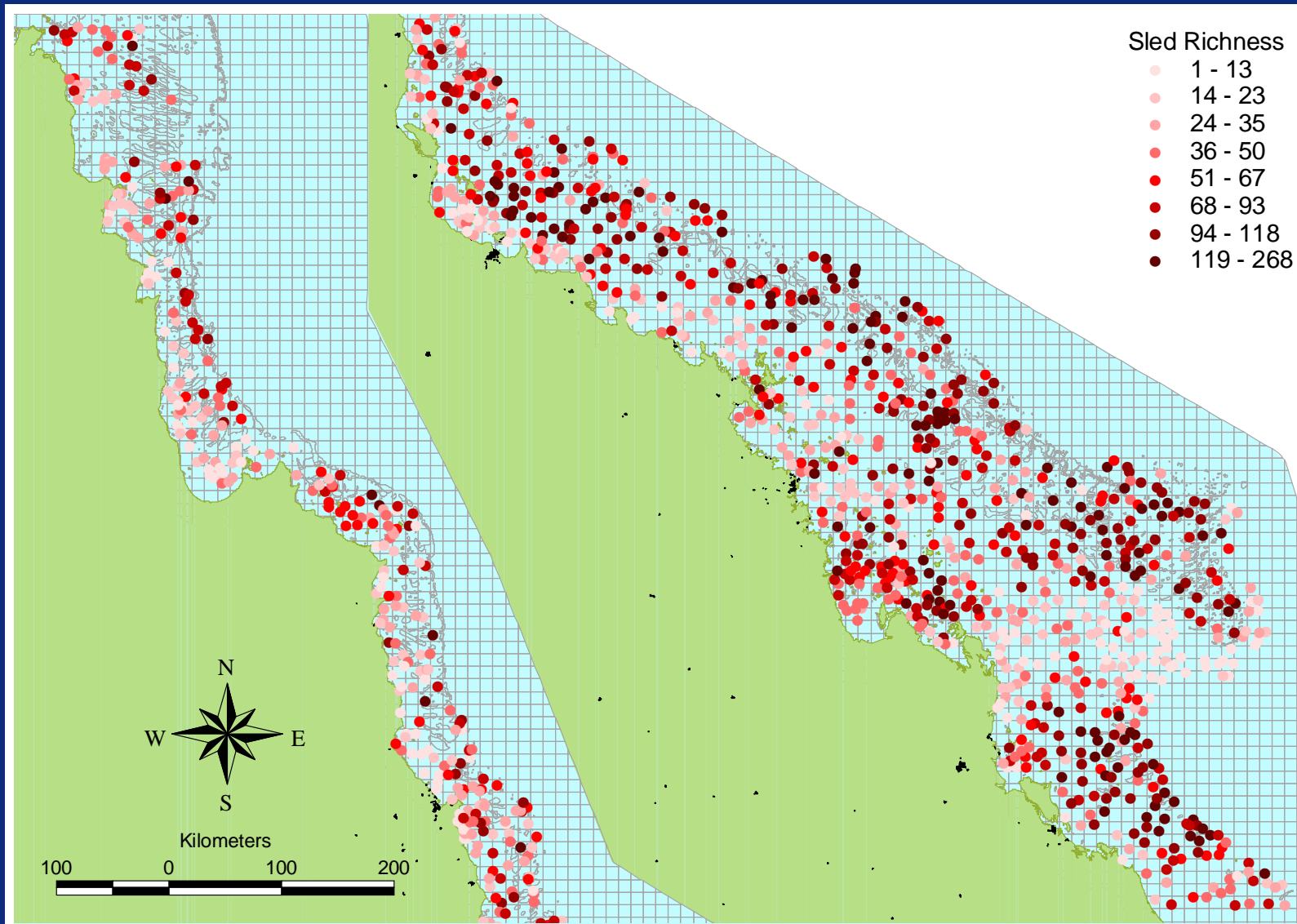


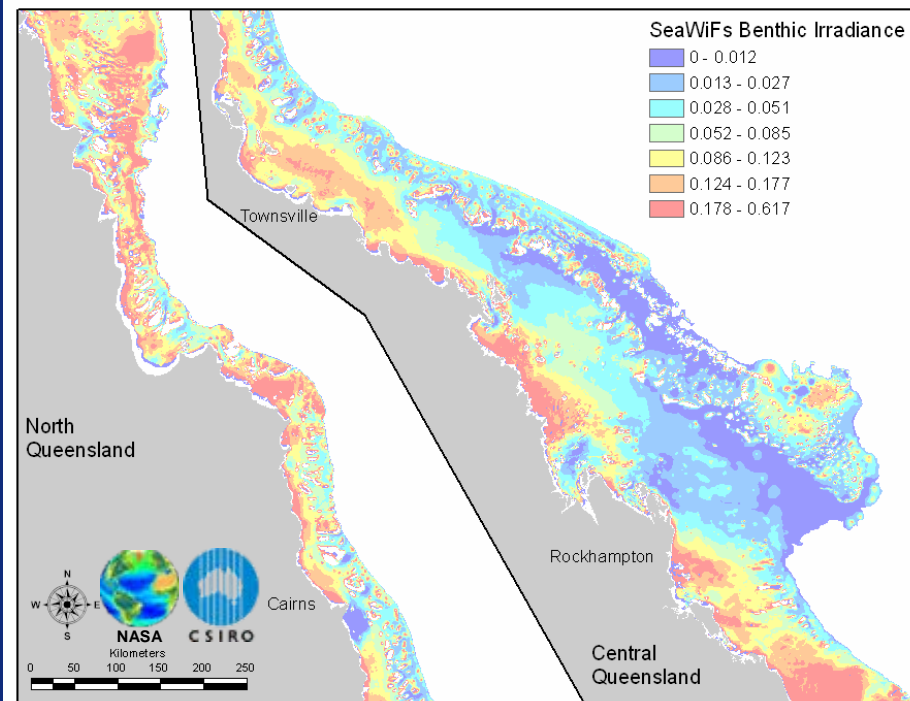
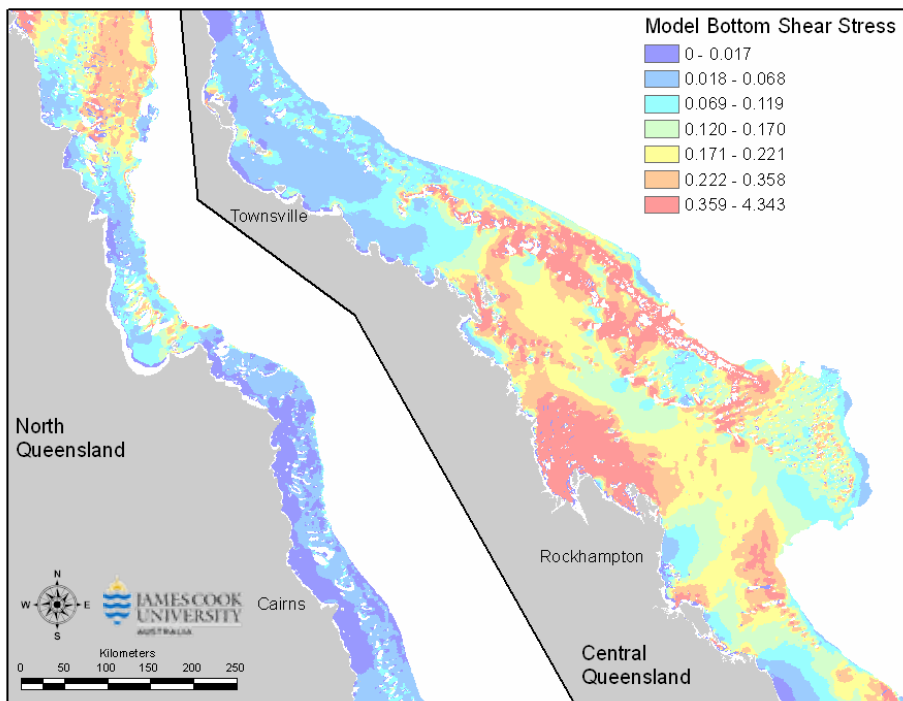
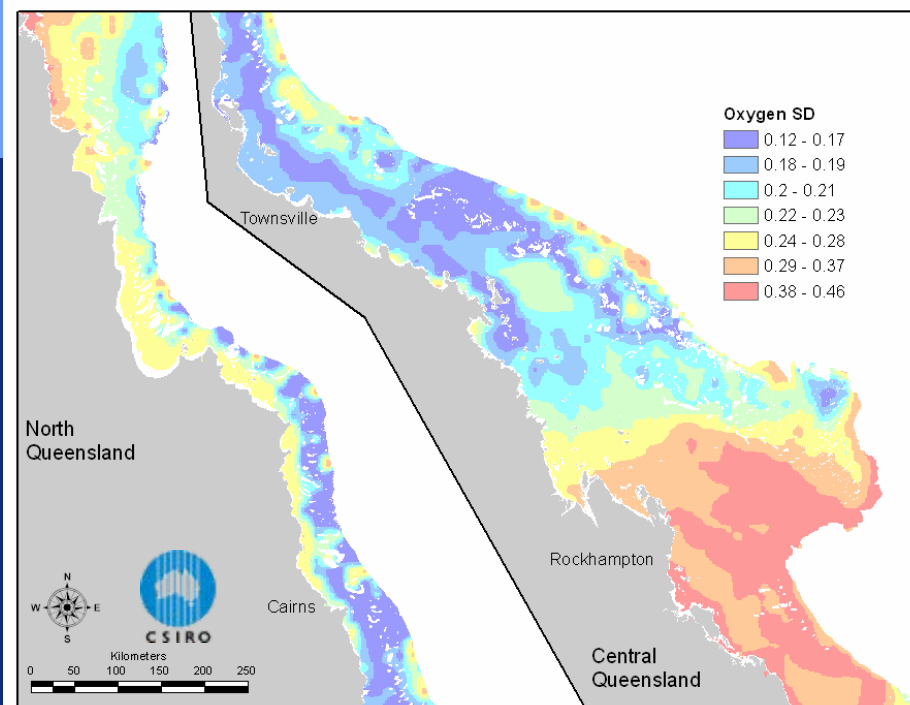
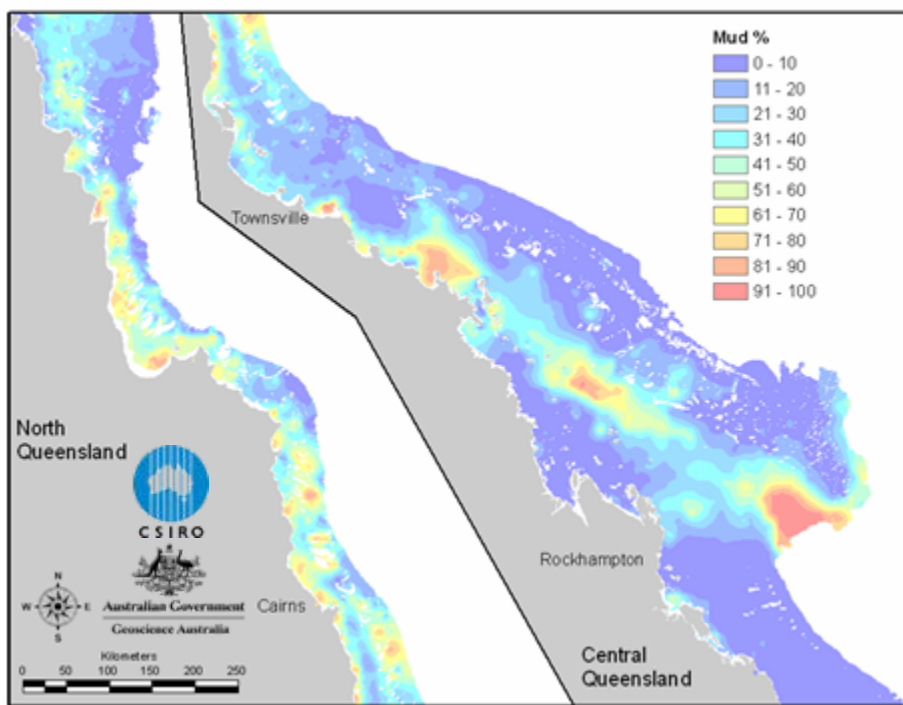
