NATIONAL OCEANS OFFICE

National Work Program (Torres Strait)

Torres Strait Seabed & Water-Column Data Collation, Bio-physical Modeling and Characterization

Project Number: NOOC2003/010

Roland Pitcher, Scott Condie, Nick Ellis, Ian McLeod, Mick Haywood, Scott Gordon, Tim Skewes, Jeff Dunn, Darren Dennis, Liz Cotterell, Malcolm Austin, Bill Venables, Tom Taranto

CSIRO
MARINE RESEARCH

Final Report

Published August 2004

National Library of Australia Cataloguing-in-Publication data:

Torres Strait Seabed & Water-Column Data Collation, Bio-physical Modeling and Characterization

Bibliography.
ISBN 0 XXX XXXXX X.

1. Subject keywords A……
2. Subject keywords B……
I. Roland Pitcher. II. CSIRO Marine
Research. III. National Oceans Office
(Australia).

xxx.xxxxxxxx

August 2004

Torres Strait Seabed & Water-Column Data Collation, Bio-physical Modeling and Characterization

Project Number: NOOC2003/010

Roland Pitcher, Scott Condie, Nick Ellis, Ian McLeod, Mick Haywood, Scott Gordon, Tim Skewes, Jeff Dunn, Darren Dennis, Liz Cotterell, Malcolm Austin, Bill Venables, Tom Taranto,

CSIRO Marine Research
Castray Esplanade
(GPO Box 1538)
HOBART, Tas. 7001
Australia

CSIRO
MARINE RESEARCH

ISBN 0 XXX XXXXX X.

NATIONAL OCEANS OFFICE National Work Program (Torres Strait) Final Report

ACKNOWLEDGEMENTS

This Project was supported by funding from the National Oceans Office and CSIRO Marine Research. Data for the Project was provided by CSIRO Marine Research, GeoScience Australia, Ocean Sciences Institute (Sydney University), James Cook University / CRC-Reef, and SeaWiFS.

TABLE OF CONTENTS

FIGURES

TABLES

EXECUTIVE SUMMARY

The main aims of this project were to compile and assess the current state of knowledge of the broad-scale seabed and water-column ecosystem of Torres Strait, provide a preliminary characterisation of the region for use by the National Oceans Office (NOO), and design a sampling strategy for the Ecosystem Mapping Task of the Cooperative Research Centre for the Torres Strait. The project successfully collated and assessed relevant seabed & water-column data-sets, examined relationships between the biological and physical data, developed a bio-physical stratification of the seabed for the Torres Strait region, and completed a sampling design for the mapping project in time for the first field survey in January 2003. The major beneficiaries of the information include the National Oceans Office (NOO) and the Torres Strait people, and the Australian Fisheries Management Authority (AFMA) and the Torres Strait fishing industries. Funding was provided by the NOO and CSIRO.

Significant information on the physical environment was available from existing data. The Project collated 17 major datasets of physical and biological data for the region. Available relevant information included: physical environment (bathymetry, sediment grain-size and composition, water attributes & chemistry, ocean colour); basic seabed habitats; seagrass and algae; and some trawl samples. After checking quality and redundancy among sources, 32 physical variables were identified and mapped as potentially useful for modelling & stratification. A 0.01 degree resolution (~1.1 km) grid was established for analyses and sampling design, and the physical variables were re-sampled to this grid and mapped (interpolated where required), to provide a consistent set of full-coverage covariates at ~45,000 grid cells for the Project. The biological data was sourced from multiple legacy projects, each with different objectives, and was reconciled to useable common-denominator formats.

The broad-scale physical factors important in structuring patterns in the biological data were identified. Seabed current stress was the most important variable, and others included: chlorophyll, turbidity, oxygen, salinity, nutrients, sediment grain size, and depth. These bio-physical relationships were used to predict and map the categorical biological data to the whole Torres Strait region, with an estimate of the uncertainty. The Torres Strait region was characterised by weighting each physical covariate by its biological importance, then grouping the 0.01° grid cells into strata that had similar physical attributes. The stratification was mapped and represents an interim surrogate characterisation of Torres Strait. Sites for future sampling by the CRC-TS Mapping Task were selected from the bio-physical strata to provide representative coverage of the Torres Strait environment.

The Project described the current state of knowledge of the physical marine environment and the seabed habitats and biota, identified the major knowledge gaps and the key information needs for regional marine planning and ecosystem-based management. These are summarized here.

Torres Strait is a shallow area of continental shelf with complex topography comprising numerous reefs and islands; the eastern area includes deeper water but is more complex; the northwestern area is very shallow limiting navigation. For navigation reasons, the bathymetry of the main shipping channels is well known, but much of the region has not been surveyed and is poorly known. Consistent coverage of bathymetric data over Torres Strait, at resolution sufficient for navigation purposes, is required for reliable circulation modelling and bio-physical mapping.

Tides and currents dominate the physical oceanography of Torres Strait, with strong tidal currents in channels between reefs. Quantitative knowledge of tides and currents is largely from the output of models, as there are few tidal and current monitoring stations. The tides and currents in most of Torres

Strait have not been measured, leading to model uncertainty and lack of knowledge of the broader circulation, dispersion and connectivity. Additional tidal and current monitoring stations are needed at key locations and periods across Torres Strait to provide data to validate circulation models.

Limited knowledge of basic hydrographic conditions indicates that: water temperature peaks broadly over the summer and during winter, shallow areas are cooler; salinities fall during the monsoon season, when a low salinity feature occurs along the PNG coastline, and increase again during the dry season; strong tidal mixing generally prevents vertical stratification and oxygen levels are relatively high in the well-mixed water; nutrients may increase during the monsoon and decrease during the trade wind season, consistent with riverine inputs; and limited time-series data indicate inter-annual variability in the monsoonal and trade-wind influences. The coverage of basic hydrographic data is extremely sparse, both spatially and temporally, and there is a critical need for additional moorings to be deployed at key locations, to provide knowledge of interannual variability and environmental change and to develop any understanding of productivity processes in Torres Strait.

Sediments become suspended by the strong spring tide cycles and wind stress, particularly in shallow areas and near rivers, causing local areas of high turbidity. Turbidity has been estimated from the SeaWiFS satellite data but is confounded in shallow areas like Torres Strait. Very few direct measurements have been made and knowledge of their tidal and seasonal patterns is inadequate. Measurements of suspended sediments are required at key locations and spanning spring-neap tidal and seasonal cycles.

There is little knowledge of biogeochemical cycles in Torres Strait, or of the role of suspended sediments in those processes, and studies *de novo* are required.

The phytoplankton in Torres Strait is not known from any direct measurements and estimates based on satellite ocean colour are confounded by turbidity and shallow water. Nevertheless, indications are that chlorophyll levels are higher during the monsoon season and decrease towards winter. The primary productivity processes and plankton community structure in Torres Strait are unknown, and again studies *de novo* are required.

The seabed sediments of Torres Strait cover the full range from fine terrestrial muds near rivers to coarse carbonate sands and gravels among coral reefs further from land. Over that pattern, the strong tidal currents scour fine sediments from narrow channels, leaving coarse gravels and rocks, and deposit them in calmer areas. The currents also create and move dunes of sand. However, the sediment grain size attributes of most of Torres Strait have been sampled only patchily and there are extensive gaps in east/southeastern and northwestern Torres Strait. Sediment types typically showed little similarity over distances of more than 5 km, perhaps 10 km maximum, providing a criterion (spatial autocorrelation) for assessing that the existing coverages were inadequate for significant parts of the region. Adequate fine-scale sampling of sediment grain size and composition is required for understanding sediment processes and biogeochemical cycles, and as surrogates for biological assemblage prediction. Sediment organic content is almost unknown, but would also contribute to these knowledge needs. Acoustic data from several vessel tracks was shown to be a useful surrogate for seabed substratum and can provide continuous along-track coverage between actual sediment samples.

Prawn trawling is largely confined to a relatively narrow strip in central eastern Torres Strait and extremely intense effort was aggregated into an area of about 200 km². The trawl logbook data coverage is quite complete; however, the resolution is coarse compared with actual trawling activities. Fine-scale trawl effort data needs to be acquired from Vessel Monitoring Systems installed on Torres Strait trawlers to provide effort data at a resolution needed to assess and manage the environmental sustainability of trawling.

The basic seabed habitats of Torres Strait (substratum type, megabenthos gardens, presence of algae and seagrass) are relatively well known compared with many other areas of seabed within Australia. The region has significant large areas of structured habitat and benthic gardens, as well as extensive seagrass beds and algae. From this knowledge, it has been observed that seabed current stress appears to be very important in structuring habitat patterns, as are sediment attributes. Moving sediments may smother benthic habitat and expose bare substratum for colonisation. Sediment dynamics may also be involved in seagrass diebacks in north west Torres Strait, though the causes are unknown.

The broad habitat characterisation is known for about two-thirds of the region, at a resolution comparable with the spatial autocorrelation distance for habitat similarity (ie. ~10 km). Unknown areas include extensive areas of north-eastern and western Torres Strait. While, this broad information has been useful, it is inadequate to properly characterize biodiversity assemblage, to develop bio-physical models and for quantitative management applications. For these purposes, species biomass data are required but are largely unavailable in Torres Strait. Where more detailed biological information is available for some biota, the spatial coverage is very limited. Thus, a broad spectrum of seabed species need to be sampled in Torres Strait, accurately identified, quantified and mapped. Careful species identification is essential because Torres Strait is a biogeographic boundary due to past periodic separation of east & west faunas — an important concern for regional marine planning.

There is also very little knowledge of assemblage dynamics or of ecosystem processes in Torres Strait. Primary productivity, whether benthic or planktonic, has not been studied in Torres Strait nor have secondary productivity and higher trophic relationships, other inter-species inter-actions, or coupling between the benthic and pelagic ecosystems. This kind of ecosystem-level knowledge is required in order to progress towards ecosystem-based management of multiple uses of the Torres Strait marine environment, and needs to be synthesized by dynamic modeling approaches such as Management Strategy Evaluation.

The Torres Strait CRC program will address a number of these data issues at varying levels of detail. Field measurements to be made by the “Bio-Physical” and “Seabed Mapping” Tasks will provide additional hydrographic data and develop hydrodynamic modelling. Data on seabed and suspended sediments will be collected and sediment transport will be modelled; the issue of seagrass dieback will also be examined. Bathymetric data will be collected by research vessels. Data on seabed habitats will be recorded and a broad range of seabed assemblage species will be sampled, identified, quantified and mapped. However, the mapping will not be able to cover the entire region, and the program does not currently include studies of biogeochemical cycles or of biological ecosystem processes.

This Project has provided an essential foundation for several CRC-TS Tasks that will address priority issues related to assessment of the effects of trawling, development of trawl sustainability risk indicators, seagrass dieback, bioregionalisation for marine planning, and multiple-use management. The preliminary characterisation of the Torres Strait region provided by the project will support the planning needs of NOO and other management agencies, in the interim before the results of TS-CRC Tasks become available.

1. INTRODUCTION

1.1. BACKGROUND

There has been an urgent need to complete preparatory work prior to the start of the Torres Strait CRC in July 2003, so that fieldwork for the Task Ecosystem Characterisation may proceed in the required timeframe. This project aimed to collate and assess relevant seabed & water-column data, examine bio-physical relationships, develop a stratification for the Torres Strait and to design a seabed & water column sampling strategy for the TS CRC ecosystem survey task. Other benefits include assessment of the current state of knowledge relevant to this task and provision of a preliminary characterisation of the region for use by the National Oceans Office (NOO). These fundamental datasets and analyses will support present and future research and monitoring needs, conservation planning, and management — ultimately contributing to the preservation of the unique values of this region.

Available relevant information included: physical environment (bathymetry, sediments, water attributes & chemistry, ocean colour); seabed habitats; seagrass; and limited trawl samples. The approach has been to collate & integrate seabed & water-column data from disparate sources to common useable formats, identify any broad-scale physical factors important in structuring patterns in the biological data, characterise and stratify the Torres Strait region based on extension of the bio-physical relationships to the whole region. A cost-effective & optimised sampling strategy needs to be designed to representatively sample the identified bio-physical strata. Information gaps need to be identified in relation to data type, bio-physical strata, prediction uncertainty – as well as spatial grid coverage. Outputs in the form of digital GIS layers will be provided to the NOO.

This Project has assessed the state of knowledge of seabed habitats, seagrasses, benthic biodiversity, and the water-column in the complex ecosystem of the Torres Strait; provided interim spatial characterisation information for management and planning needs until more complete information is available (including issues related to anthropogenic impact in seabed ecosystems eg. trawling); and design future research surveys to optimally & cost-effectively address gaps in the current knowledge. The broad-scale objectives of this Project have been met by specialists from multiple disciplines and experienced Torres Strait researchers, based on CSIRO Marine Research’s (CMR) significant Torres Strait data holdings and access to other relevant datasets. The approach is relevant to the national objectives of NOO as outlined by Australia's Oceans Policy.

1.2. NEED

This Project was required to provide an essential foundation for TS-CRC Tasks that will address issues related to seagrass dieback, assessment of the effects of trawling, development of trawl sustainability risk indicators, bioregionalisation for marine planning, and multiple-use management. These issues were identified as priorities at several client and stakeholder forums. The NOO and CMR identified the outputs of this Project as a pre-requisite for the CRC-TS Task Ecosystem Characterisation to proceed from July 2003. Another need to be provided by the Project is an interim characterisation of the Torres Strait region to support the planning processes of NOO and other management agencies, before the results of TS-CRC Tasks become available.

1.3. OBJECTIVES

- Collate available existing physical data relating to the seabed & water-column environment of the Torres Strait, eg. sediments, bathymetry, currents & stress, water physical & chemical attributes, ocean colour, trawl effort etc.

- Collate available existing biological data relating to the seabed environment of the Torres Strait, eg. habitat type, flora and fauna species distribution and abundance data.

- Conduct exploratory analysis of bio-physical relationships, spatial modelling and stratification based on available biological and physical data of the seabed & water-column from the Torres Strait.

- To the extent possible, characterise, describe & map patterns in seabed & water-column biological and physical attributes, including those potentially vulnerable to trawling and other seabed activities.

- Assess the current state of relevant knowledge of the seabed & water-column environment, habitats, and biological assemblages of the Torres Strait

- Identify key information needs, including areas of seabed & water-column where additional sampling / mapping / survey work is required and design a sampling strategy for that survey work (including PNG seabed in the Torres Strait);

- Provide a report and GIS information to the NOO and Torres Strait Reef CRC / Torres Strait Fisheries Scientific Advisory Committee to support the marine planning and research in the Torres Strait

2. METHODS & RESULTS

This Project characterised the major patterns in the seabed habitats & water-column of Torres Strait (Figure 2.0-1), at spatial scales relevant to regional conservation and management needs, and planning/design of future research surveys. The information included: seabed habitat distribution in inter-reef areas; water chemistry; and physical attributes that may drive patterns within the system.

The approach was to collate and integrate the available biological, habitat, physical and water-column data; analyse bio-physical relationships to identify important environmental variables; stratify the Torres Strait seabed based on these variables weighted by their biological importance; attempt spatial-prediction of seabed biological assemblage information to provide interim characterisation of the Torres Strait seabed and estimates of prediction uncertainty; design sampling for future seabed surveys to achieve representative inclusion of important biological components, major habitat strata, and areas of uncertainty.

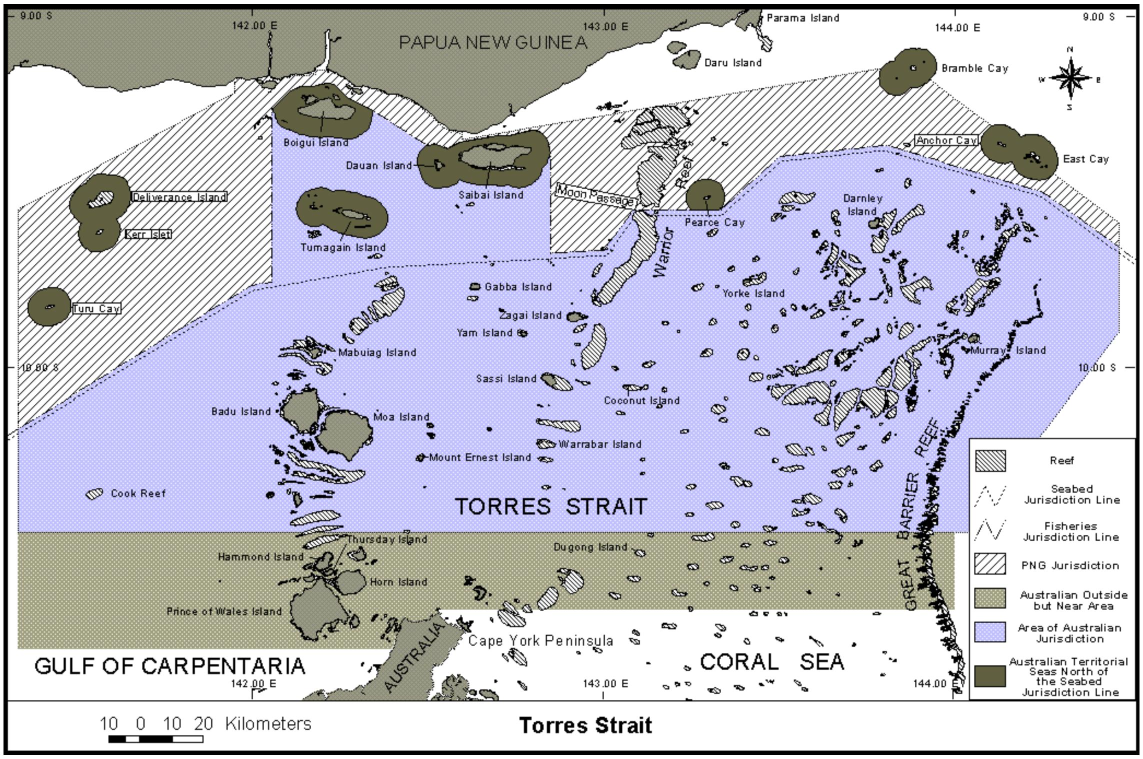


Figure 2.0-1 Map of Torres Strait showing the spatial scope of the Project, including all continental shelf marine seabeds and water column within the Torres Strait Protected Zone (corresponding to the scope of the CRC proposal) and relevant adjacent waters (extent indicated by map).

2.1. COLLATION OF AVAILABLE DATASETS

Available datasets of biological and physical data were collated from several internal & external sources. Some data was already held by CMR and simply needed to be mapped onto a 0.01º grid for analysis by this Project, some previously held data required updating before importing into GIS and mapping onto the grid, thirdly, some data needed to be sourced. These datasets included those outlined below.

2.1.1. Datasets Collated

- **Torres Strait Jurisdiction Zones**/Lines for Fisheries, Management, Protected Zone – these datasets were available from previous AFMA funded research in TS.
- **Bathymetry** – this dataset was updated from CMR (digital soundings & imagery) and GA sources, incorporated into GIS and modelled to produce a DEM and mapped onto a 0.01º grid. The coverage of the TS is extensive but not complete. Slope and aspect variables were derived.
- **Seabed sediment composition** – this dataset was updated from CMR and Ocean Sciences Institute (Chris Jenkins, formally OSI, Sydney University, auSeabed sediment database)

sources, incorporated into GIS and mapped onto a 0.01º grid. Coverage is the most extensive available in TS, but is not complete. The dataset included: characteristic grain size (phi), sorting, %mud, %sand, %gravel, %rock, %carbonate, with varying degrees of spatial reliability.

- **Torres Strait Prawn Trawl Fishery Logbook** – this dataset was provided by Industry–QDPI/QFS and has been updated post 1997 to cover the years 1989-2002. TS trawl effort (boat-days) data has been summarised annually at 6' (0.1º) resolution (~11.1 km).

- **Seabed current-stress** – this dataset was provided by Bode & Mason (JCU/Reef-CRC). The data are root mean square (RMS) stress (Pascals (N/m^2)) output from a circulation model run over period of approx 6 months. The modelled coverage is for the entire TS region, at 1 minute of arc resolution (~1.8 km), but is dependent on bathymetry data (which is incomplete) and other model assumptions.

- **CSIRO Atlas of Regional Seas (CARS)** – this dataset is an Australia wide database of water-column physical and chemical attributes maintained by CMR. The dataset included temporal series at fixed stations in Torres Strait and additional data has been collated to provide broader spatial coverage. All available measurements of water column properties were mapped at the the near-surface and the seabed to provide full-coverage of the TS region at 1/8 degree resolution by weighted averaging that takes into account bathymetry and seasonality. Water properties evaluated included:
  - temperature - degrees C, mean and standard deviations
  - salinity - psu, mean and standard deviations
  - oxygen - ml/l, mean and standard deviations
  - silicate - uM, mean and standard deviations
  - phosphate - uM, mean and standard deviations
  - nitrate – uM, mean and standard deviations

- **Ocean Colour** – this dataset includes estimates of mean and standard deviations for chlorophyll concentration and turbidity, processed by CMR based on SeaWiFS satellite data and calibration and validation algorithms. The SeaWiFS coverage has been updated with more than a year of additional data (more than 4 years in total). Relative benthic irradiance has been calculated, based on K490, latitude, and depth. The data provides full coverage of the TS region, with 0.01º resolution ~(1.11 km).
  - chlorophyll-a (mg/m3) concentration, mean and standard deviations
  - K490 diffuse attenuation coefficient at wavelength 490nm, m-1, mean and standard deviations

- **Seabed substratum types and living habitat** – this dataset includes physical substratum, epibenthos, seagrass, algae from several CSIRO diver and towed-video surveys of western, central and southeast TS over the period 1987-2002. While some surveys provided species abundance or percent cover detail, this level of information was not available with sufficient spatial coverage for analysis and a common set of lower level information was extracted. Broad scale coverage was provided primarily by 6 Projects/Surveys:
  - Western TS Seagrass Survey 1987
  - Lobster Abundance Survey 1989
  - South Eastern TS Pipeline Survey 1996
  - Central TS Pipeline Survey 1997

- PNG Lobster Abundance Survey 1998
- Lobster Benchmark Abundance Survey 2002

The following common set of lower level data was available for all surveys:
- SUB_CODE classification: 1= >50% rocky, 2= 10-50% rock, 3= rubble, 4= sand, 5= mud
- BIO_CODE classification:
  1= dense epibenthos, fauna separated by a few metres or less, covering >50% of the transect
  2= sparse, fauna separated by more than a few metres, covering 10-50% of the transect
  3= very sparse, patches separated by 10s-100s of metres, covering <10% of the transect
  4= no epibenthos, virtually no epibenthic fauna present
- PA_ALGAE: Presence/Absence Algae
- PA_SEAGRASS: Presence/Absence Seagrass

A slightly higher level of information was available with somewhat reduced spatial coverage:
- PCT_TOT_ALG: Estimated total cover of all Algae on transect %
- PCT_TOT_SGRS: Estimated total cover of all Seagrass on transect %
- PCT_SUB_COMP: Estimated mud/silt, sand/gravel, rubble, consolidated, rock on transect %

- **Torres Strait Effects of Trawling Study** – this dataset includes seabed fish abundance by species by station from a series of trawl surveys conducted in the region of the trawl fishery in central eastern Torres Strait during the mid 1980s.

- **Acoustics** – this dataset includes RoxAn Hardness and Roughness indices (and Depth) acquired during various surveys conducted 1986-2002.

2.2. OCEANOGRAPHIC DATA

2.2.1. Tides and currents

The bathymetric complexity of Torres Strait severely limits exchanges of both water and wave energy between the Gulf of Carpentaria and Coral Sea. The semidiurnal and diurnal tides contain most of the tidal energy on both sides of the strait, the dominant constituents being O_1, K_1, M_2, S_2, and N_2. However, bottom friction within the straits dissipates much of the tidal energy. Sea-level is therefore not coherent across the strait and the spring-neap cycles are not aligned. This mismatch can result in sea-level differences across the strait of up to 6 m (Wolanski et al. 1988), which drives tidal currents of almost 2 m s^{-1} in the major shipping channel (Clarke 1990).

Non-linear interactions with tidal flows also cause significant dissipation of low-frequency (sub-tidal) signals through the straight. Low-frequency sea-level is incoherent across the strait with differences of up to 0.3 m (Wolanski et al. 1988). Residual currents through the strait are typically less than 0.15 m s^{-1} (Wolanski et al. 1988, Harris 1991). They also tend to reverse with the seasonal change from the summertime northwesterly monsoon winds to the winter south-easterly trade winds (Harris 1991). The long-term average currents may therefore be as small as 0.01 m s^{-1}, corresponding to a through strait transports as low as 10^4 m^3 s^{-1} (Wolanski et al. 1988).

Modelling of sea-level and currents in Torres Strait began with highly simplified "channel flow" models representing the balance between sea-level difference, bottom friction, and local acceleration.

While these models can yield realistic estimates of local tidal currents through tuning of bottom friction parameters (Clarke 1990), the inclusion of nonlinear interactions is critical in reproducing the observed increase in transmission through the strait with frequency (Wolanski et al. 1988). Depth-integrated nonlinear models appear to adequately reproduce tidal elevations throughout the strait (Bode and Mason 1995). However, more recent model development has focused on three-dimensional solutions of the full equations over realistic bathymetry (Hemer et al. 2003). This model included realistic wind and wave forcing. Freshwater inputs representative of the Fly River were also included, but surface freshwater and heat fluxes were neglected. The results showed good agreement with sea-level observations, although the observed currents were more difficult to reproduce.

2.2.2. Hydrographic conditions

Seasonal hydrographic conditions, such as temperature, salinity, dissolved oxygen, and nutrients, can be estimated for Torres Strait from historical cast data. A least-squares mapping of this data has recently been developed in the form of the CSIRO Atlas of Regional Seas or CARS (Ridgway et al. 2002). The mapping methodology explicitly accounts for separation of water masses by land and complex bathymetry (Dunn and Ridgway 2002), and is therefore well suited to regions such as Torres Strait. However, the spatial and temporal coverage of the data is quite restricted in the strait. For example, temperature and salinity casts were nearly all taken during the monsoon (mainly March) and there has very little nutrient or oxygen data collected in the local region (Figure 2.2-1).

Climatological temperatures from CARS show a broad peak over the summer monsoon period of between 29 and 30°C, with waters in the neighbouring Gulf of Carpentaria up to 1°C warmer than those in the northwestern Coral Sea (Figure 2.2-2 and Figure 2.2-3). During winter, a cold-water anomaly (≈ 25°C) forms over the shallow Torres Strait, presumably in response to local heat loss to the atmosphere.

Salinities from CARS fall quite rapidly from around 35 to 32 psu under the influence of the monsoon rains (Figure 2.2-2). A strong freshwater anomaly along the PNG coastline is evident during March (Figure 2.2-3), when the data coverage is relatively dense (Figure 2.2-1). While salinities gradually increase again following cessation of the monsoon, there is evidence that the trades can occasionally introduce freshwater from the Fly River plume to the northeast (Harris et al. 1993, Wolanski et al. 1995, Hemer et al. 2003). For example, low salinity water (24 psu) observed for around two weeks in the Great North East Channel was attributed to repeated exposure of Gulf of Papua coastal water to the Fly plume during reversals in the alongshore current, prior to being advected southwestward into Torres Strait (Wolanski et al. 1999). While the Fly plume generates a strong halocline in the Gulf of Papua (Wolanski et al. 1995), tidal mixing generally prevents the development of vertical stratification within Torres Strait.

CARS suggests that dissolved oxygen levels in the well-mixed waters of Torres Strait tend to be relatively high (Figure 2.2-3), with limited seasonal variability (Figure 2.2-2). However, the data coverage is extremely sparse (Figure 2.2-1) and local enhancements in northern Torres Strait and the Gulf of Papua are largely artefacts of the mapping (Figure 2.2-3).

Nitrate, phosphate, and silicate all follow a similar trend, increasing over the course of the monsoon, before diminishing under trade wind conditions (Figure 2.2-2). Such patterns are consistent with enhancement through riverine inputs. On the basis of these distributions, it might be concluded that nutrients are unlikely to limit primary productivity in Torres Strait at any time of the year. However, it should be emphasized that the existing data coverage (Figure 2.2-1) is insufficient to support the apparent enhancement of nitrate and phosphate in Torres Strait and further northwest relative to neighbouring Gulf of Carpentaria and Coral Sea waters (Figure 2.2-4).

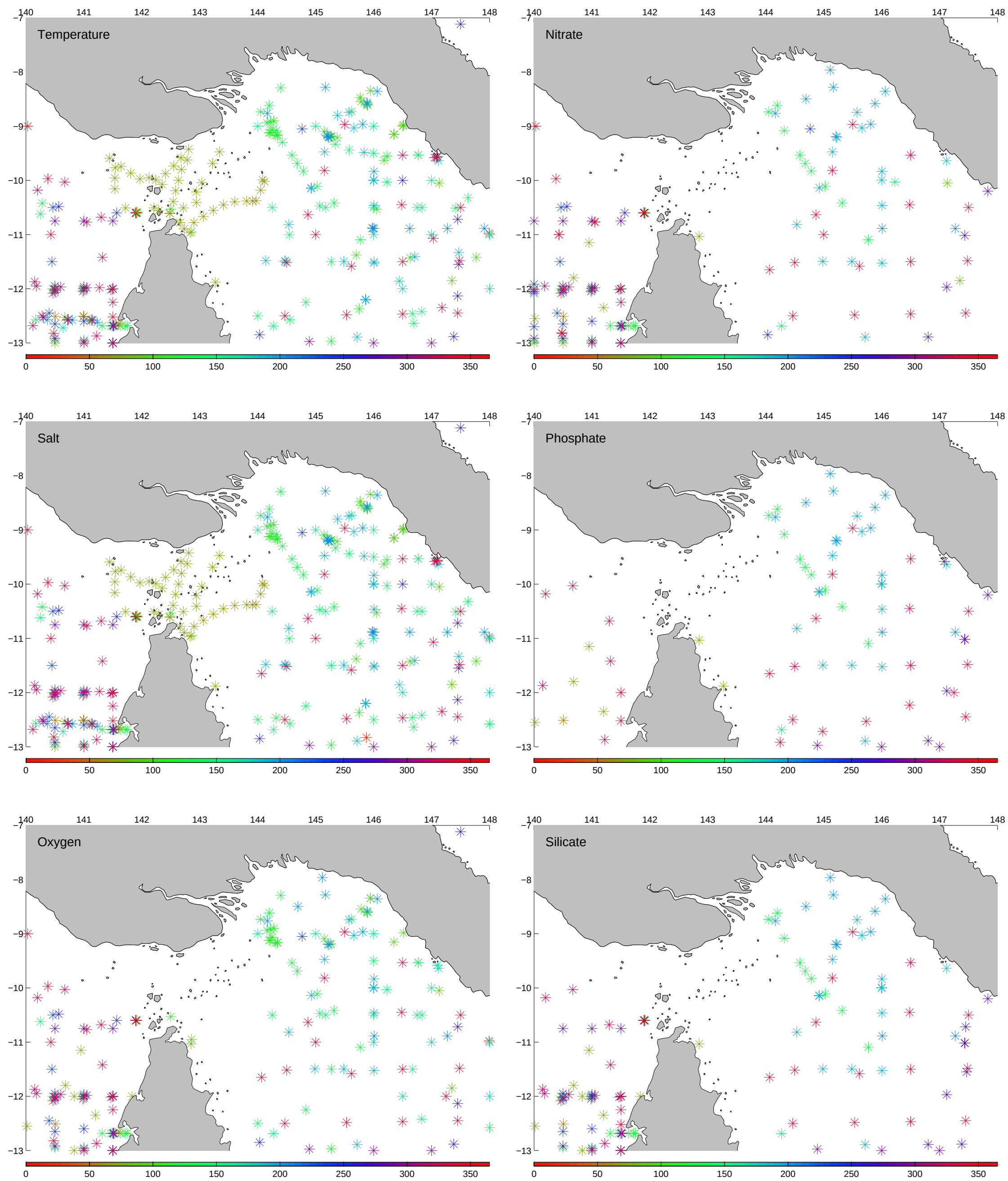


Figure 2.2-1 Distribution of hydrographic casts used to derive the CARS maps. The colours indicate the day of the year that individual casts were taken and thereby provide an indication of potential seasonal biases.

In southwestern Torres Strait, near Booby Island, a hydrographic station was regularly occupied from 1977 to 1983 as part of the CSIRO Coastal Monitoring Program. This data provides some insight into the interannual variability of the region (Figure 2.2-5). Temperature follows a very regular annual cycle over the monitoring period, which closely matches the seasonal trend estimated from CARS for the site. Salinity also tends to follow a regular seasonal pattern following CARS, although there is some evidence of low-salinity anomalies around the end of the monsoon in 1978 and slightly earlier in 1982. In contrast, salinities remained relatively high during the 1982-83 monsoon.