Australian Government

Department of the Environment
and Water Resources

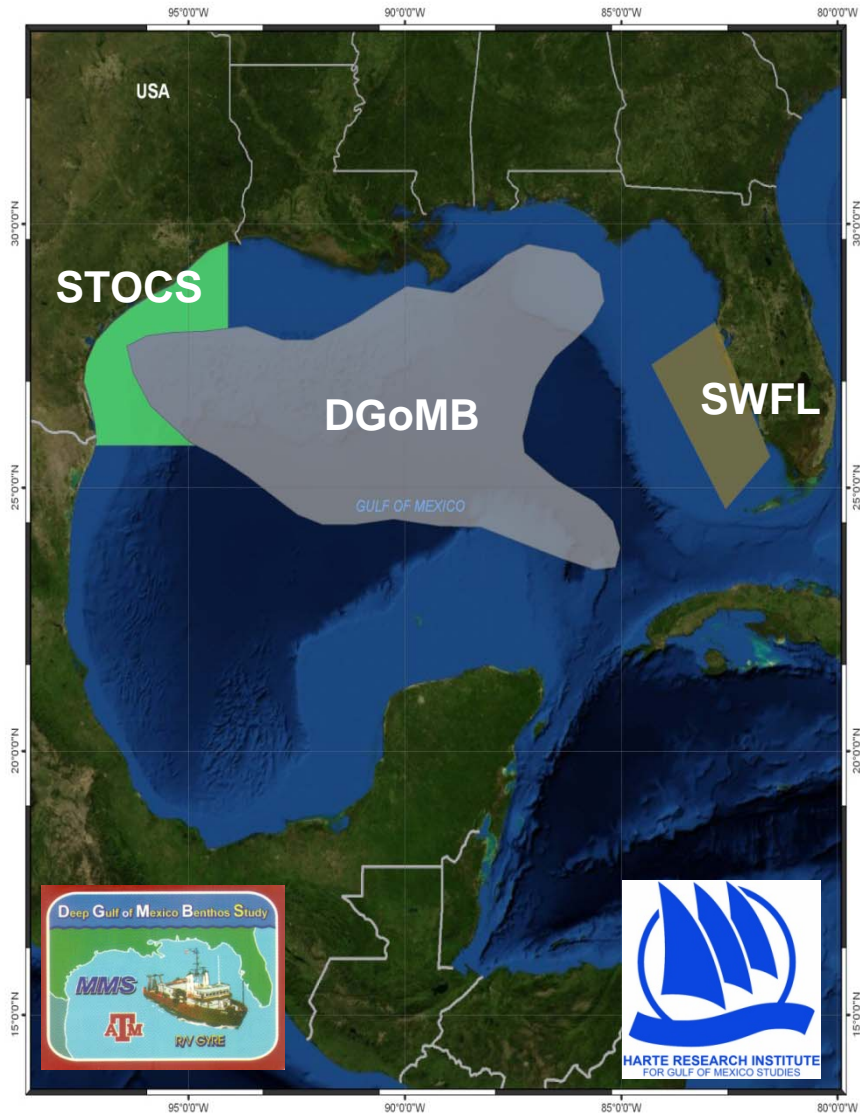


Scope: Great Barrier Reef





Scope: Gulf of Mexico



DGoMB - Deep Gulf of Mexico Benthos Program:

- 51 locations; 271 boxcores; 2079 taxa; 200-3800m; conducted 2000 - 2002

STOCS – South Texas Outer Continental Shelf Study:

- 1671 grabs; 799 taxa; < 200m; conducted 1975 - 1977

SWFL – Southwest Florida Shelf Study:

- 2975 grabs, 2274 taxa, < 200m; conducted 1980 - 1985,

STOCS and SWFL form two of four benthic surveys recently incorporated into OBIS (www.iobis.org) in 2007 (Carney, R.S. 2007 OCS Study MMS 2007-030)

Scope: Gulf of Maine



Theroux and Wigley 1998 . U.S. Dep. Commer.,NOAA Tech. Rep. NMFS 140:

-1,076 grab samples, 567 taxa; 3 to 3,974 m; mid-1950's to the mid-1960's; included S. New England

- Data for GoM selected, leaving 626 grab samples with genus or species- level information, sample biomass and sediment grain size

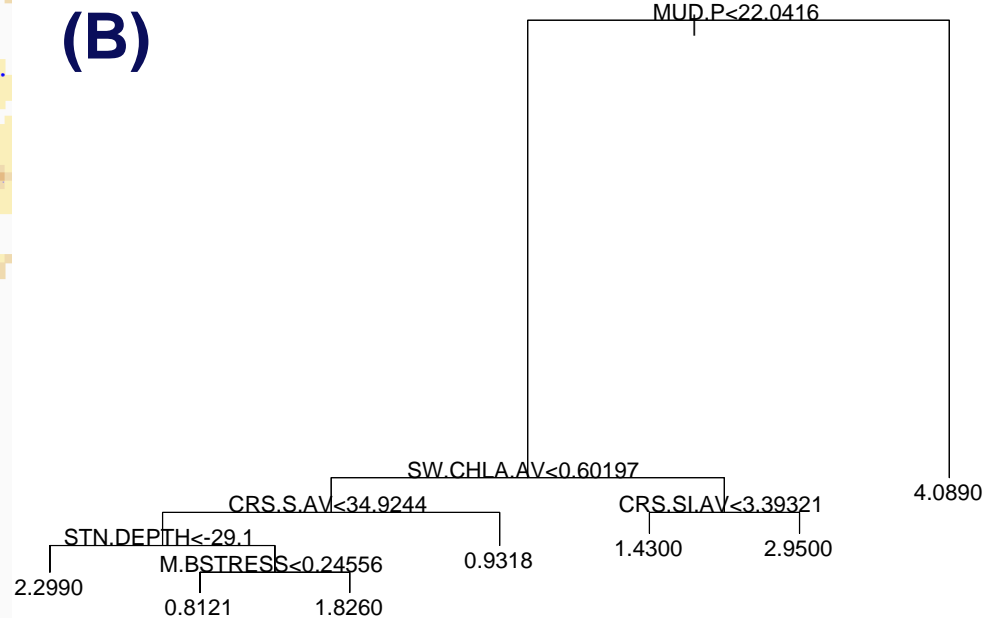
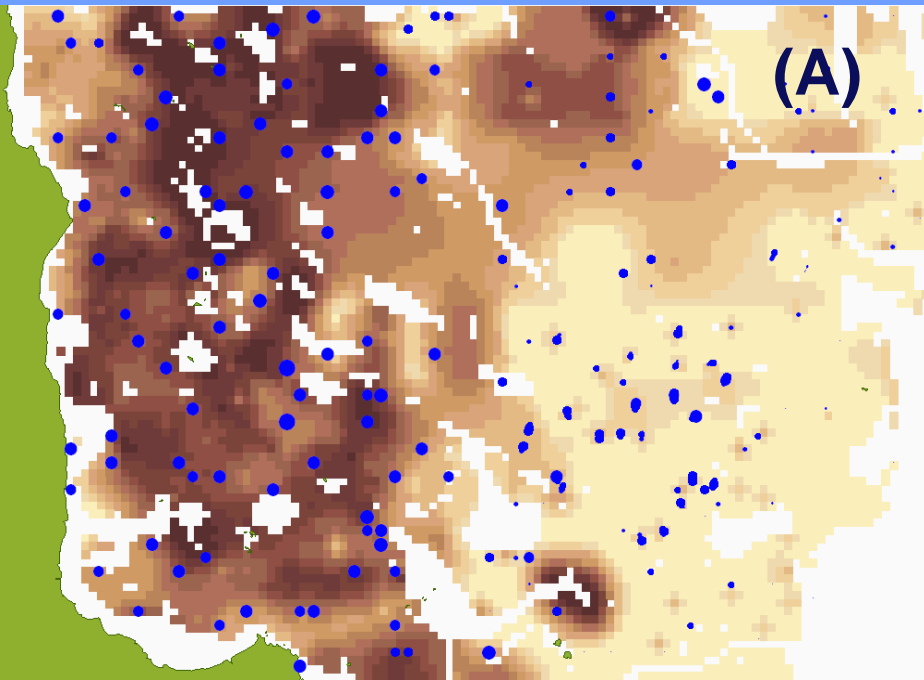
- Contained in benthic database provided by NOAA. Has been updated to capture accepted names in WoRMS and/or ITIS:
(<http://www.marinespecies.org/>)
(<http://www.itis.gov/>)