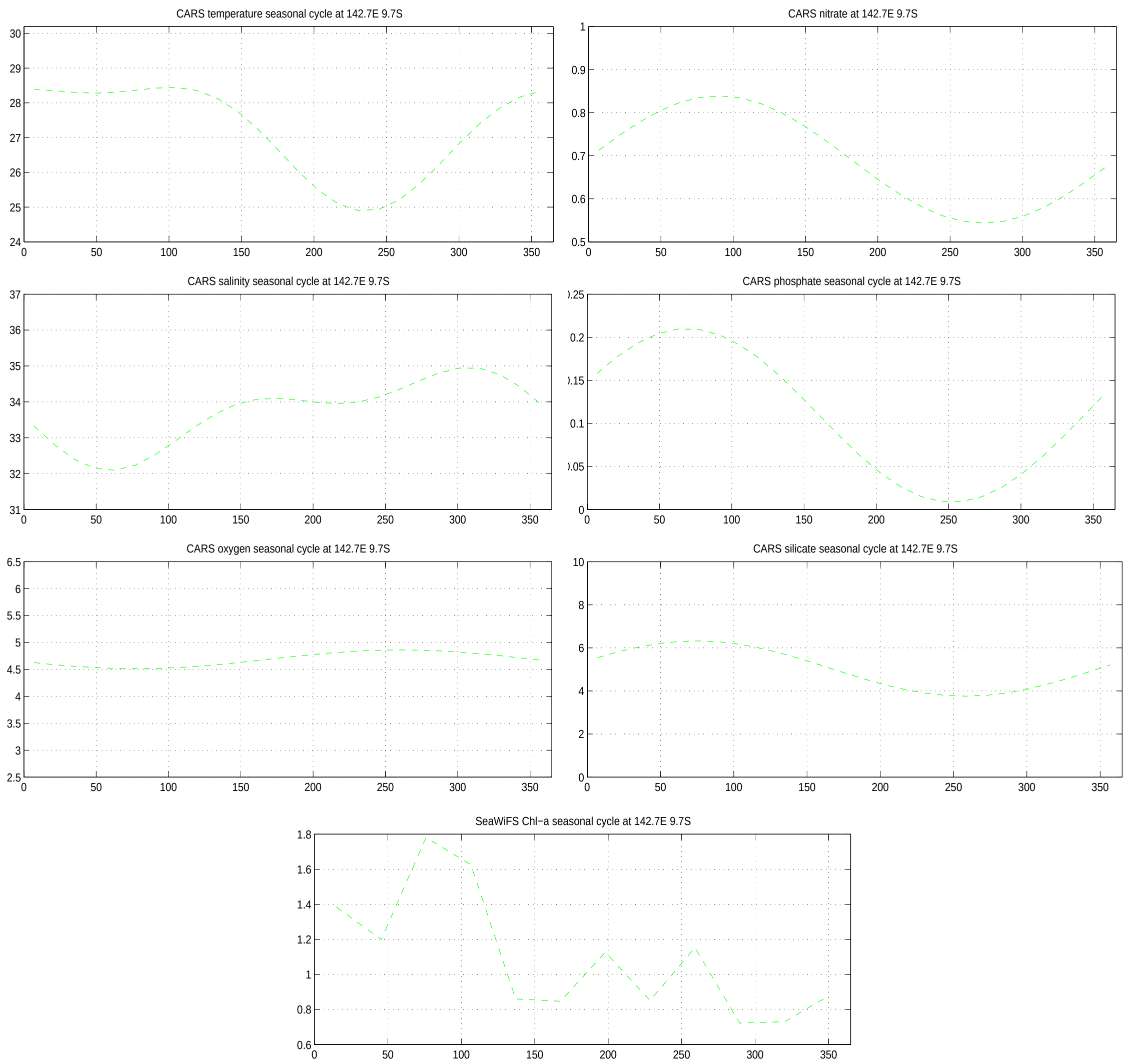


Figure 2.2-2 Seasonal trends in near-surface temperature (°C), salinity (PSU), dissolved oxygen (ml l^{-1}) (left) and nutrients (ml l^{-1}) (right) in central northern Torres Strait (142.7°E, 9.7°S) from CARS. Monthly averaged chlorophyll estimates based on SeaWiFS at the same location is also shown (mg m^{-3}) (bottom centre).

Dissolved oxygen exhibits much higher levels of interannual variability than temperature or salinity (Figure 2.2-5). While CARS correctly predicts that oxygen levels rise during the trade winds then fall during the monsoon, the observed range is at least three times that of CARS (even ignoring possibly suspect records from 1979). Nitrate levels also show high interannual variability, including an extended period of enhancement in 1977-78, and low values from late 1979 through 1980. With the exception of 1980, 1982 and 1983, strong peaks regularly occurred near the end of the monsoon period. The average seasonal trend confirms that the CARS nitrate distributions are not yet reliable in Torres Strait.

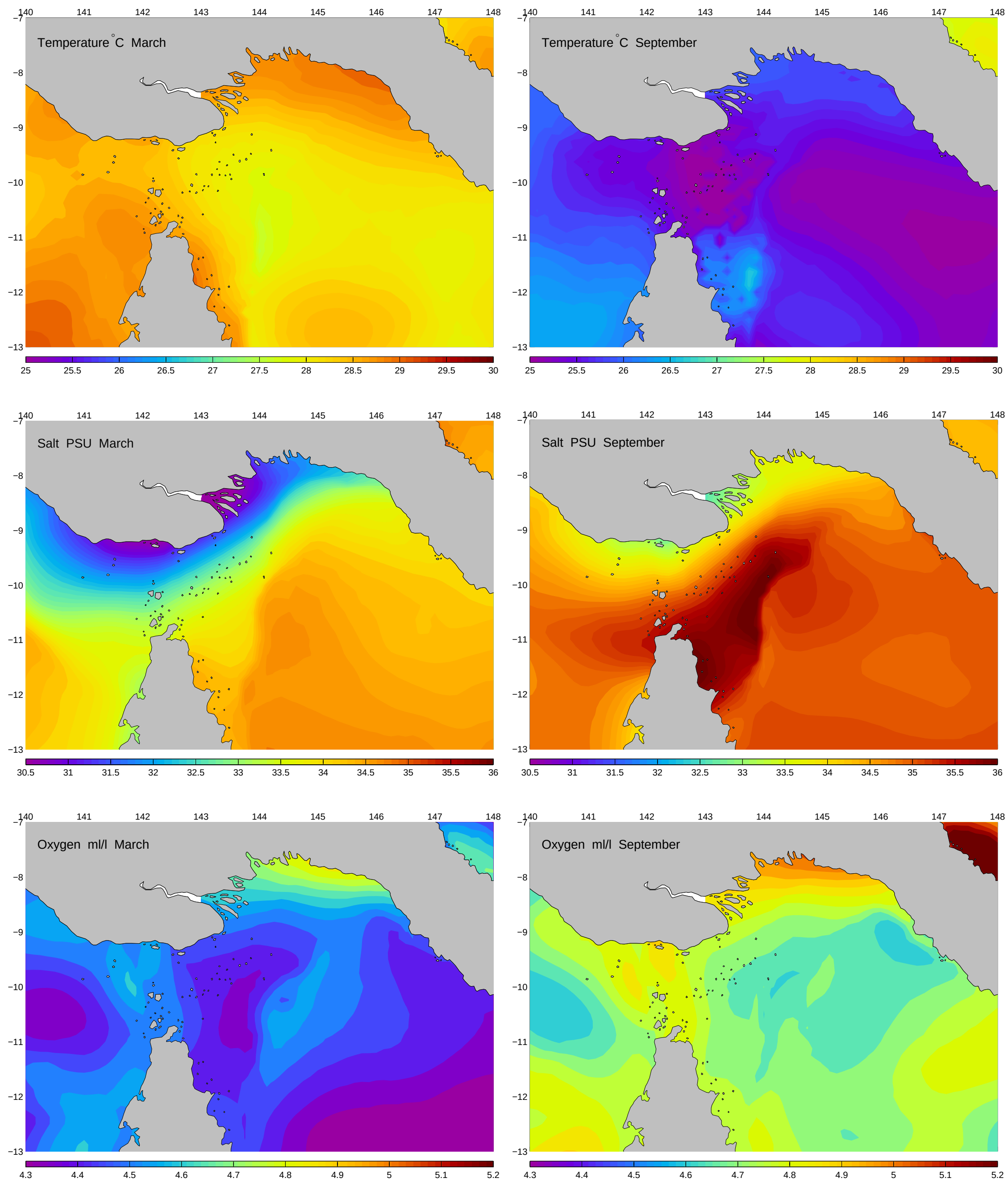


Figure 2.2-3 Seasonal maps of near-surface temperature (top), salinity (center), and dissolved oxygen (bottom) for March (left) and September (right) from CARS.

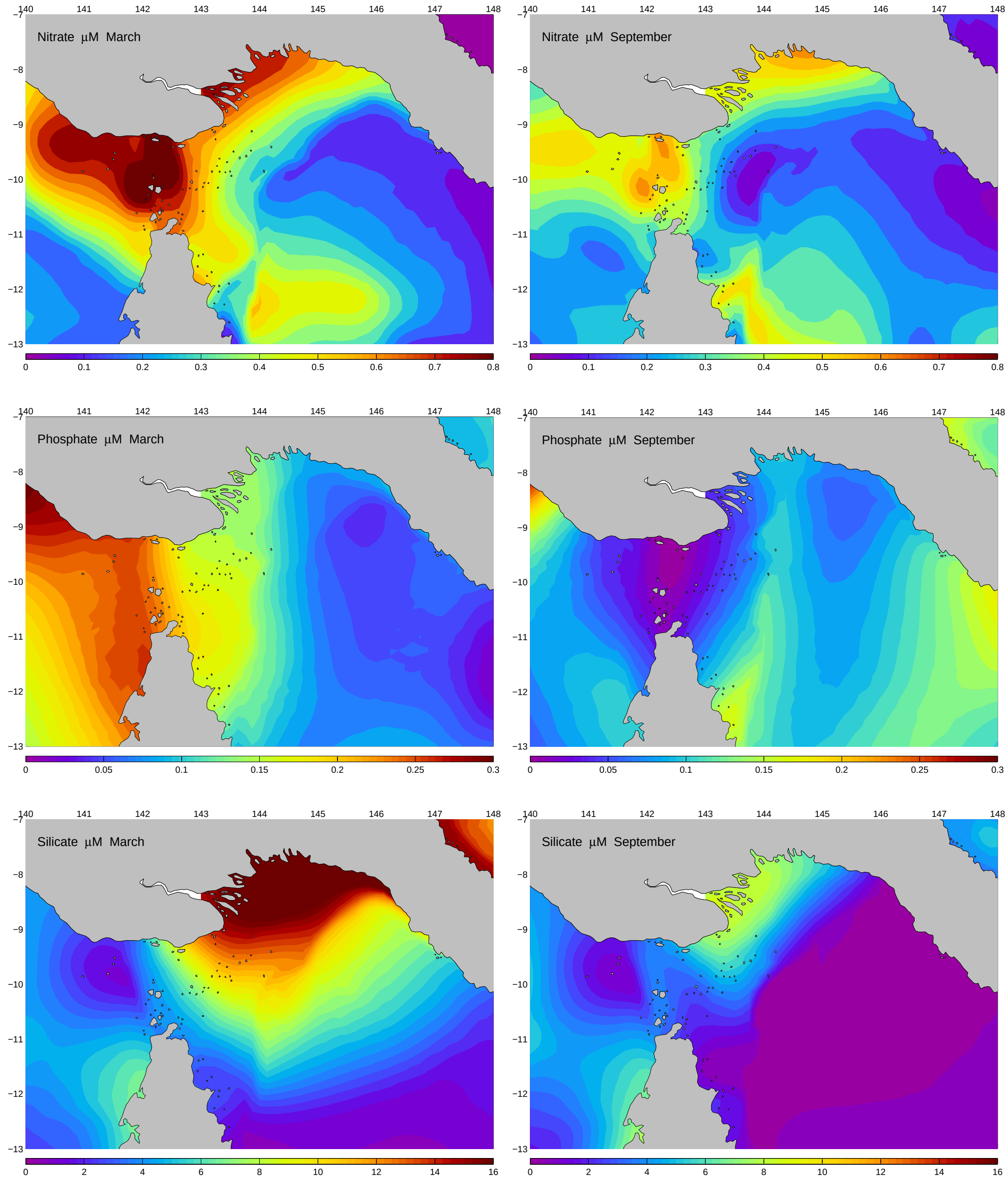


Figure 2.2-4 Seasonal maps of near-surface nitrate (top), phosphate (center), and silicate (bottom) for March (left) and September (right) from CARS. There are large uncertainties associated with nutrient distributions along the PNG coast and within Torres Strait due to the poor data coverage (Figure 2.2-1).

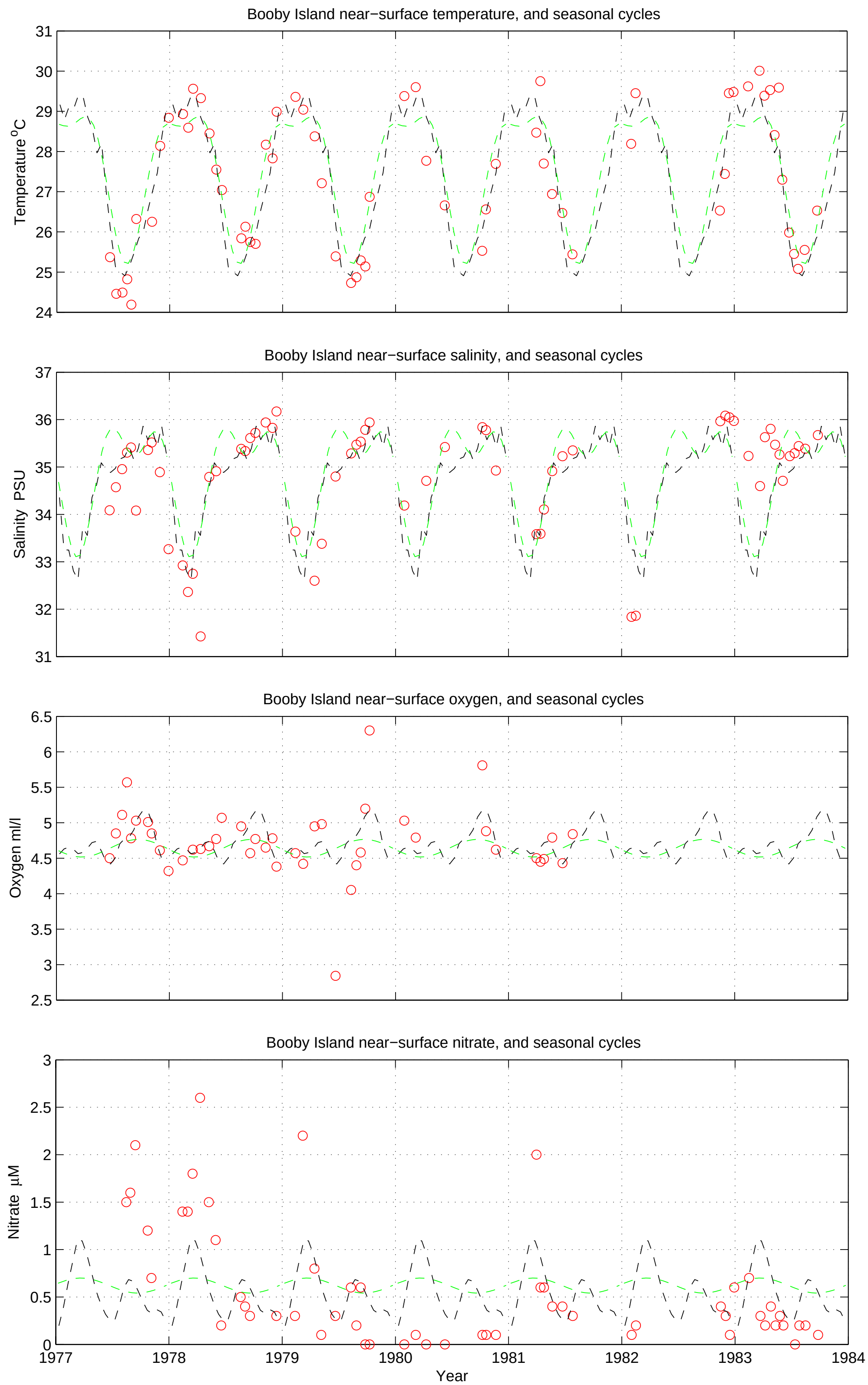


Figure 2.2-5 Measurements of temperature, salinity, dissolved oxygen, and nitrate from the Boobie Island Coastal Station (o), compared to the seasonal trend averaged across years using the same dataset (---), and CARS estimates for the same site (---). Note that the CARS analysis includes the Boobie Island data.

2.2.3. Suspended sediments

Tidal currents and locally generated surface waves are responsible for resuspension of sediments, giving rise to a turbidity maximum in central Torres Strait (Harris and Baker 1991). Suspended sediment concentrations in Missionary Passage have also been observed to increase from under 10 mg l^{-1} during neap tide to 20-30 mg l^{-1} during spring tide (Harris 1999). Model results suggest that sediments from the Fly River plume are likely to enter the strait, with increased loads during the trade-wind season (Hemer et al. 2003). This conclusion is consistent with observations of a low salinity event during September 1994 (Wolanski et al. 1999).

2.2.4. Chlorophyll

There are no in situ plankton measurements available from Torres Strait region. It is also likely that high suspended sediment loads, coupled with extensive regions of shallow water and persistent cloud cover, will introduce significant errors into chlorophyll estimates based on satellite ocean colour (Figure 2.2-6). Within these limitations, the seasonal trends in ocean colour suggest that highest chlorophyll levels occur during the monsoon season (Figure 2.2-2 and Figure 2.2-7). As winter approaches, chlorophyll peaks in the Gulf of Carpentaria (Burford and Rothlisberg 1999) and Coral Sea, but decreases in Torres Strait, where it remains relatively constant until the return of the monsoon (Figure 2.2-7).

There is currently insufficient information to determine what factors control primary production in Torres Strait. The seasonal peak in chlorophyll coincides with that in nutrients (Figure 2.2-2) and nitrate levels at the Boobie Island station are sometimes depleted (Figure 2.2-5). However, it seems likely that light availability will also be a factor in the turbid waters of Torres Strait, as found in the neighbouring waters of the Gulf of Carpentaria (Burford and Rothlisberg 1999). The phytoplankton community in Torres Strait may also have similarities to the Gulf (Rothlisberg et al. 1994). For example, the availability of both nitrate and silicate suggests that diatoms will form a major part of the phytoplankton biomass. However, there is currently no data available on the plankton community structure in Torres Strait.

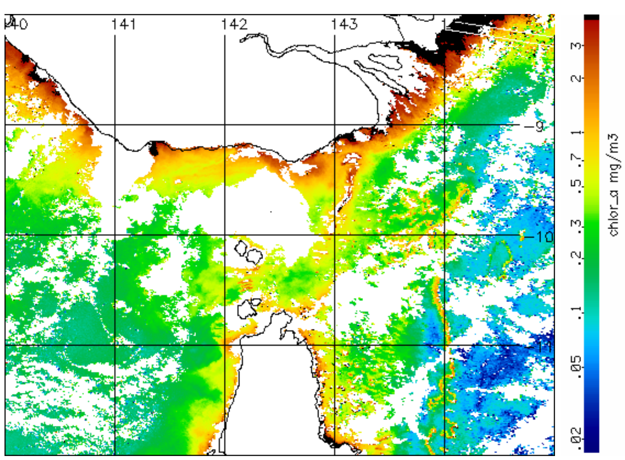


Figure 2.2-6 An example of chlorophyll distribution in Torres Strait derived from SeaWiFS ocean colour data (31 Mar 2000). The white patches over water correspond to unreliable data that has been excluded. For instance, the solid patches near the centre of the strait correspond to shallow water (< 15 m) where the bottom is likely to contribute to the signal, while the more diffuse patches indicate cloud cover. Despite this filtering major reefs, such as the northern Great Barrier Reef and Warrior Reefs, still appear as pseudo high chlorophyll patches.

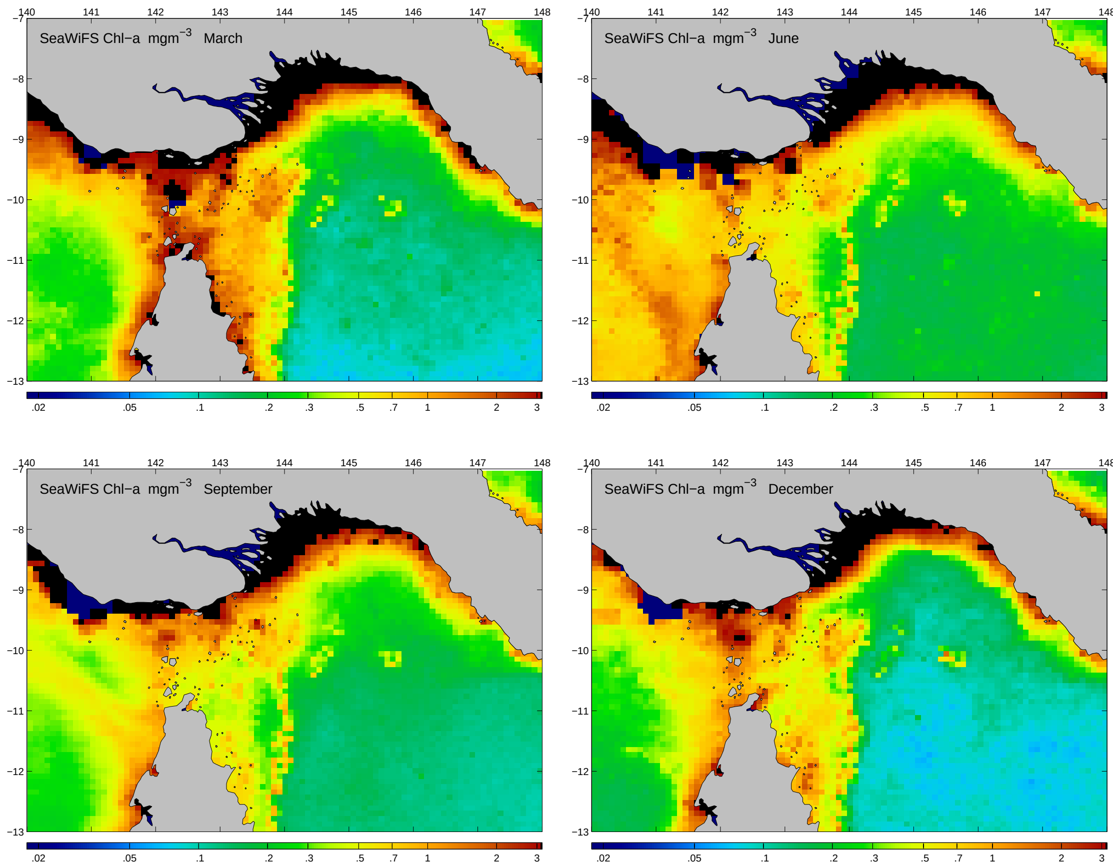


Figure 2.2-7 Seasonally averaged chlorophyll estimates derived from SeaWiFS ocean colour data for March, June, September, and December (mg m^{-3}). The black patches over water indicate that no reliable data was received for that location.

2.3. PHYSICAL COVARIATE DATA

2.3.1. Data Processing

After accounting for redundancy among the collated data, and collapsing the monthly SeaWiFS data to annual average and variability, there were 21 covariates (+8 measures of variability in the CARS & SeaWiFS attributes) for developing bio-physical models of biological survey data and for stratification of the TS non-reef region. These datasets were checked and imported into an ArcInfo GIS.

We constrained the covariates to the continental shelf by establishing a base study area bounded by the Torres Strait Protected Zone and adjacent areas but excluding those areas beyond the continental shelf.

A 36-arc-second grid (0.01 decimal degree, (~1.11 km) was generated for this area. Each grid cell was assigned a unique identifier that was subsequently used as the key to this dataset. As the collated data were of various spatial resolutions, we resampled those data to the 36-arc-second grid framework by a discrete thin plate spline technique (Wahba, 1990) using the TOPOGRID module in ArcInfo, to provide a consistent set of full-coverage covariates for the Project. As many of the covariates were not available for every grid cell, a "reliability indicator" was calculated that represented the distance to the nearest source data.

The TS wide coverage of all of the collated covariates was thematically mapped using a colour range appropriate to the individual distribution.

The interpolated physical data for each grid cell were exported out of ArcInfo for statistical analysis. This physical data set was also geographically matched to the location of each sampling station in the Biological Survey datasets. These were also exported from the ArcInfo GIS into a database suitable to provide physical covariates matching biological sample data for statistical analyses of bio-physical relationships.

2.3.2. Maps of Physical Covariate Data

2.3.2.1. Bathymetry

Torres Strait is a complex shallow area of continental shelf between Cape York and Papua New Guinea (Figure 2.3-1) (Harris, 1995). The main features are two central ridge lines extending from Cape York to PNG, coming to the surface at numerous places as reefs and islands, and dissected by numerous channels. In the far east, the Great Barrier Reef extends northward onto the PNG shelf at the western extremity of the Gulf of Papua. Outside the barrier, the slope drops very steeply into 2,000-4,000 m depths. In eastern Torres Strait, behind the barrier, there are numerous shallow reefs and in the northeast, these form large complexes dissected by deep channels. These deep channels, and those just south of the PNG coastline, are old river beds that continue to be scoured by tidal currents. Western Torres Strait grades gently into the Gulf of Carpentaria. Northwest Torres Strait is very shallow and mostly un-navigable by hydrographic vessels.

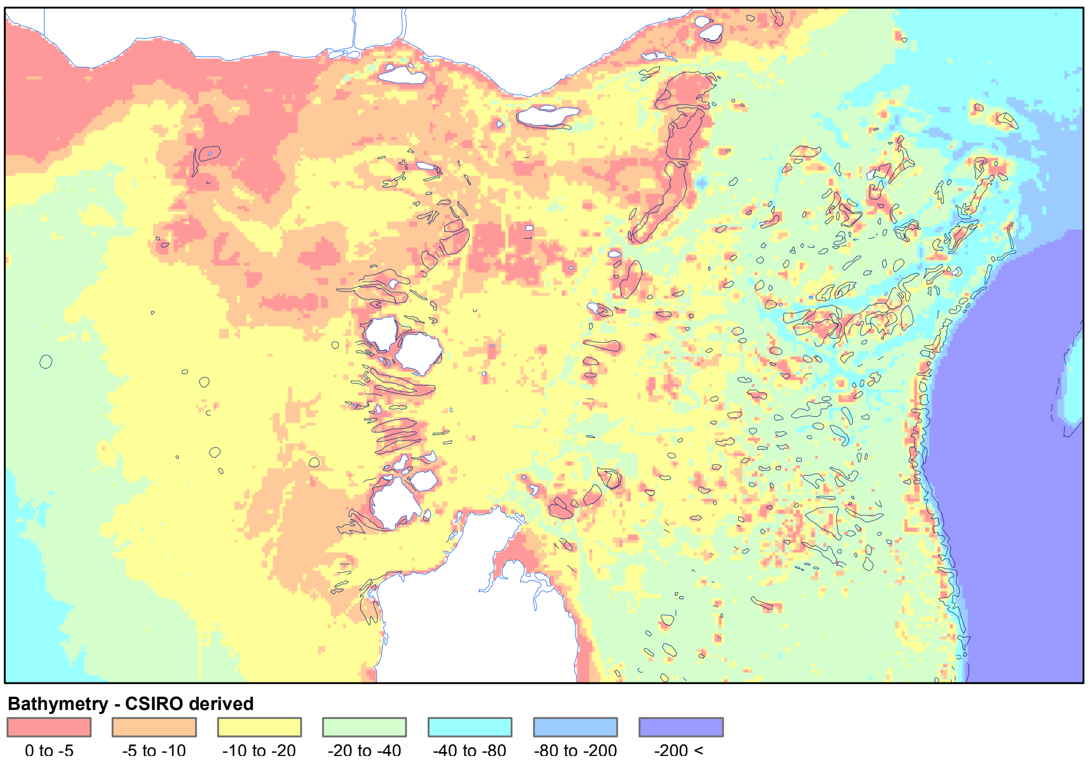


Figure 2.3-1 DEM of the bathymetry (m) of Torres Strait, mapped onto a 0.01° grid, from various sources (see section 2.1.1).

The bathymetry of the major shipping channels of Torres Strait is well surveyed for navigation purposes (Figure 2.3-2, dark blue areas through Prince of Wales, Adolphus and Great NE Channels). Much of the remainder of Torres Strait, however, is poorly mapped and potentially unreliable for bio-physical mapping. Areas of particular concern in this regard include: NE Torres Strait, which is very complex with large formations of reefs and shoals, deep areas and even deeper channels; and NW Torres Strait, which is very shallow with complex sand ridges and shoals mostly uncharted.

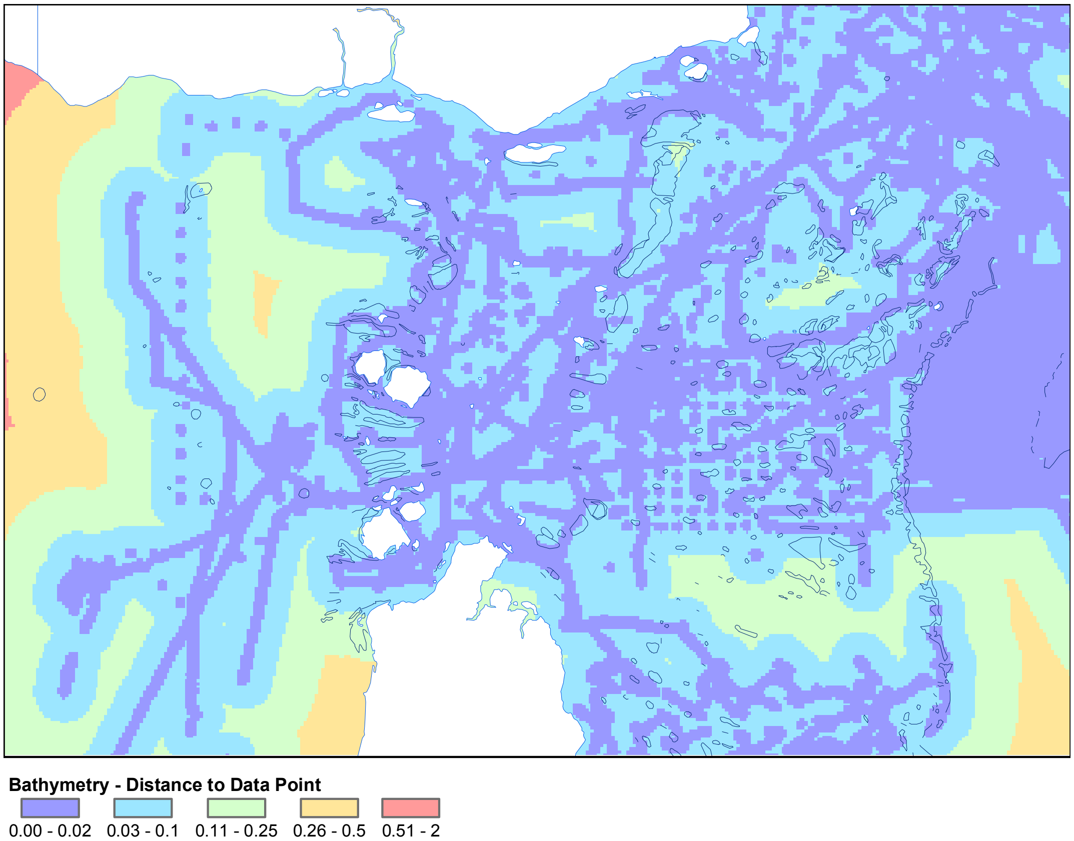


Figure 2.3-2 Map of distance (decimal degrees) to soundings, mapped onto a 0.01º grid, as an indication of bathymetric reliability and data gaps for Torres Strait.

2.3.2.2. Sediment Attributes

The surface sediments of Torres Strait are a reflection of the terrestrial inputs, particularly from major river systems such as the Fly in the NE, and biogenic production of carbonate skeletons by foraminiferans, bryozoans, algae and corals, modified by strong tidal currents, particularly in narrow channels between reefs, and sea level change during periods of glaciation (Harris, 1991, 1995). High mud areas include: off the Fly River delta, west of Warrior Reef, the Great NE Channel, and east of Cape York (Figure 2.3-3). Gravel especially dominates areas between reefs and islands where strong tidal currents scour finer sediments away, depositing them in dunes beyond the channels, leaving gravel and/or pavement (Figure 2.3-3) (Harris, 1991). Rock is distributed similarly to gravel, but is more constrained (Figure 2.3-3). The sand fraction is most ubiquitous, dominating wherever mud, gravel and rock do not (Figure 2.3-3). Characteristic grain size reflects the distribution of the sediment fractions coarse (red) to fine (blue), and grain sorting indicates the range of grain size from homogeneous (blue) to widely mixed sizes (red) (Figure 2.3-3). The composition of most of Torres Strait sediments is carbonate, with low carbonate areas close to Cape York and PNG indicating the input of terrestrial sediments (Figure 2.3-5).

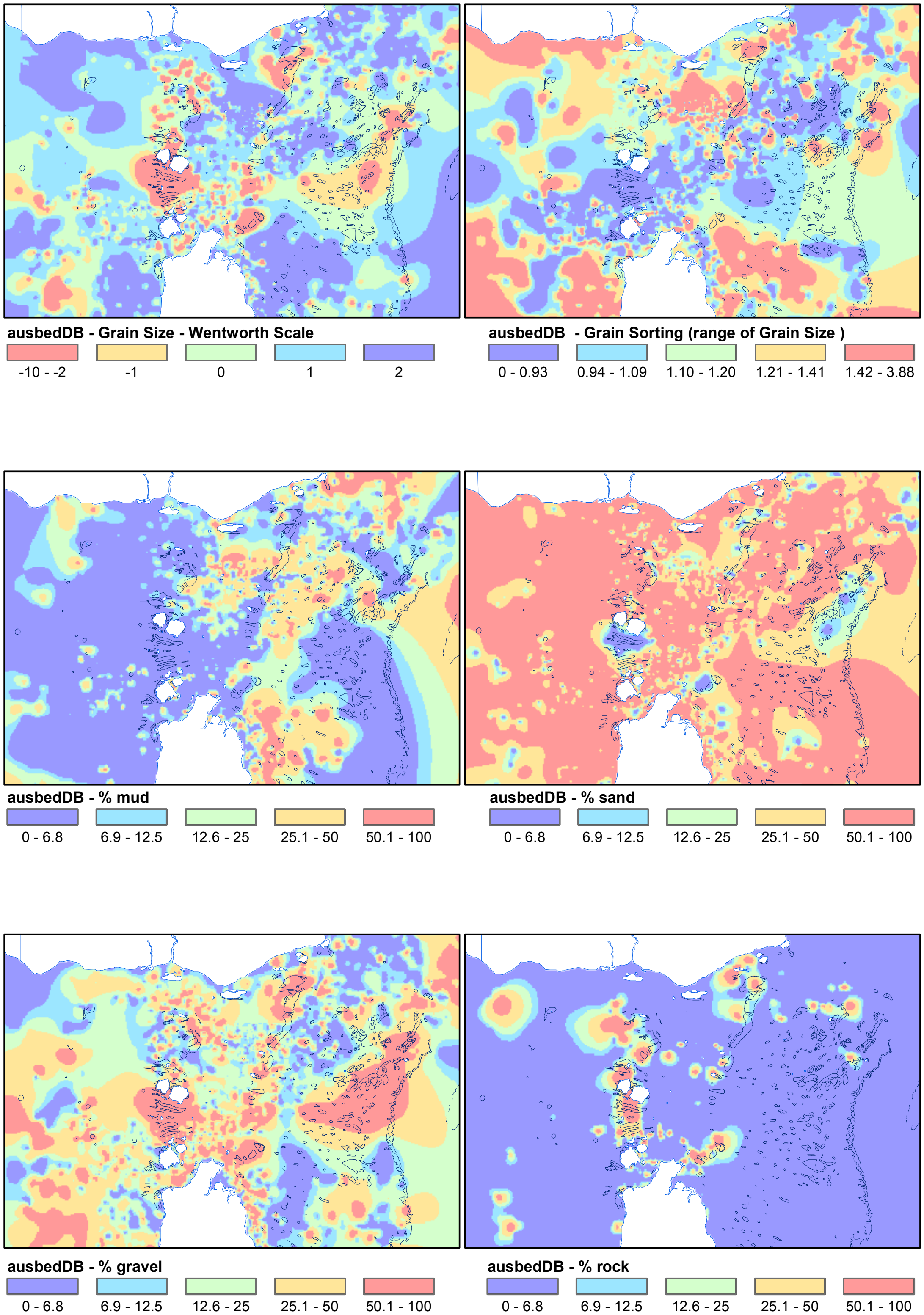


Figure 2.3-3 Maps of sediment grain size attributes for Torres Strait: characteristic grain size, sorting, and percent mud/sand/gravel/rock fractions (source, see section 2.1.1).