- 14 sets of physical variables are being compiled using Canadian and US sources

Selected Analysis Approach

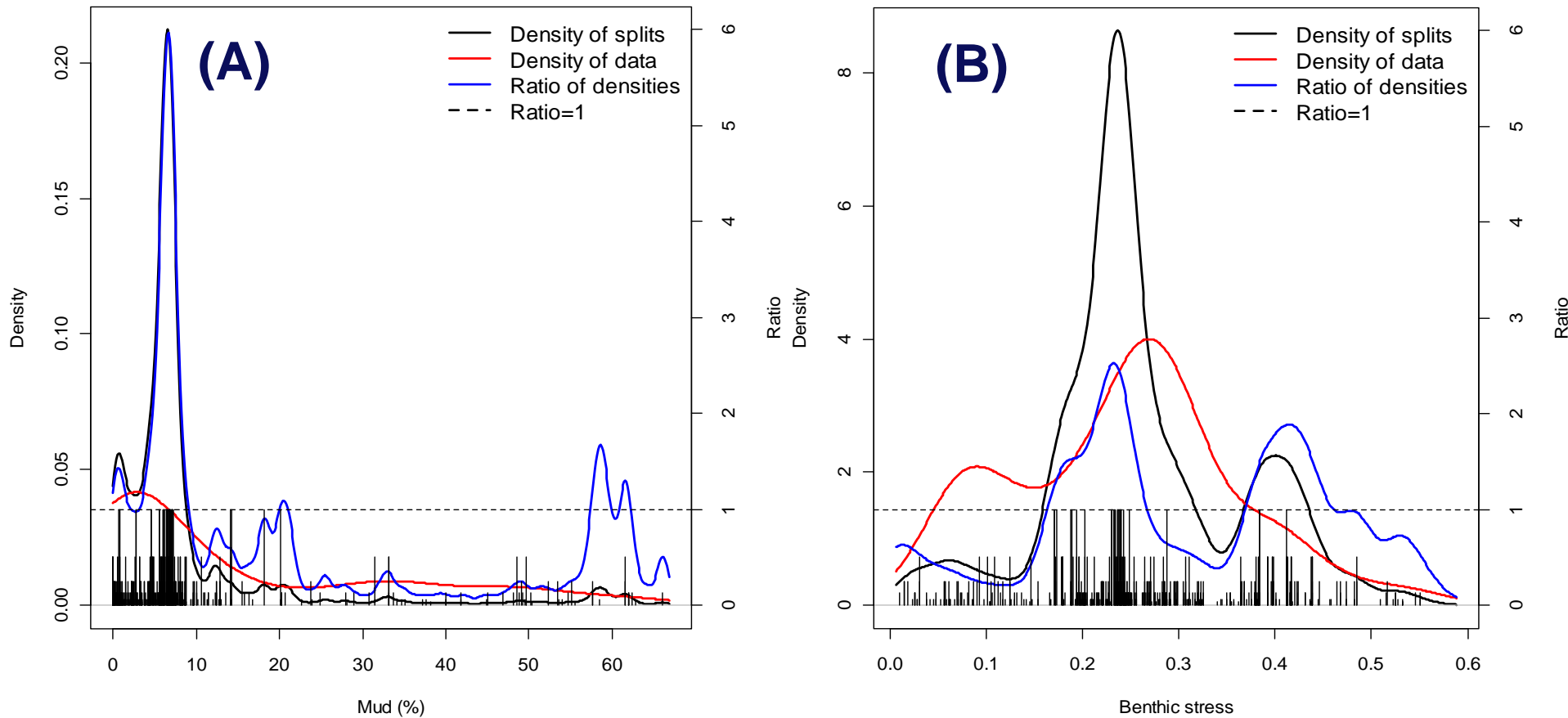
- Bootstrapped randomized regression tree techniques — random forests — to identify and rank variable importance of physical factors on observed biological assemblage patterns
- Technique development and existing application derive from prior Australian studies on Great Barrier Reef system and research currently underway in Australian CERF Marine Biodiversity Hub
- CoML synthesis project led Nick Ellis to modify the random forest methodology, by extracting additional information to enable comparison of the influence of physical variables across different sampling approaches and different geographical regions

Analysis Overview (1)



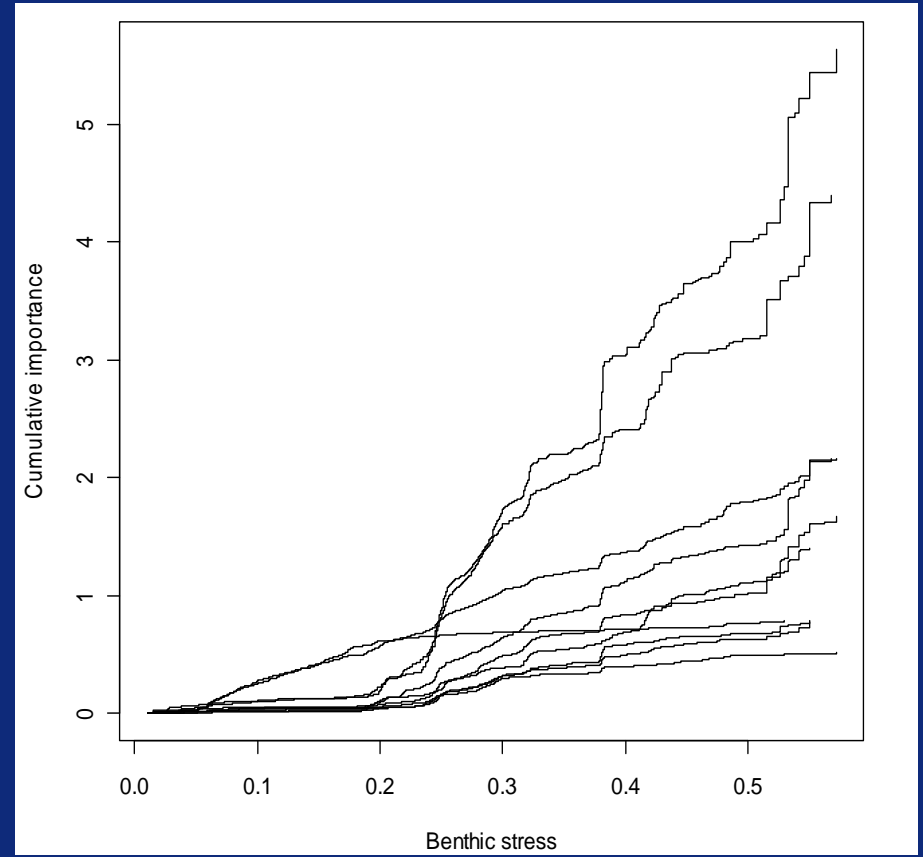
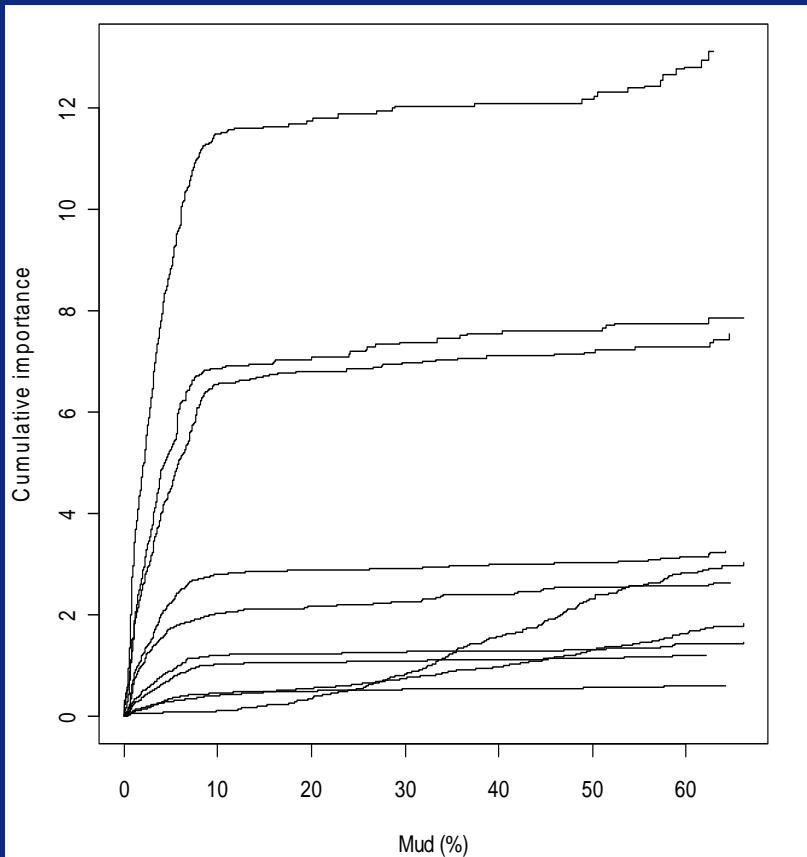
- (A) Biomass distribution for 1 species displayed on coverage of percentage mud
- (B) A single regression tree model for same species against set of physical variables (build randomized forest of ~500 of these)
- Examples from Effects of Trawling dataset from far northern GBR

Analysis Overview (2)



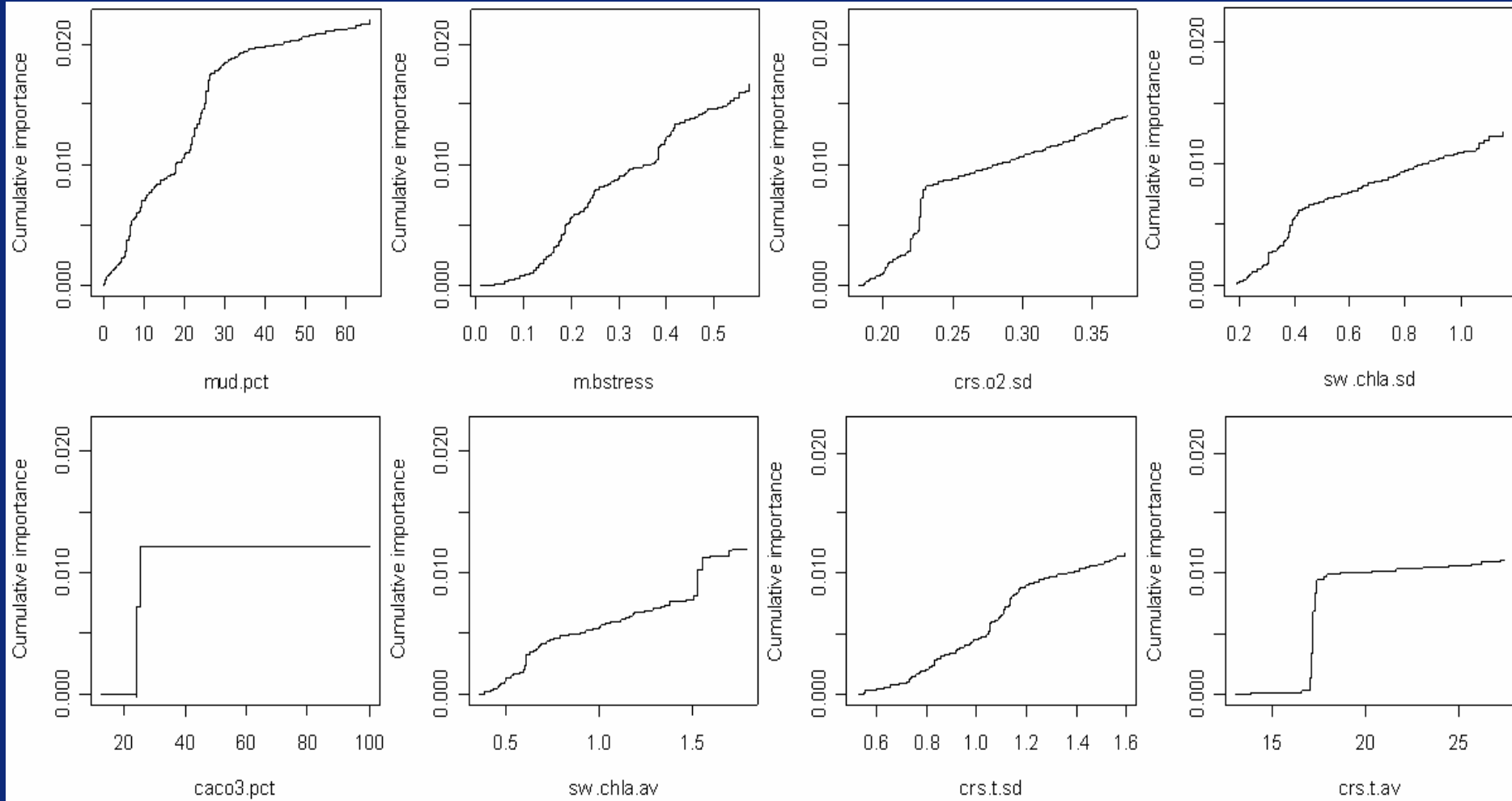
“Importance” weighted split points and data frequency for two physical variables (A: mud; B: benthic stress) for 10 most common species

Analysis Overview (3)



Cumulative distribution of weighted splits for same variables (A: mud; B: benthic stress) for 10 species

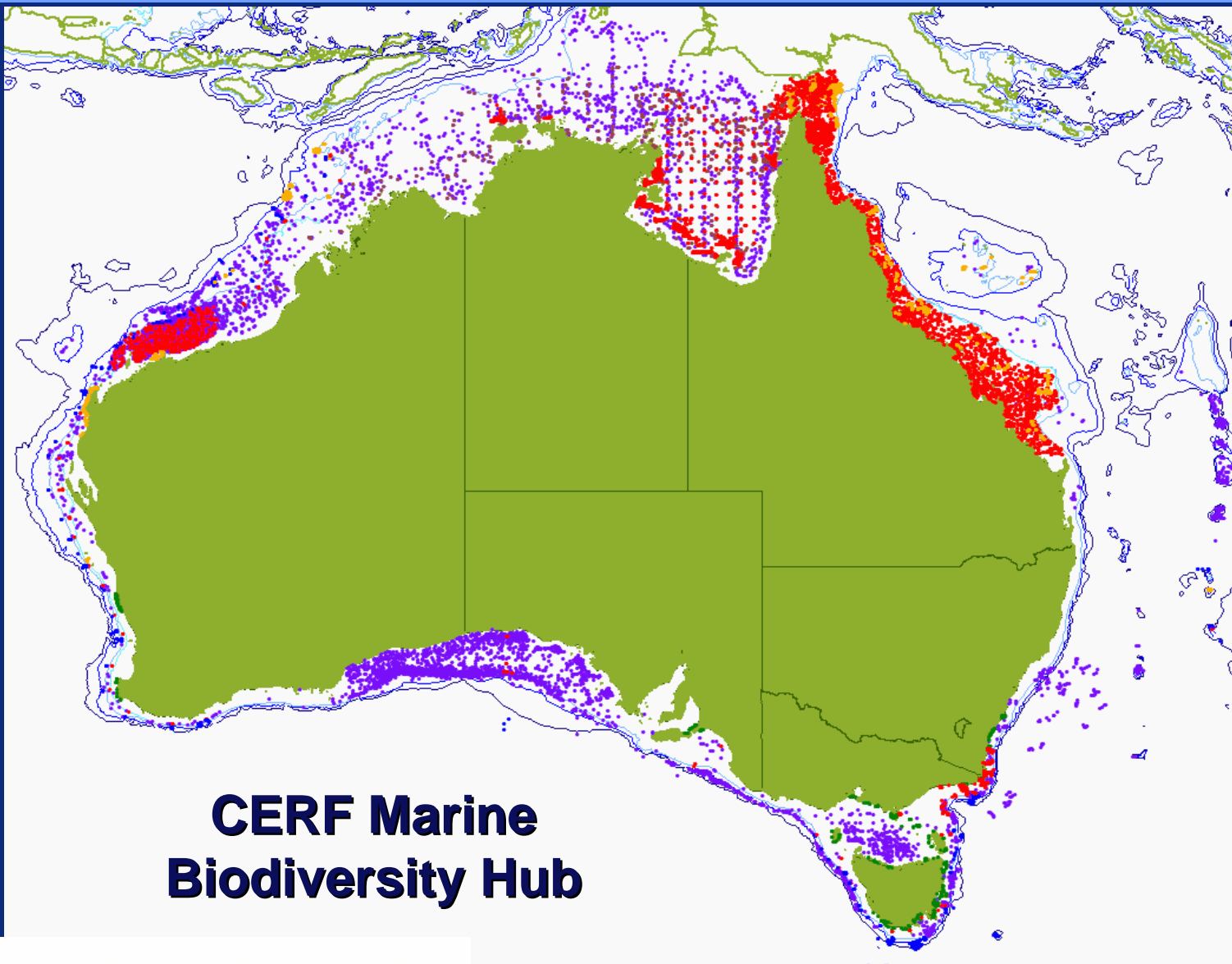
Analysis Overview (4)



Pattern of biological change-response on the top 8 physical gradients for 93 species, scaled by aggregate variable importance

Can use these outputs to transform physical data to ecological gradients to characterise (& map) a region directly from physical surrogates...