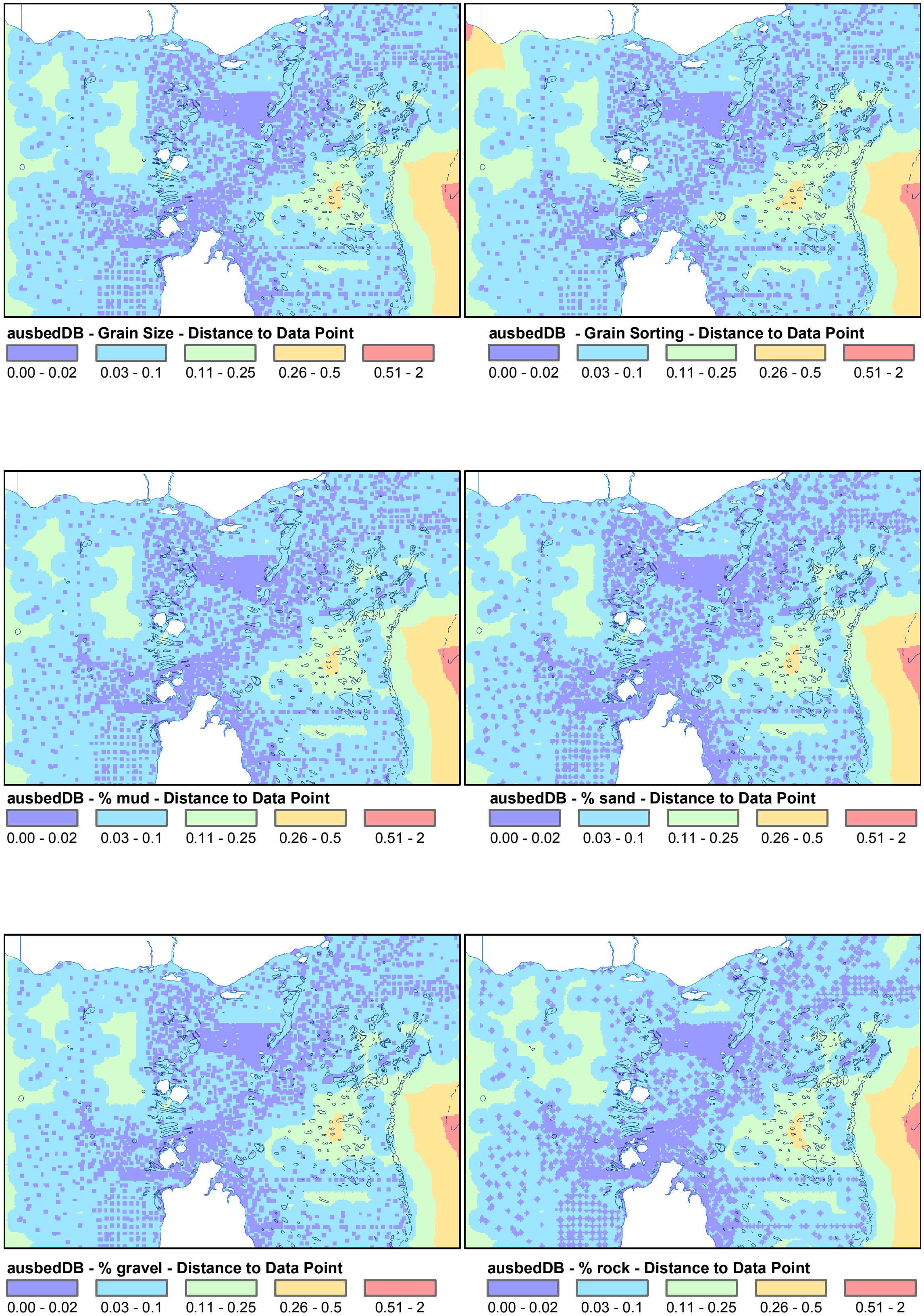


Figure 2.3-4 Map of distance (decimal degrees) to sample sites for sediment attributes, mapped onto a 0.01º grid, as an indication of reliability and data gaps for sediment attributes in Torres Strait.

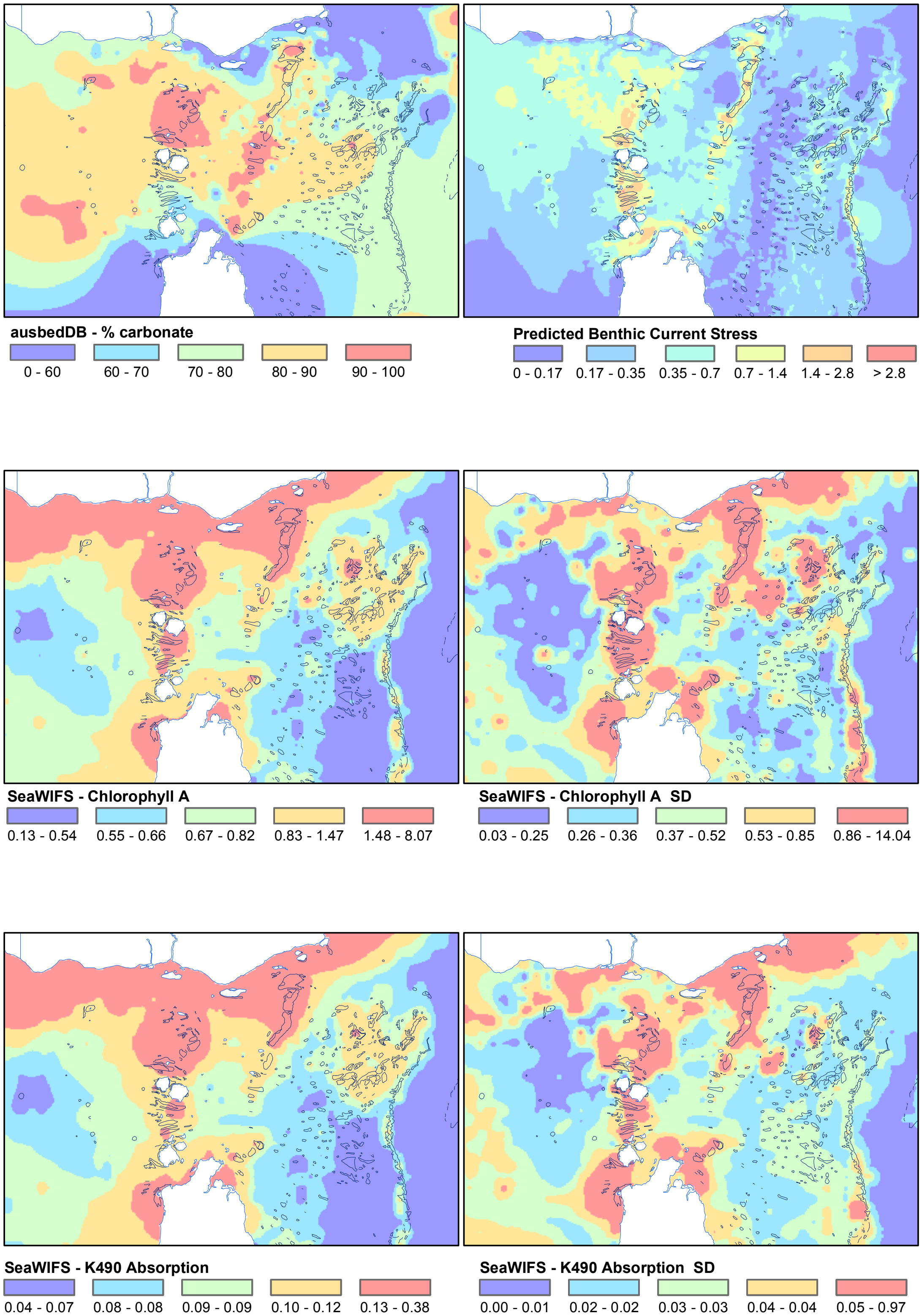
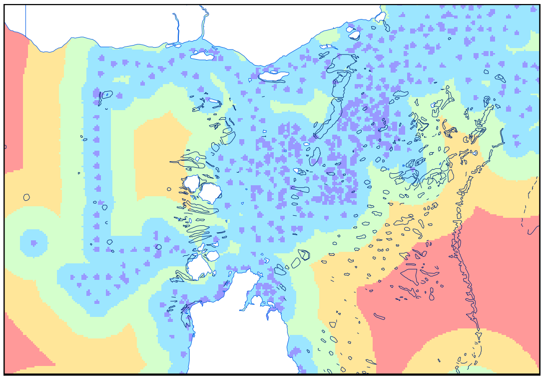


Figure 2.3-5 Maps of sediment carbonate composition, mean modeled seabed current stress, SeaWiFS predicted chlorophyll-a and standard deviation, and light absorption (attenuation coefficient) at 490 nm and standard deviation for Torres Strait (sources, see section 2.1.1).



The sediment grain size attributes of the main shipping channels and of an area west of the southern warrior reefs have been relatively well sampled (Figure 2.3-4), whereas as most of eastern and northwestern Torres Strait are largely unsampled. The remainder has been only patchily and inadequately sampled. The gaps for carbonate data are even more significant (Figure 2.3-6).

Figure 2.3-6 Map of distance (decimal degrees) to sample sites for sediment carbonate, mapped onto a 0.01° grid, as an indication of reliability and data gaps for this sediment attribute in Torres Strait.

2.3.2.3. Seabed Current Stress

The physical oceanography of Torres Strait is dominated by the tidal regime, which generates extremely strong currents. These tidal currents are driven by the Coral Sea/Gulf of Papua and Gulf of Carpentaria/Arafura Sea tidal cycles, which are largely out of phase causing large sea level gradients across Torres Strait (Bode & Mason, 1995; section 2.2.1). The tidal currents exert a force on the seabed, that in turn causes friction to the flow of water, known as seabed current shear stress, which redistributes sediments (Harris, 1991, 1995) and appears to influence biotic assemblages (section 2.7).

The areas of highest seabed stress occur where the tidal currents are forced through the narrow channels between the reefs and islands of western Torres Strait, the Warrior Reefs system, and to a lesser extent in eastern Torres Strait and the outer barrier (Figure 2.3-5). The shallows of northwestern Torres Strait are an extensive area of moderately high current stress (Figure 2.3-5) that transports sediments and forms large dunes and sand banks (Harris, 1991, 1995). The higher stress areas correspond with larger grain size fractions and, conversely, low stress areas correspond with finer grain size fractions (cf. Figure 2.3-3). The higher stress areas also correspond with the occurrence of benthic sponge and gorgonian gardens (Figure 2.4-2, Figure 2.4-3, Figure 2.4-4).

The reliability of the seabed current stress data is dependent on the availability of bathymetric data, which has significant gaps (section 2.3.2.1), and on the resolution and accuracy of the current modelling (Bode & Mason, 1995). There are few tidal and current monitoring stations in Torres Strait (section 2.2) against which to test model results, nevertheless, those that do exist correspond well (Bode & Mason, 1995).

2.3.2.4. Ocean Colour (chlorophyll & turbidity)

The use of SeaWiFS ocean colour data for estimating chlorophyll is discussed in section 2.2.4, and seasonal patterns are presented there. The annual mean and standard deviation (indicating seasonal variability) of estimated chlorophyll and turbidity (attenuation coefficient at 490 nm) are shown in Figure 2.3-5. The strong correlation between these remote sensed estimates can be readily seen, and is an issue for the reliability of this data (section 2.2.4).

2.3.2.5. Bottom Water Attributes

The CARS database of hydrographic measurements was discussed in detail in section 2.2.2, including seasonal and wider regional patterns at the sea surface. This section presents the CARS mappings at the seabed for the Torres Strait study area, as used in section 2.7. Unfortunately, there is very little oceanographic data for Torres Strait (Furnas, 1991; section 2.2.2), so the mappings presented here need to be considered with caution, as any apparent fine scale detail is likely to be an artefact of the interpolation.

The annual average salinity appears to be higher in western, southern and central Torres Strait, with an area of lower salinity in northern Torres Strait (Figure 2.3-7) due to riverine input from PNG, some of which originates from the Fly River (Furnas, 1991; Wolanski, 1991; section 2.2.2). The standard deviation (implied seasonal variability) appears to be higher in the northern area (Figure 2.3-7), due to the monsoonal seasonality of the riverine input. Temperature appears to be higher in western/central Torres Strait, with higher standard deviation in the same general area (Figure 2.3-7). Dissolved oxygen appears to be higher in central/northern Torres Strait, corresponding with the shallow areas of high tidal current energy. The higher oxygen areas seem to be seasonally consistent with low standard deviation in the same general area; areas with large standard deviation occur in eastern Torres Strait (Figure 2.3-7). There is likely to be tidally driven upwelling at the shelf break in eastern Torres Strait that may inject cooler water and nutrients onto the shelf seabed behind the barrier reef (Wolanski, 1991).

The concentration of silicates is an indication of the influence of terrestrial inputs (Furnas, 1991). The annual average silicate concentration tends to be higher in the area of the Great NE Channel (Figure 2.3-9) probably due to the influence of the Fly River (Furnas, 1991; Wolanski, 1991); the standard deviation of silicate also tends to be higher in this area. Nutrient levels (phosphate, nitrate) are generally low in Torres Strait (Furnas, 1991). The limited data suggest relatively higher values appear to occur in central/western areas and deep waters off the shelf, and standard deviations appear to be larger where annual means are higher (Figure 2.3-9), but these patterns need to be treated with caution (section 2.2.2).

There are broad areas of uncertainty for all bottom water attributes in Torres Strait, as indicated by the RMS residual of the mapping and scarcity of CTD casts (Figure 2.3-8, Figure 2.3-10). This is particularly so for oxygen, silicate and nutrients, but the additional once-off data for temperature and salinity offer limited improvement because off the temporal dynamics of these attributes.

2.3.2.6. Prawn Trawl Effort

Trawling for prawns occurs in central-eastern Torres Strait, with areas of highest effort in the vicinity of Yorke Island and extending towards Coconut Island (Figure 2.3-11). This area is typified by muddy-sand and low current stress, suitable for prawns. The variability in annual effort intensity corresponds closely with the amount of effort (Figure 2.3-11).