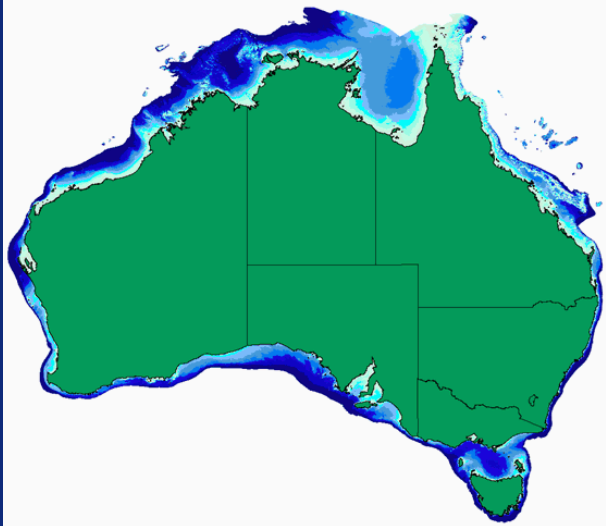
Current Australian Applications



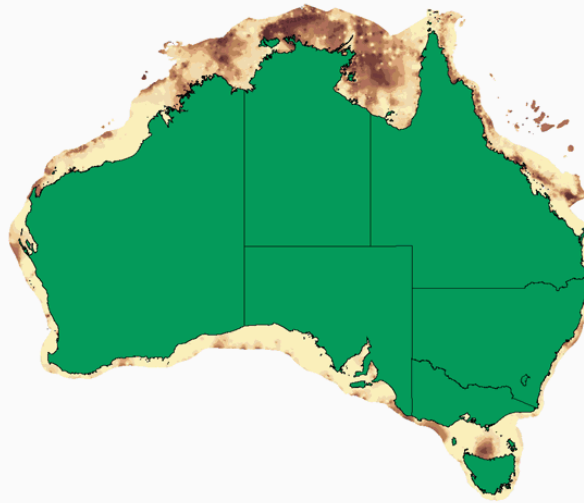
**CERF Marine
Biodiversity Hub**



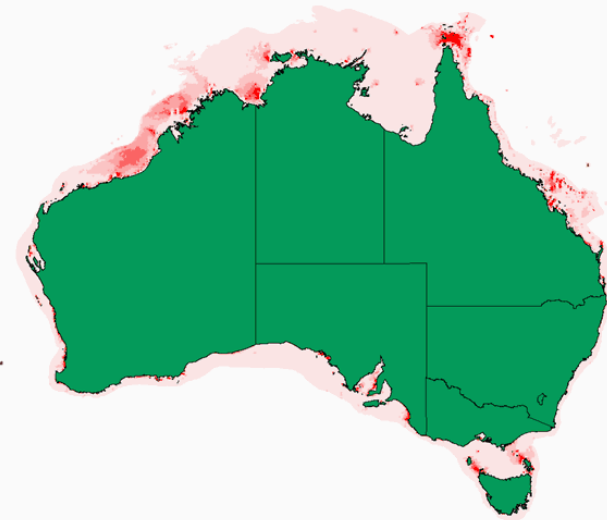
Some “Influential” Physical Variables



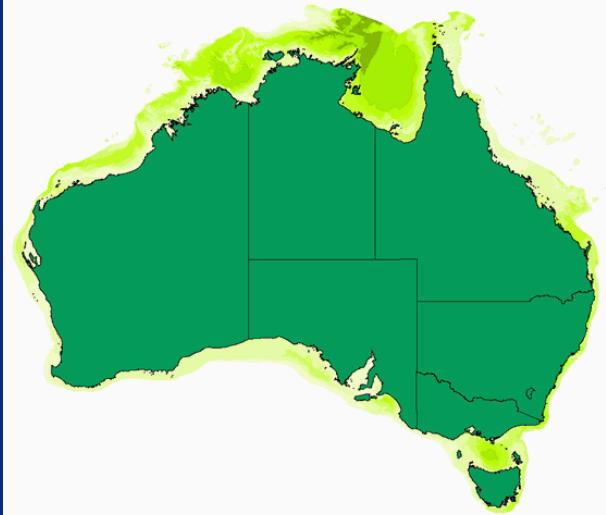
Bathymetry



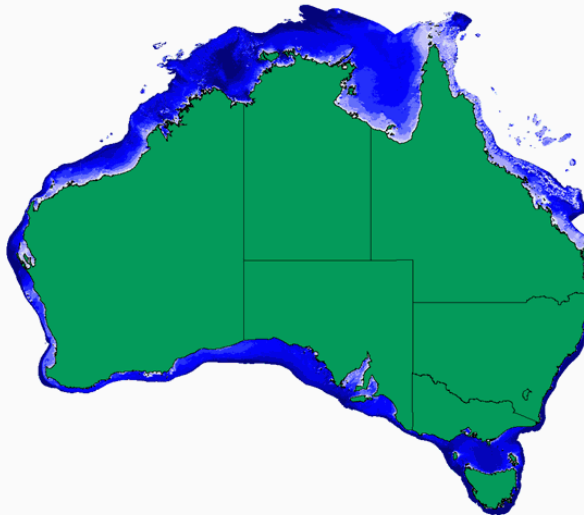
Sediment % mud fraction



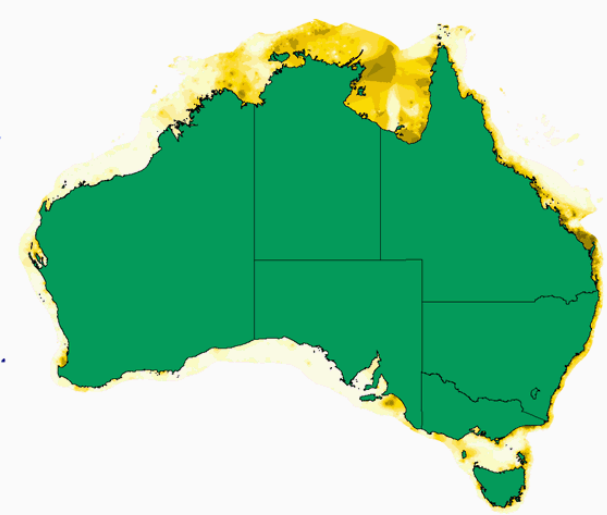