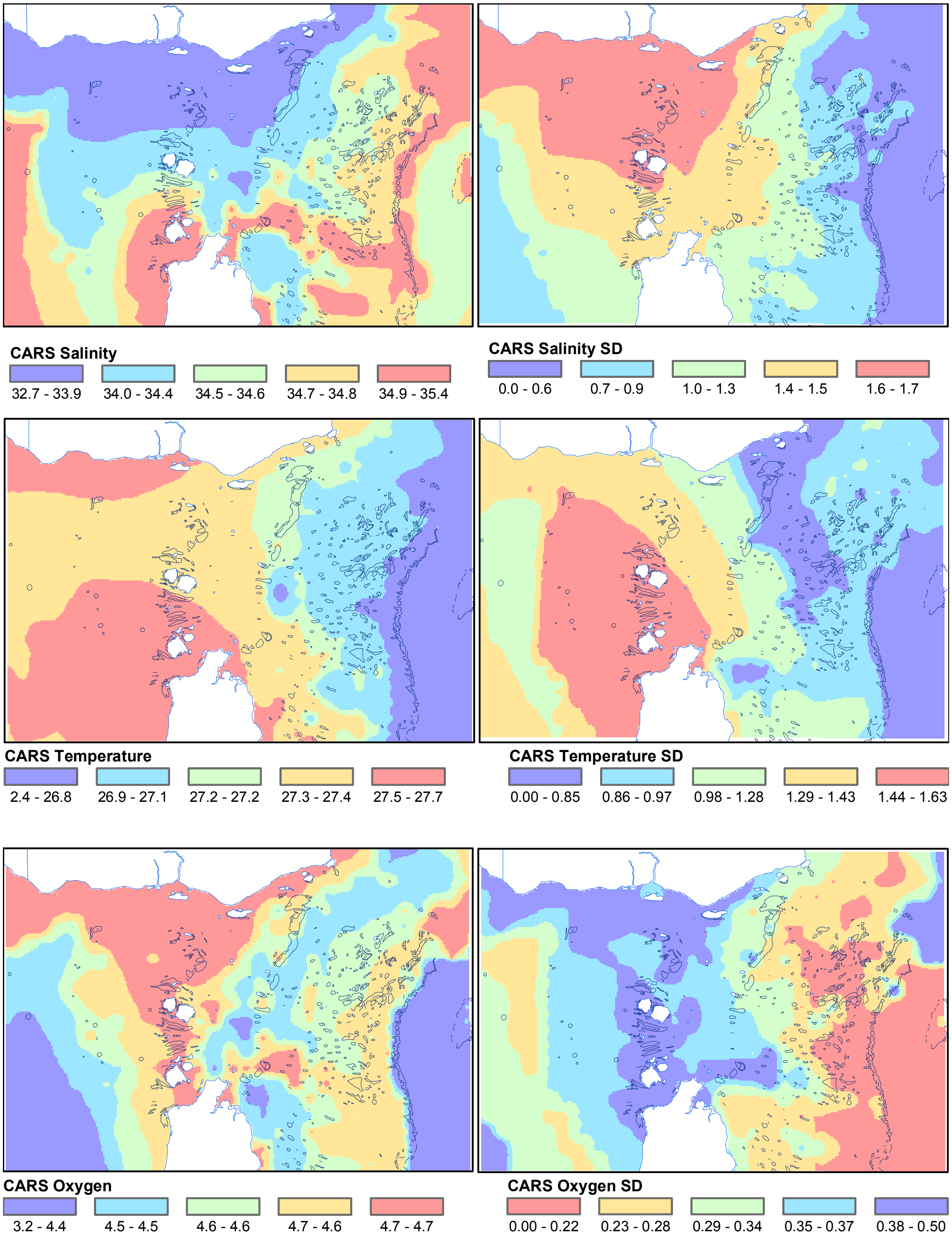


Figure 2.3-7 Maps of CARS bottom water physical attributes for Torres Strait: salinity (mean & SD), temperature (mean & SD), and dissolved oxygen (mean & SD), (source, see section 2.1.1).

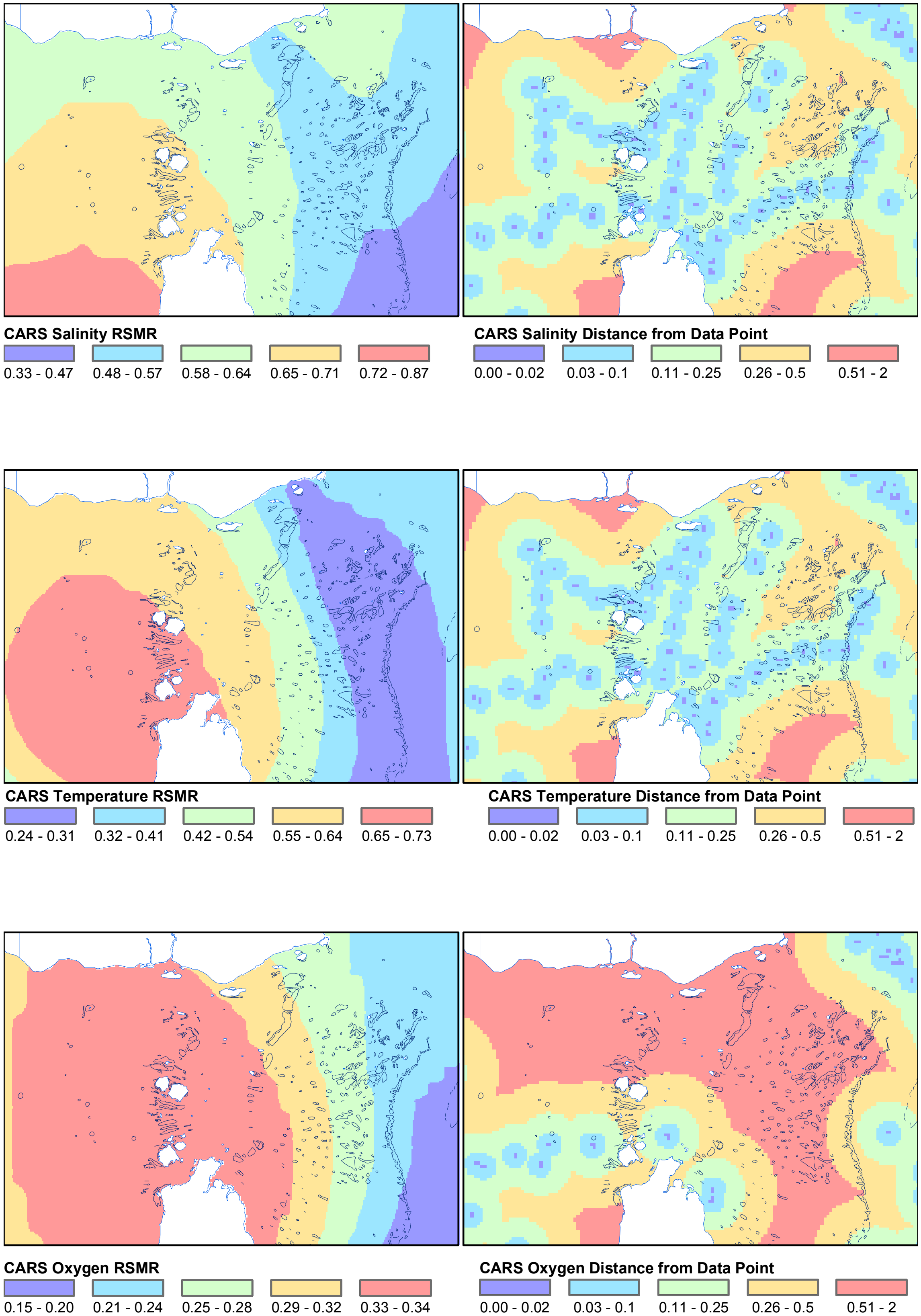


Figure 2.3-8 Maps of root-mean-square residual of CARS mapping, and distance (decimal degrees) to CTD casts for CARS attributes, mapped onto a 0.01° grid, as an indication of reliability and data gaps for physical water attributes at the seabed in Torres Strait.

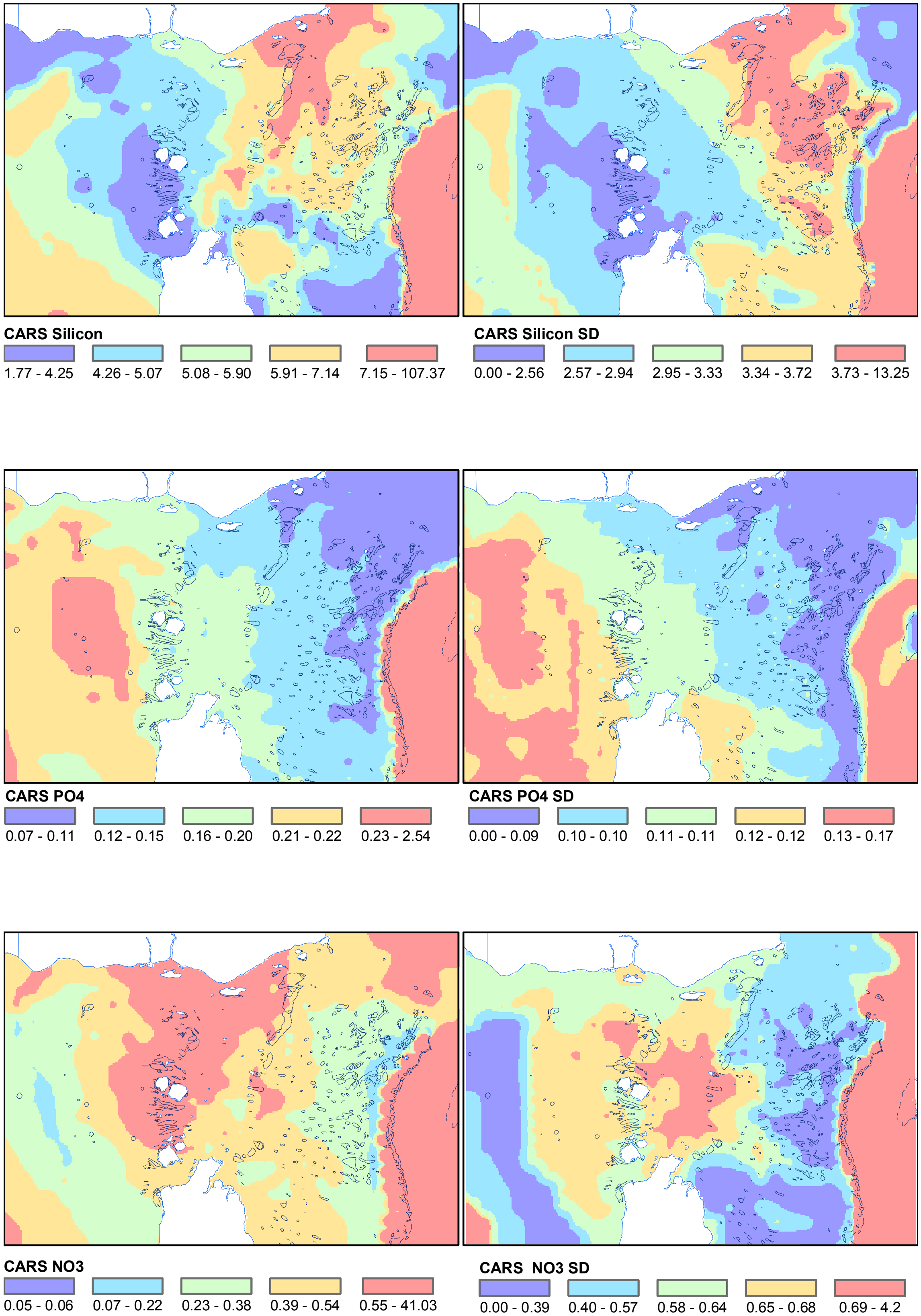


Figure 2.3-9 Maps of CARS bottom water nutrient attributes for Torres Strait: silicate (mean & SD), phosphate (mean & SD), and nitrate (mean & SD), (source, see section 2.1.1).

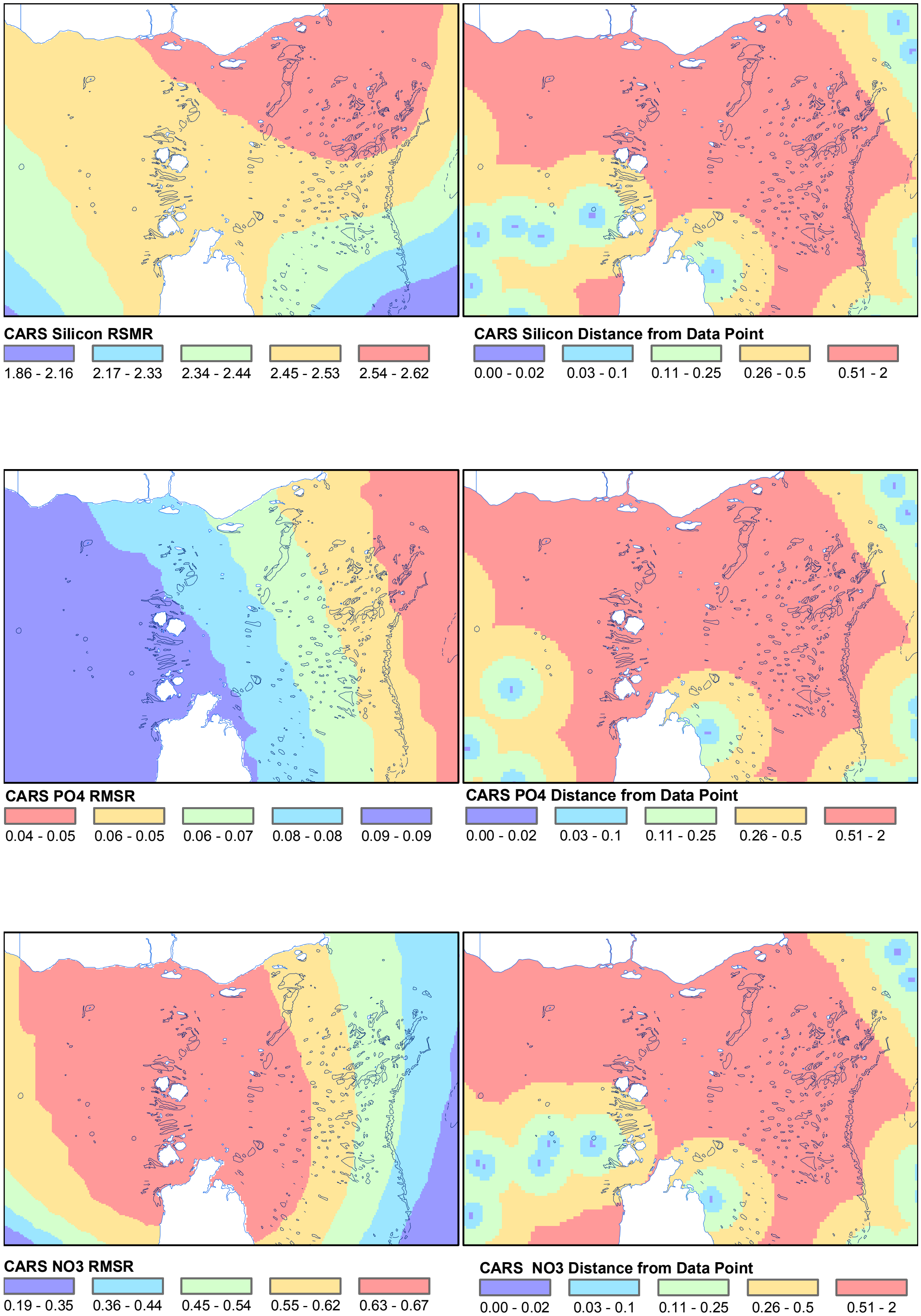
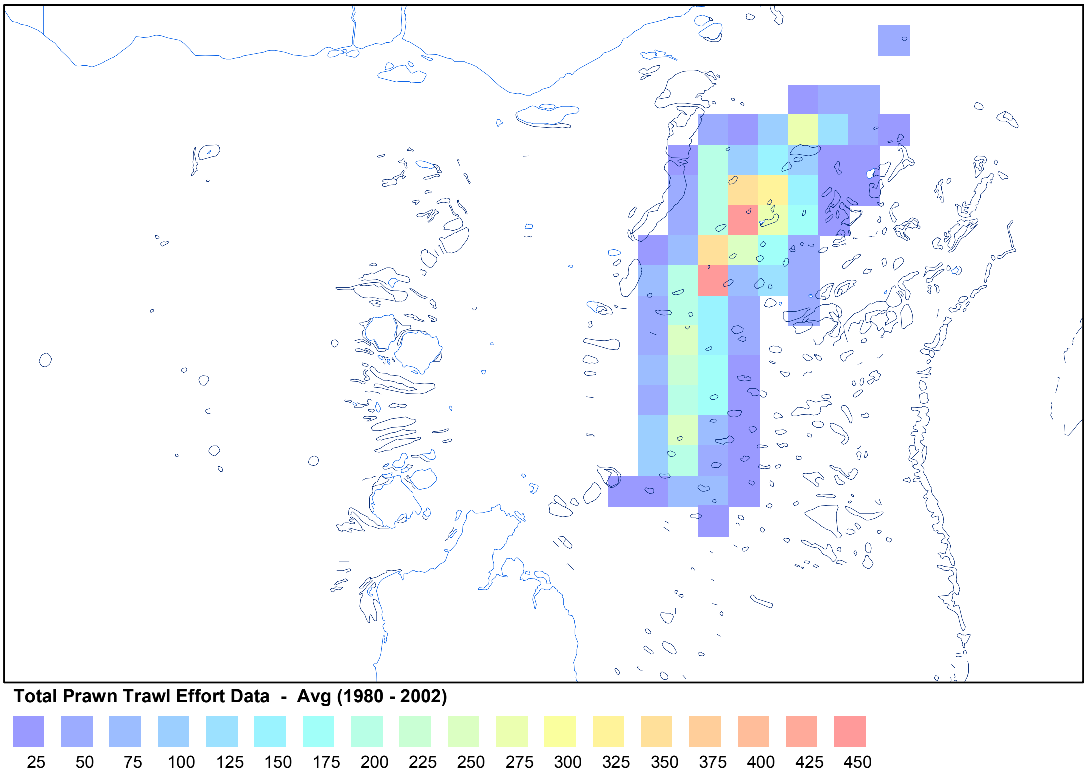


Figure 2.3-10 Maps of root-mean-square residual of CARS mapping, and distance (decimal degrees) to CTD casts for CARS attributes, mapped onto a 0.01° grid, as an indication of reliability and data gaps for water nutrient attributes at the seabed in Torres Strait.