Seabed current stress



Bottom water Oxygen std dev



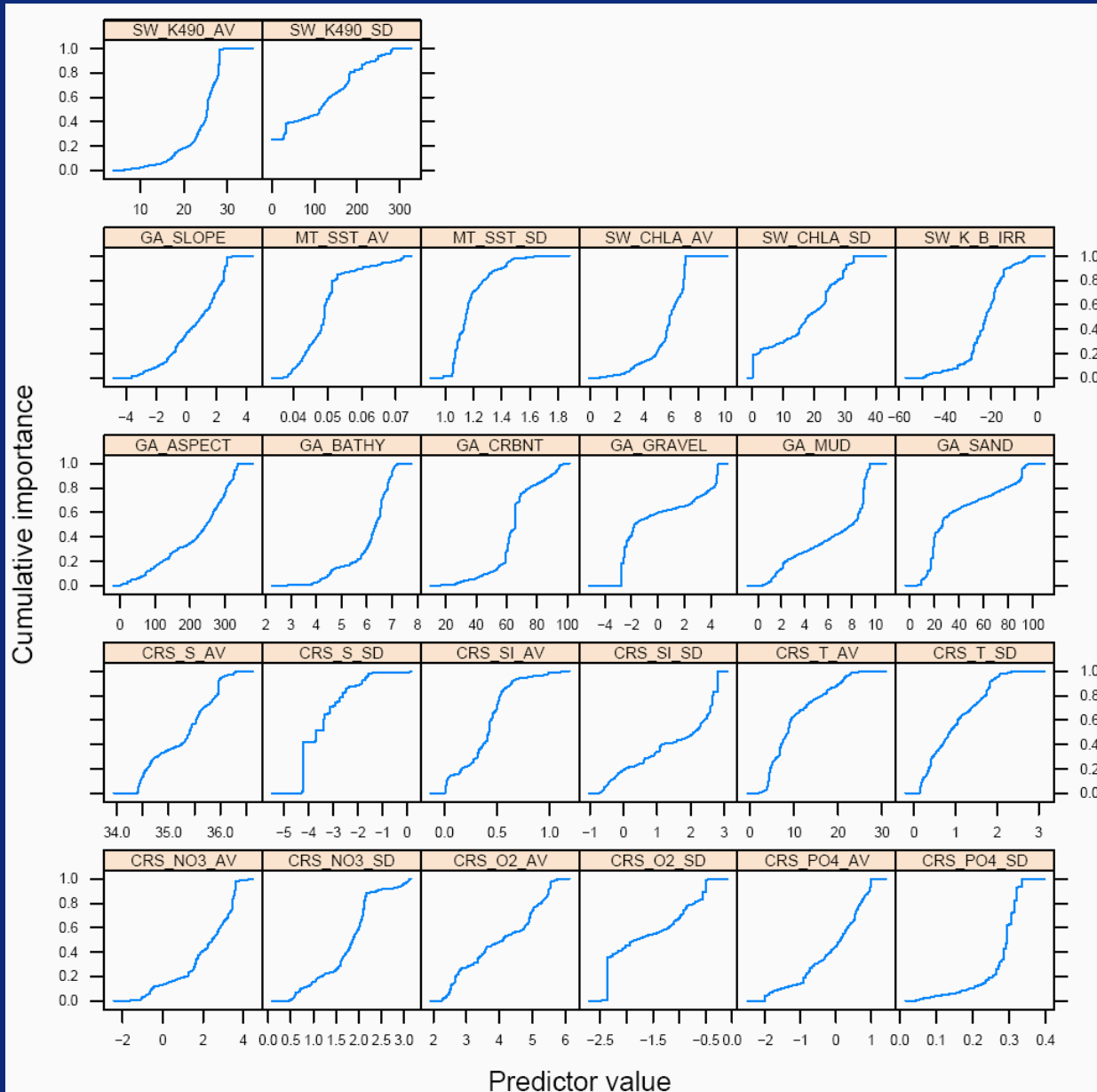
Benthic Irradiance



Sediment carbonate composition

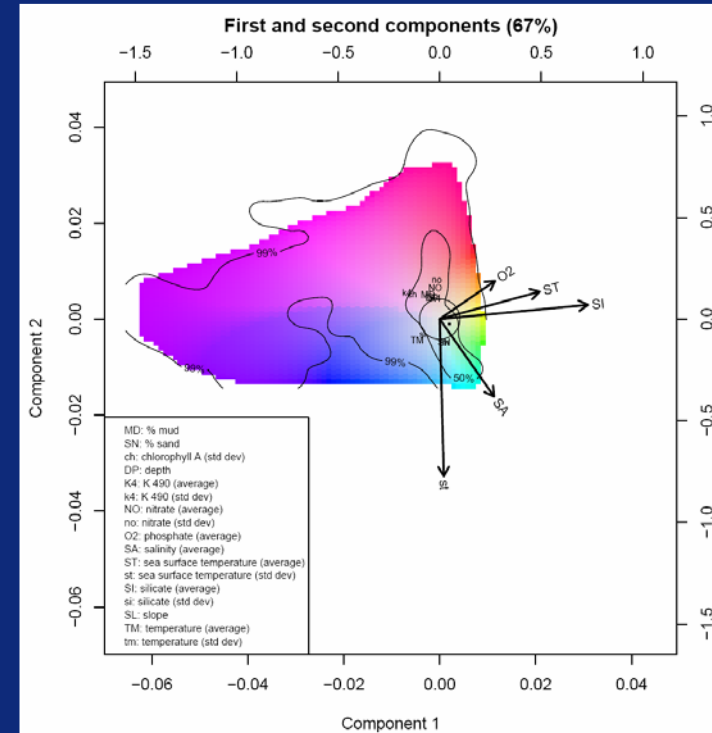
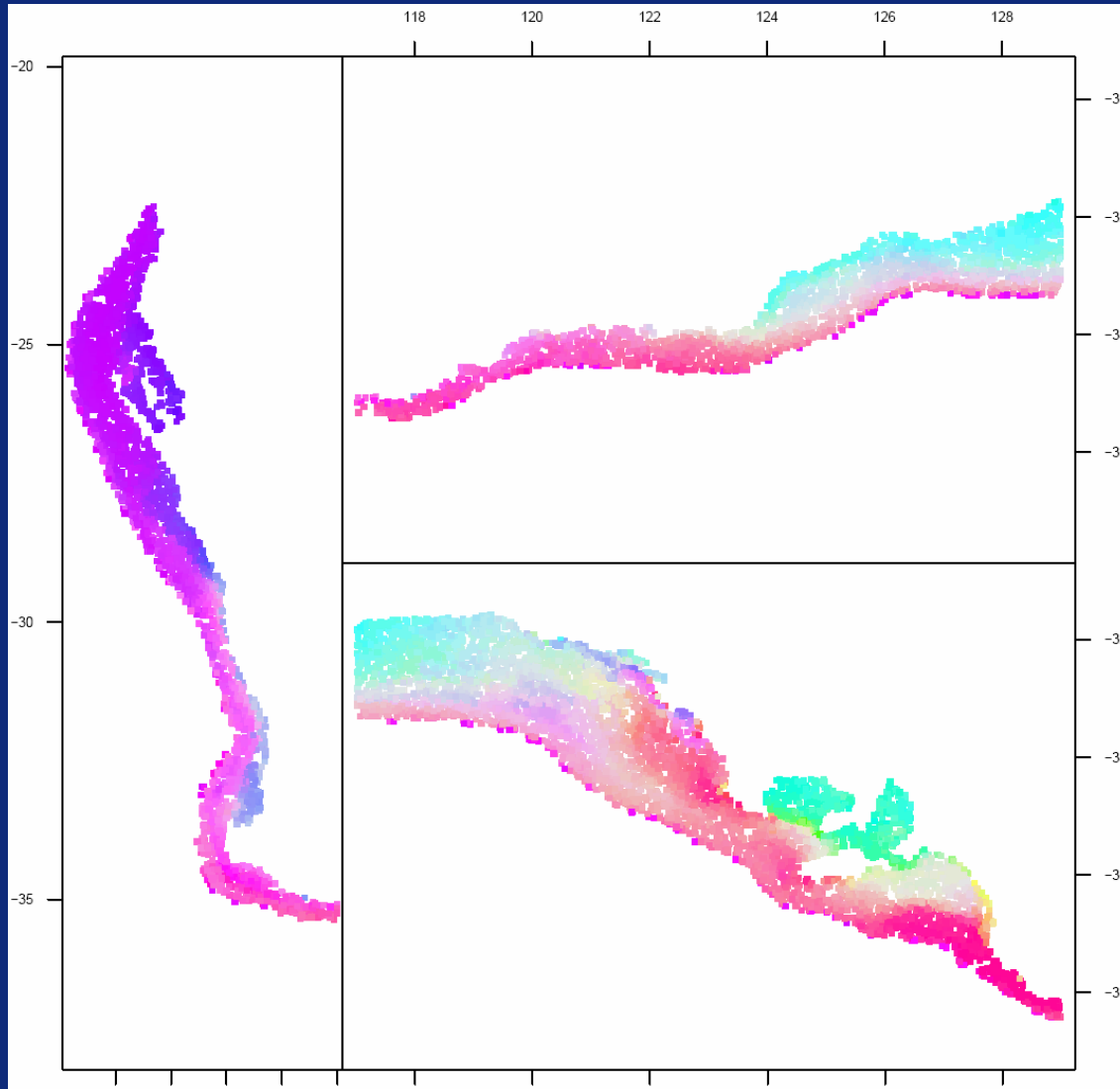
Maps of selected physical variables on Australia's continental shelf that have been associated with patterns in seabed biodiversity data.