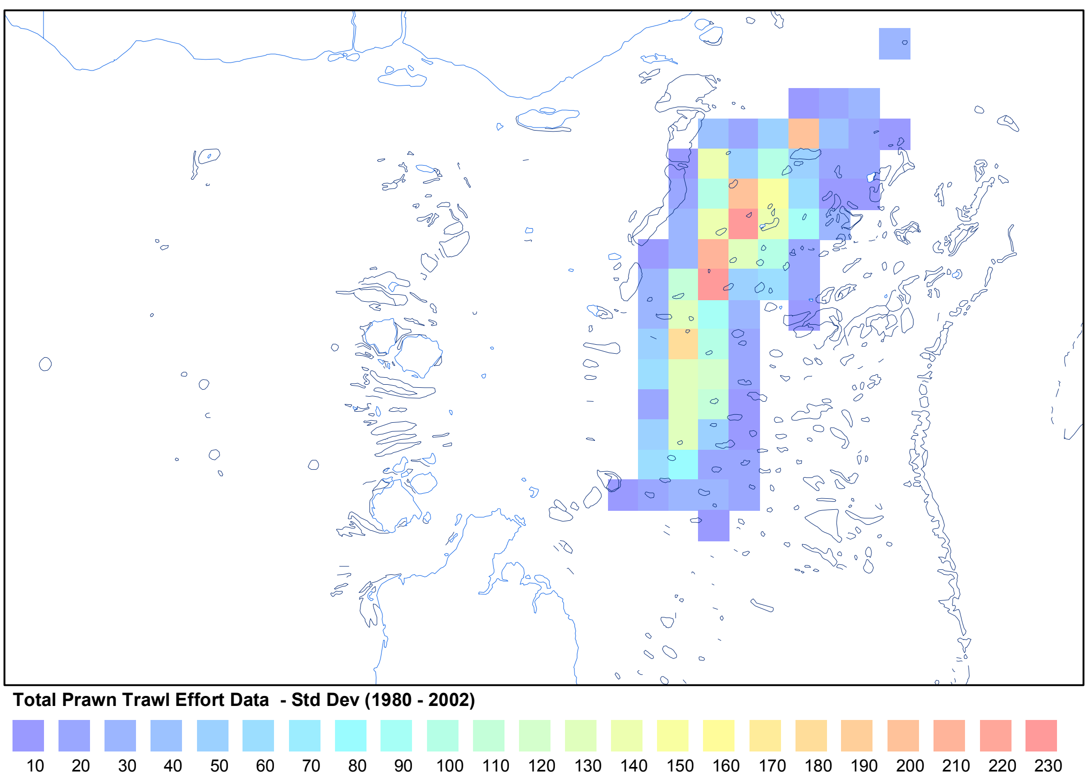


Figure 2.3-11 Maps of Torres Strait prawn trawling effort (boat-days per 6 min grid cell), average and standard deviation for years 1980-2002 (source, see section 2.1.1).

2.4. BIOLOGICAL SURVEY DATA

2.4.1. Data Processing

Biological data was amalgamated from 16 data sets from surveys in the Torres Strait study area conducted for a variety of purposes ranging from stock assessment of tropical rock lobster and Bech-de-Mer to environmental seabed mapping surveys. The surveys were carried out between 1987 and 2002. The information has been extracted from diver or video transects: divers recorded habitat components & coverage and video was coded into habitat facies in 'real-time' — both can be regarded as "Rapid Assessment" techniques. The video data included the frequency of major habitat facies such as: substratum type (mud, silt, sand, rubble, rock); epifaunal gardens (combinations of composition & density); and algae, seagrass, and bivalve shell beds. The diver data included similar estimates of percent cover of these (Pitcher *et al.*, 1992) and some additional components in more detail (some to genus or species). Altogether, the datasets contained 527 individual variables collected at over 1500 sites in the area of interest. Theses were amalgamated and combined to produce 26 standard variables that were considered as suitable for describing broad-scale environmental characteristics of the study area.

Table 2.4-1. Coverage of information common across the widest spatial coverage of Project and Cruises. (A) Widest coverage/simplest data, (B) slightly narrower coverage/more information (see section 2.1.1 for definitions of these fields). Scale: S = Small, M = Medium and L = Large.

(A) Dataset	Scale	Sub_Code	Bio_Code	Seagrass	Sub-Comp	Algae
Warraber	S	✓	✓			
KulasiWest	M	✓	✓	PA		PA
Lobster'89	L	✓	✓	PA		PA
PNG Lob'98	M	✓	✓	PA		PA
Lobster'02	L	✓	✓	PA		PA
IPC'96	L	✓	✓	PA		PA
Chevron'97	L	✓	✓	PA		PA

(B) Dataset	Scale	Sub_Code	Bio_Code	Seagrass	Sub-Comp	Algae
Warraber	S	✓	✓		✓ %	
KulasiWest	M	✓	✓	Tot%		
Lobster'89	L	✓	✓	Tot%	✓ %	Tot%
PNG Lob'98	M	✓	✓	Tot%	✓ %	Tot%
Lobster'02	L	✓	✓	Tot%	✓ %	Tot%
IPC'96	L	✓	✓		✓ %	Tot%
Chevron'97	L	✓	✓		✓ %	Tot%

The final dataset had a wide coverage for more general habitat descriptions such as coded substrate (1452/1501) and epibenthos abundance (1336/1501) categories (Table 2.4-1A). A somewhat more detailed substratum description that included the relative cover of several substrate categories were possible at 968/1501 sites, and algal & seagrass percent cover was possible at 490/1501 sites (Table 2.4-1B). The resulting biological dataset had a broad spatial coverage, north-south and east-west (Figure 2.4-1) which provided a wide match with the range of variability of physical covariates (section 2.3.2) to examine bio-physical relationships (section 2.7).

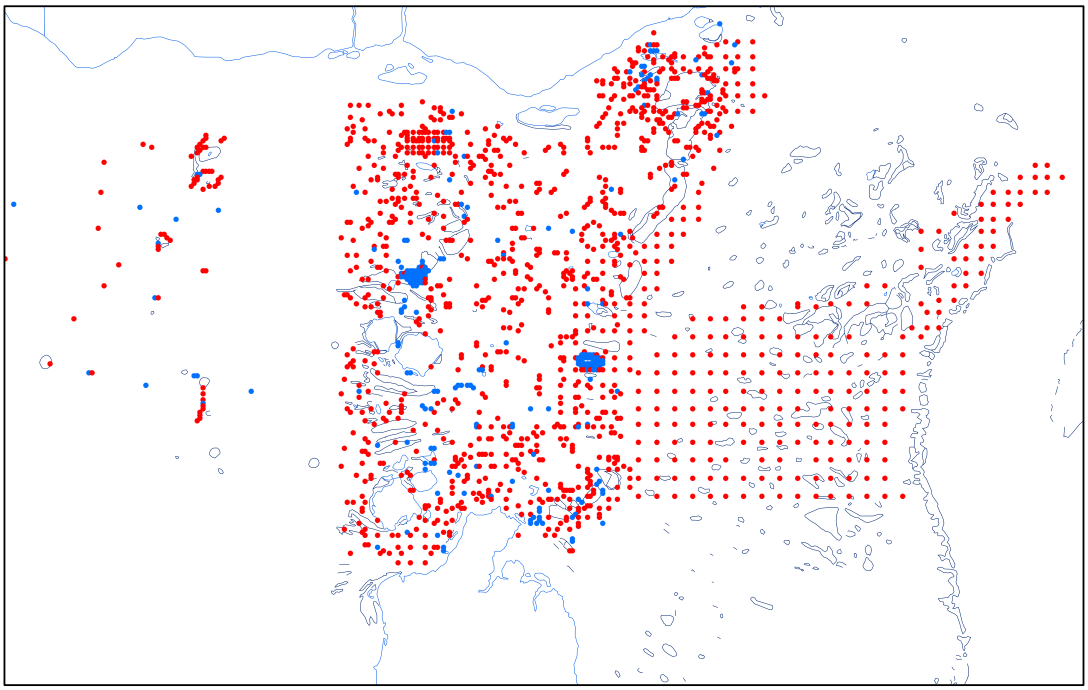


Figure 2.4-1 Map of the location of all seabed biological survey sites included in the data collated for this project (sources, see section 2.1.1).

2.4.2. Maps of Biological Survey Data

2.4.2.1. Substratum Type

The substratum type observed during dive or towed-video surveys was recorded in a number of forms: SUB_CODE (see section 2.1.1 for definition) was the most basic form and was common to all included surveys (Figure 2.4-2). During most towed-video surveys, the video was observed in real time and occurrence of mud, sand, rubble, rocks, reef, along with other seabed habitats was entered directly into a computer ("Tappity" data) an later summarized to percent occurrence on transects (Figure 2.4-4)and, with intermediate spatial coverage, estimated substratum proportions of mud, sand, rubble, consolidated and hard (Figure 2.4-4). These patterns correspond well with the sediment attribute data, with the same processes affecting the observed distributions (see section 2.3.2.2).

2.4.2.2. *Epibenthos Habitat*

The epibenthos habitat type observed during dive or towed-video surveys was also recorded in a number of forms: BIO_CODE (see section 2.1.1 for definition) was the most basic form and was common to all included surveys comprising a rough estimate of the abundance of epibenthos on a transect (Figure 2.4-2); "Tappity" data was entered in real time during most towed-video surveys and included sparse, medium and dense epibenthos gardens, along with other seabed habitats that were summarized to percent occurrence on transects (Figure 2.4-4); and one survey estimated the percent cover of epibenthos and coral (Figure 2.4-3). The patterns of epibenthos distribution correspond well with that of harder substratums (2.4.2.1) and with high seabed current stress (2.3.2.3).

2.4.2.3. *Algae and Seagrass*

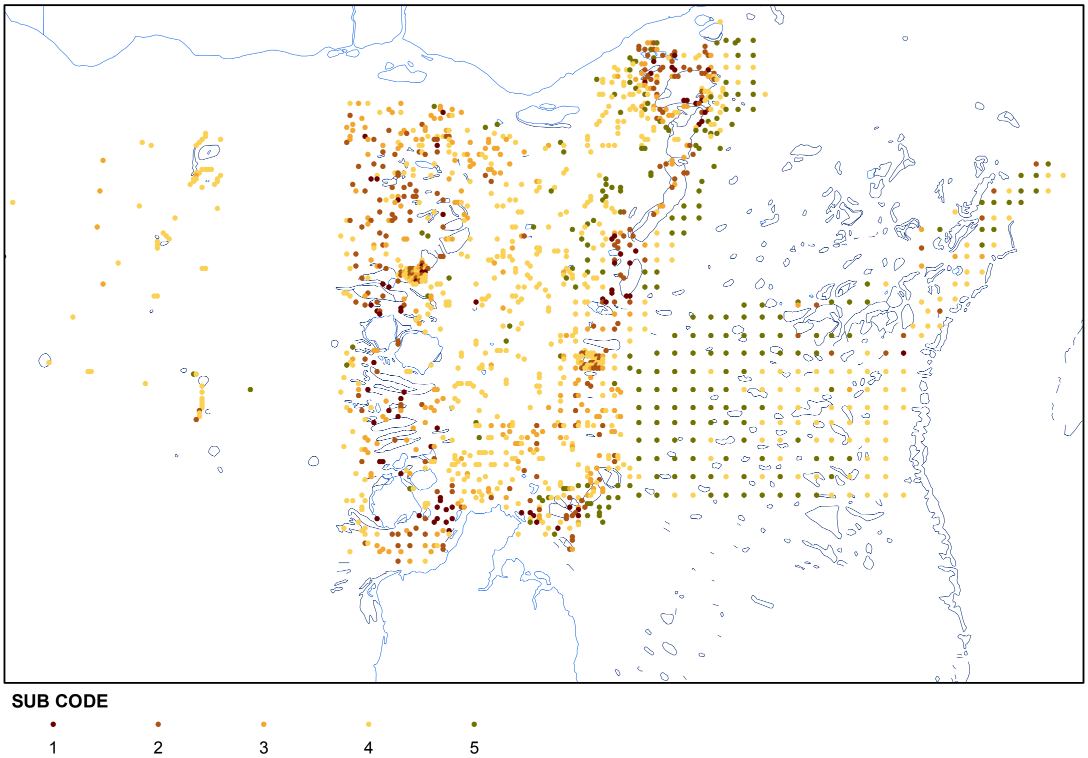
The algae and seagrass observed during dive or towed-video surveys were recorded as presence/ absence most broadly (Figure 2.4-5) and, in somewhat fewer surveys, as estimated percent cover (Figure 2.4-6). Algae were almost ubiquitous across Torres Strait and harder, higher current areas tended to have higher cover of algae (cf. 2.4.2.1 and 2.3.2.3). Of course there were many species of algae and different species occurred in different physical environments. Seagrass was more restricted to the north-west half of Torres Strait and there were few sites that had very high cover of seagrass.

2.4.2.4. *Seabed fishes*

During the mid 1980's a series of cruises was undertaken in the central eastern Torres Strait as part of a study of the effects of trawling in the region. The purpose of these cruises was to characterise the fish bycatch of the Torres Strait prawn trawl fishery and to compare fish assemblages in trawled and adjacent untrawled areas. Specifically, seven cruises were conducted between February 1985 and September 1986. The vessel towed 4 × 5 fathom prawn trawls for approximately 30 minutes at each site. The fish species in the catch were identified, weighed and measured. Catch data from each trawl was standardised as kg/ha and for this report, mean catch rates of each species at each site were calculated across all cruises. The location of the sampling sites was largely restricted to the area of the trawl fishery and a more limited area in the west of Warrior Reef closure (Figure 2.4-7).

In general, the sites inside the area closed to trawling, to the west of the Warrior Reef complex had more fish species (mean = 21.4 ±*sd* 10.7 species; range: 13 – 29.5) than those open to trawling (mean = 14.7 ±*sd* 7.9 species; range: 10.4 – 28.7) (p<0.001) (Figure 2.4-8). However, catch rates were similar inside the closure (mean = 7.1 ±*sd* 2.2 kg/ha; range:4.2 – 11.3 kg/ha) compared with outside the closure (mean = 6.3 ±*sd* 4.8 kg/ha; range: 2.4 – 32.4 kg/ha) (p<0.84), although one site within the trawl ground had a very high catch rate of fish (Figure 2.4-9).

The fish assemblages of this part of Torres Strait were characterised by clustering the sites and then these clusters were examined in relation to the physical characteristics using Linear Discriminant Function analysis (see Section 2.7.3).