“Biologically Informed” Transformations



Predictor	Importance
CRS_O2_AV	7.40
CRS_SI_AV	6.15
CRS_SI_SD	5.02
CRS_S_AV	4.97
CRS_T_AV	4.90
GA_BATHY	4.36
CRS_NO3_AV	4.29
GA_SAND	4.01
MT_SST_AV	3.83
CRS_NO3_SD	3.79
CRS_PO4_AV	3.74
MT_SST_SD	3.59
SW_K490_SD	3.58
GA_MUD	3.42
SW_K490_AV	3.37
SW_CHLA_AV	3.33
CRS_O2_SD	3.32
CRS_T_SD	3.27
SW_CHLA_SD	3.09
GA_SLOPE	2.82
CRS_S_SD	2.77
GA_GRAVEL	2.76
SW_K_B_IRR	2.66
GA_CRBNT	2.50
GA_ASPECT	2.20
CRS_PO4_SD	2.17

“Biologically transformed Environment Space”

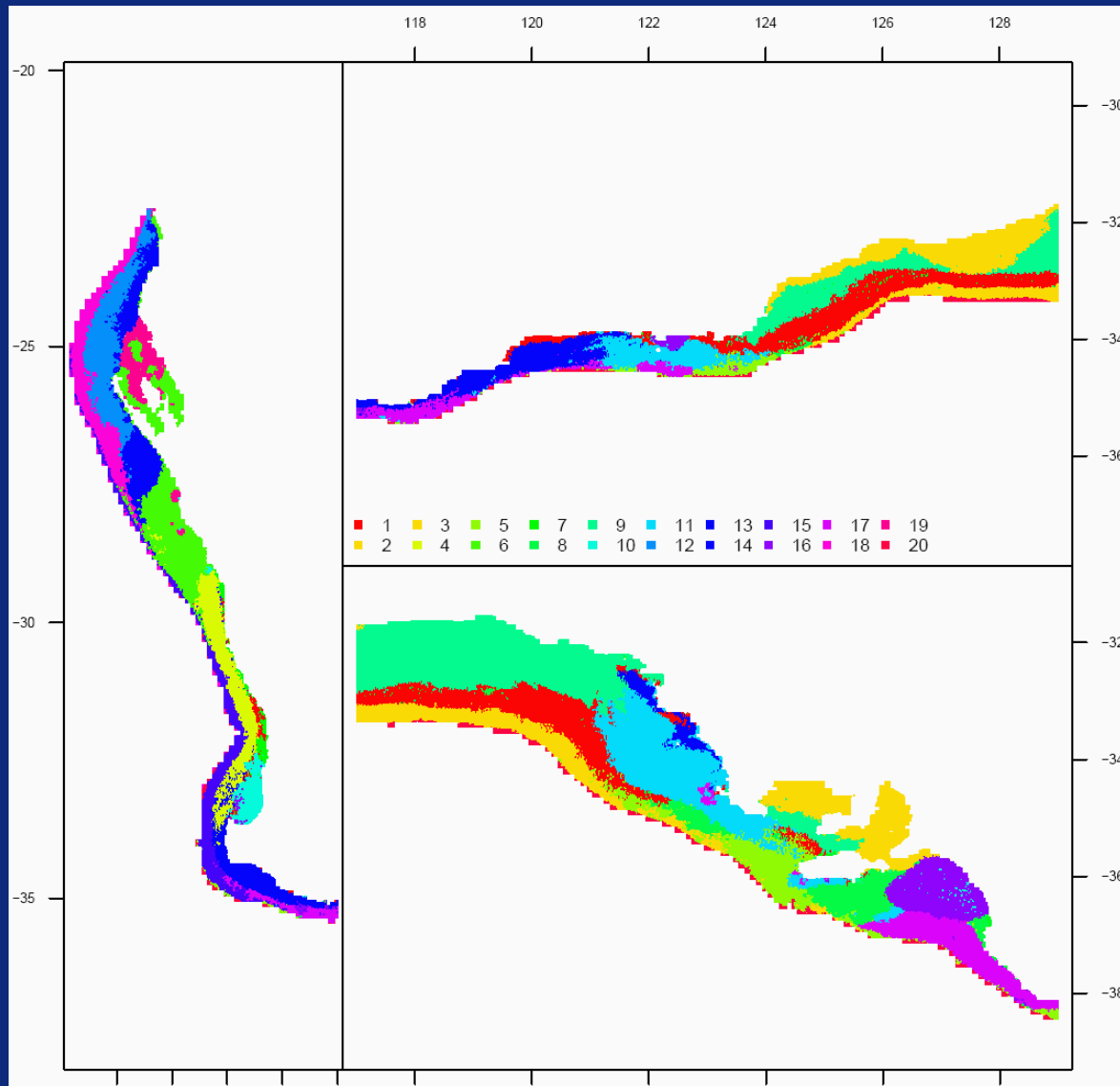


1st and 2nd PCs with biplot vectors

for major driving factors
(Preliminary analyses reported)

Continuous variation in physical environment from perspective of available biological survey data.
“Surrogate” map of expected patterns in seabed biodiversity for the SW Australia region
(primarily trawled fish in this example)

A Conservation Planning Application



Some applications (such as use of Marxan software for MPA planning) require categorical products.

These can be derived from the BTES

This example is a “Clara” mediod clustering (R package) modified to give relatively more weight to rarer combinations of important physical variables.

Somewhat arbitrarily in this example, 20 groups have been chosen

Interim product, April 2009

Summary

- CoML's Synthesis Program provided an opportunity to review approaches to evaluate the importance of physical surrogates for prediction of patterns in seabed biological assemblages
- Initial workshop led to a decision to apply Random Forest approaches (modified by Australian researchers) to other biogeographic systems that have been the focus of CoML programs
- Has helped to spur further development of these statistical techniques through ways to compare and visualize responses to physical variables across different sampling programs (Nick Ellis)
- This expanding set of techniques should be applicable to a range of marine ecological and marine conservation questions.
- Given the schedule for the CoML synthesis program, additional datasets are unlikely to be able to be accommodated, however further collaborations are welcomed