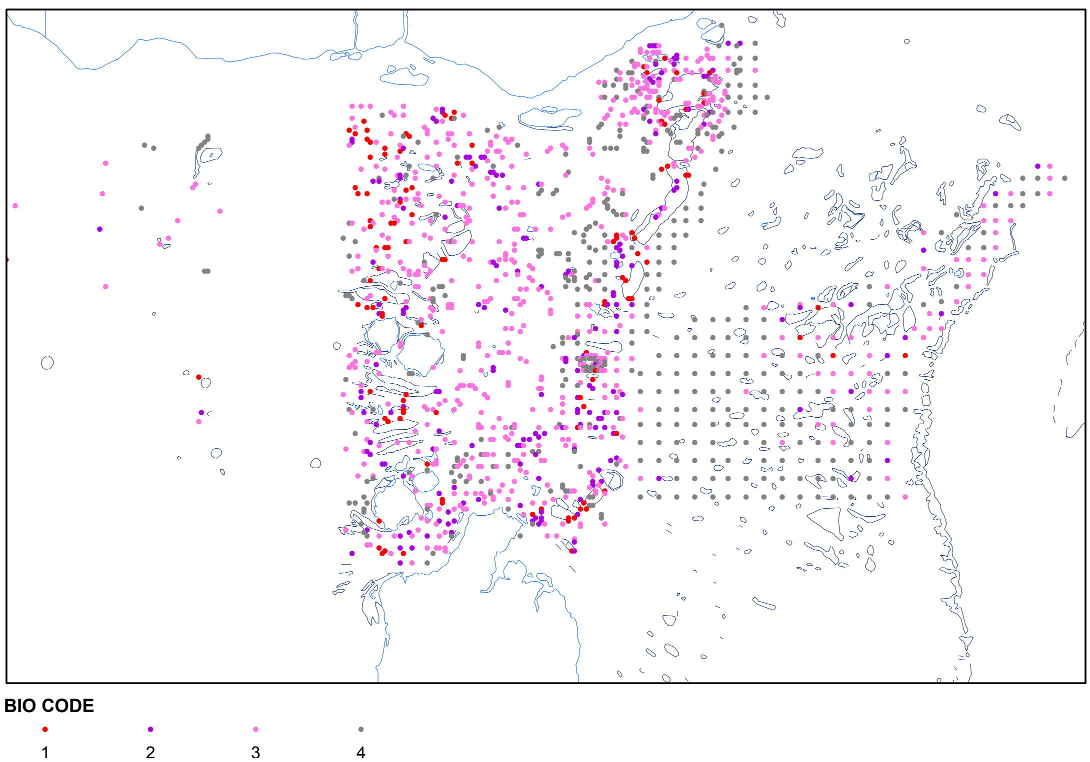
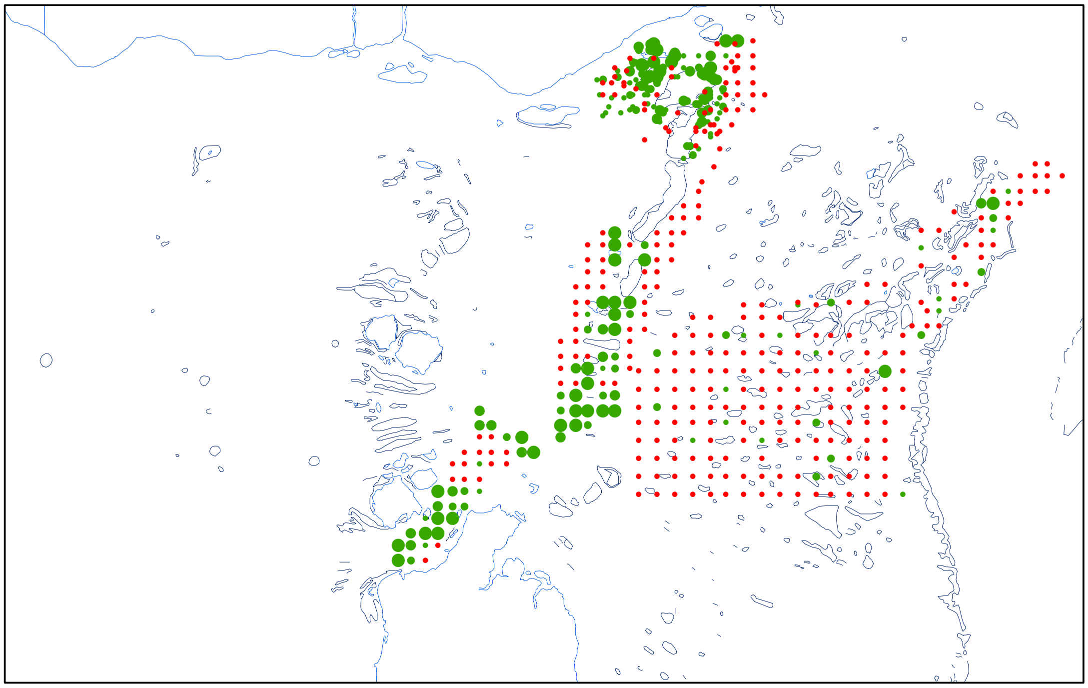
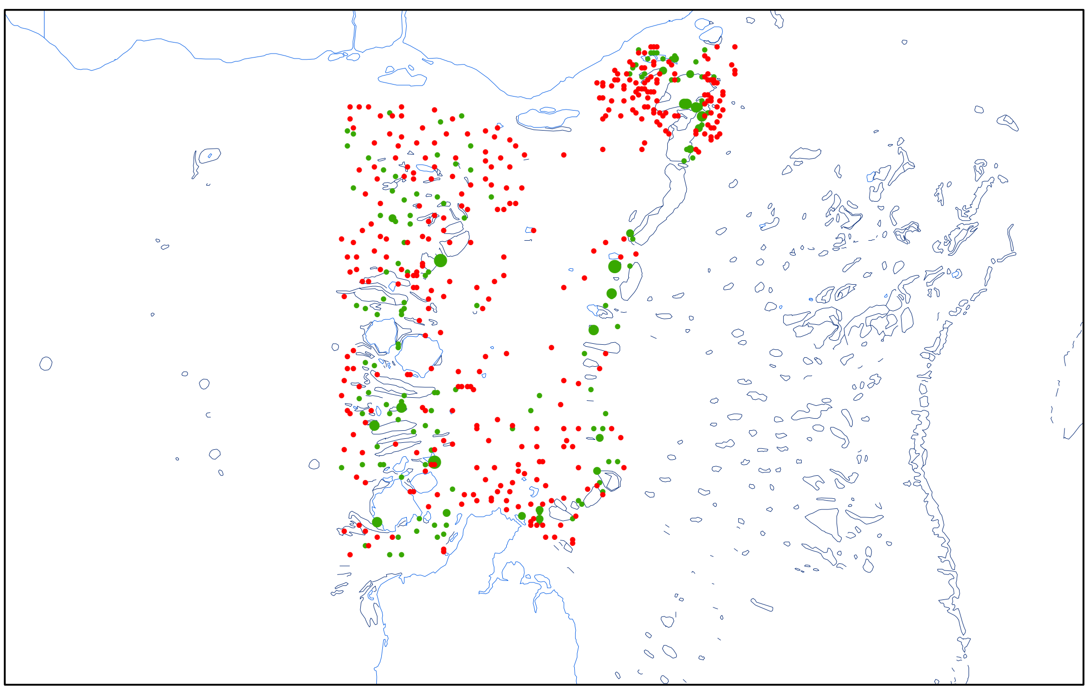


Figure 2.4-2 Maps of substratum type and epibenthic gardens from biological survey sites: Substrate classification: 1= >50% hard substrate, 2= 10-50% hard substrate, 3= rubble (<10% hard & >15% rubble), 4= sand (muddy sand & sand), 5= mud (mud & sandy mud); Epibenthos classification: 1= dense fauna, 2= sparse fauna, 3= very sparse fauna, 4= no fauna (sources, see section 2.1.1).



Megabenthos cover %

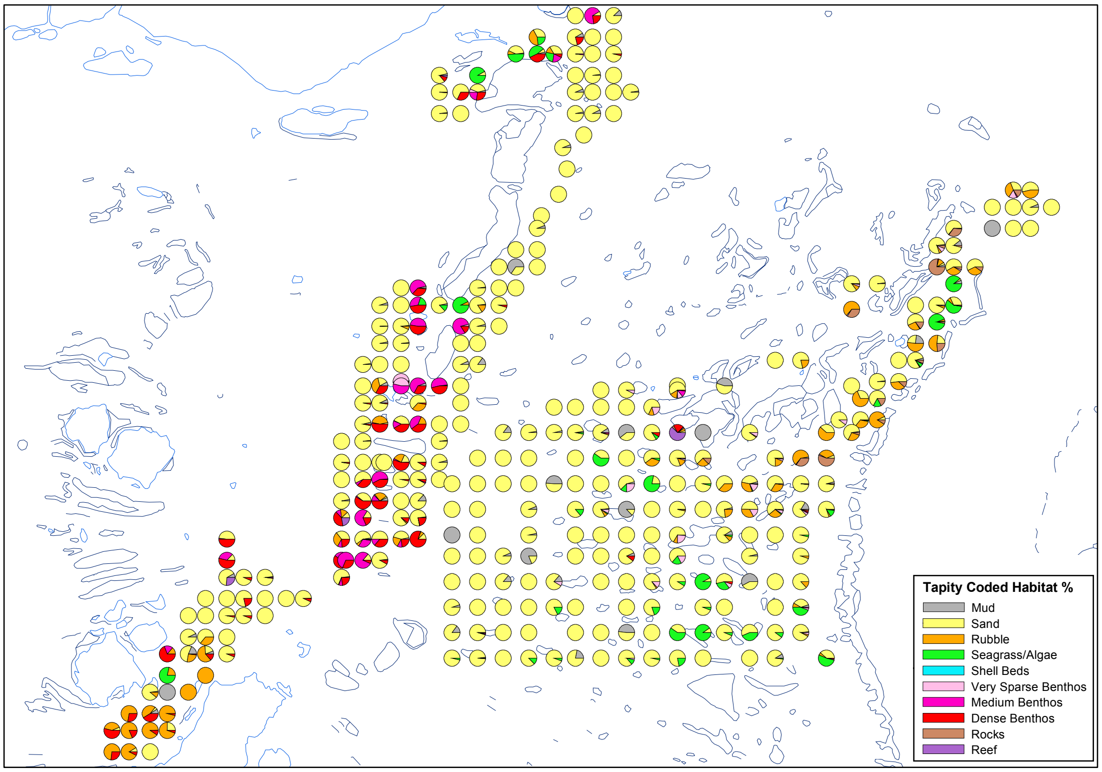
0 | 1 - 2 | 3 - 5 | 6 - 10 | 11 - 68



Hard Coral %

0 | 1 - 6 | 7 - 12 | 13 - 25 | 26 - 50

Figure 2.4-3 Maps of percent cover of epibenthos gardens and hard coral gardens of 500 m transects, from biological survey sites (sources, see section 2.1.1).



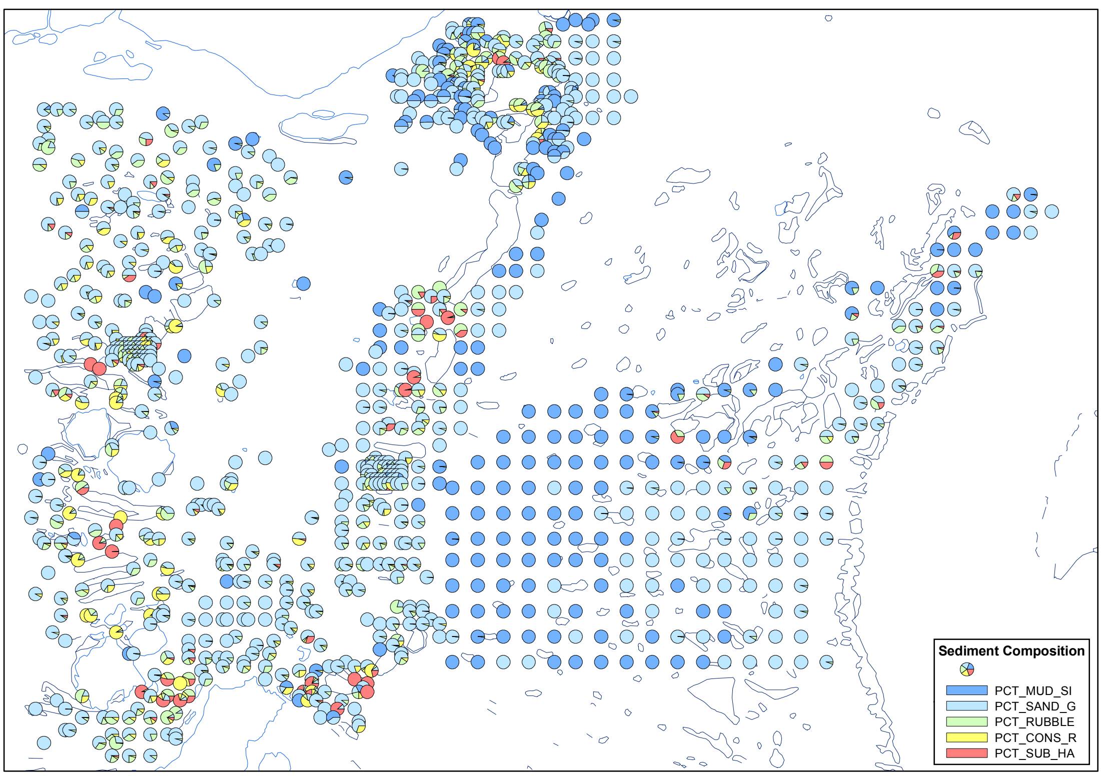
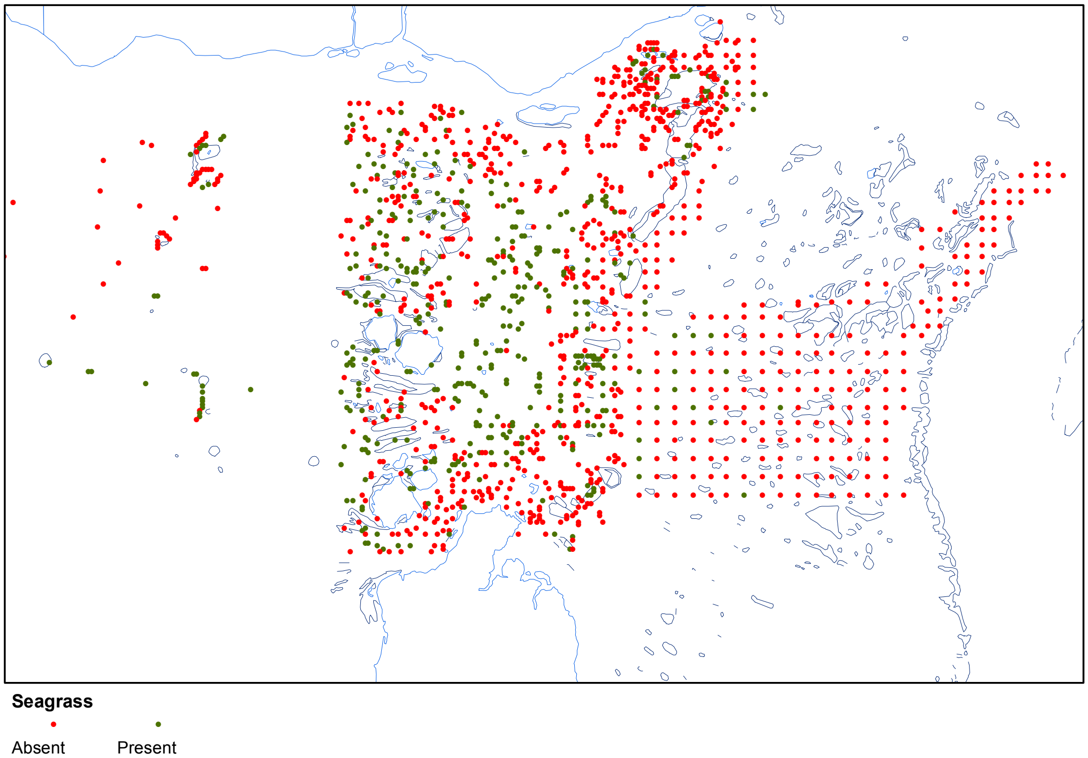


Figure 2.4-4 Maps of summary underway 'tappity' coding of seabed habitats and observed substratum composition, as a percentage of 500 m transects, from biological survey sites (sources, see section 2.1.1).



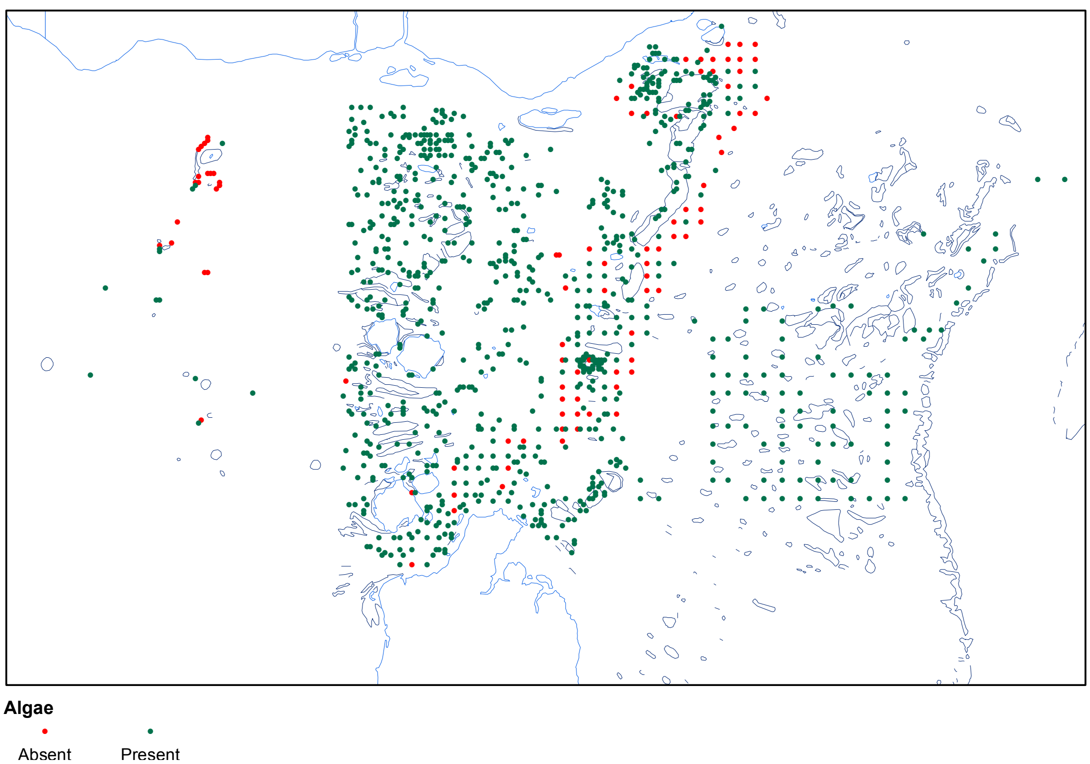


Figure 2.4-5 Maps of seagrass presence/absence and algal presence/absence from biological survey sites (sources, see section 2.1.1).

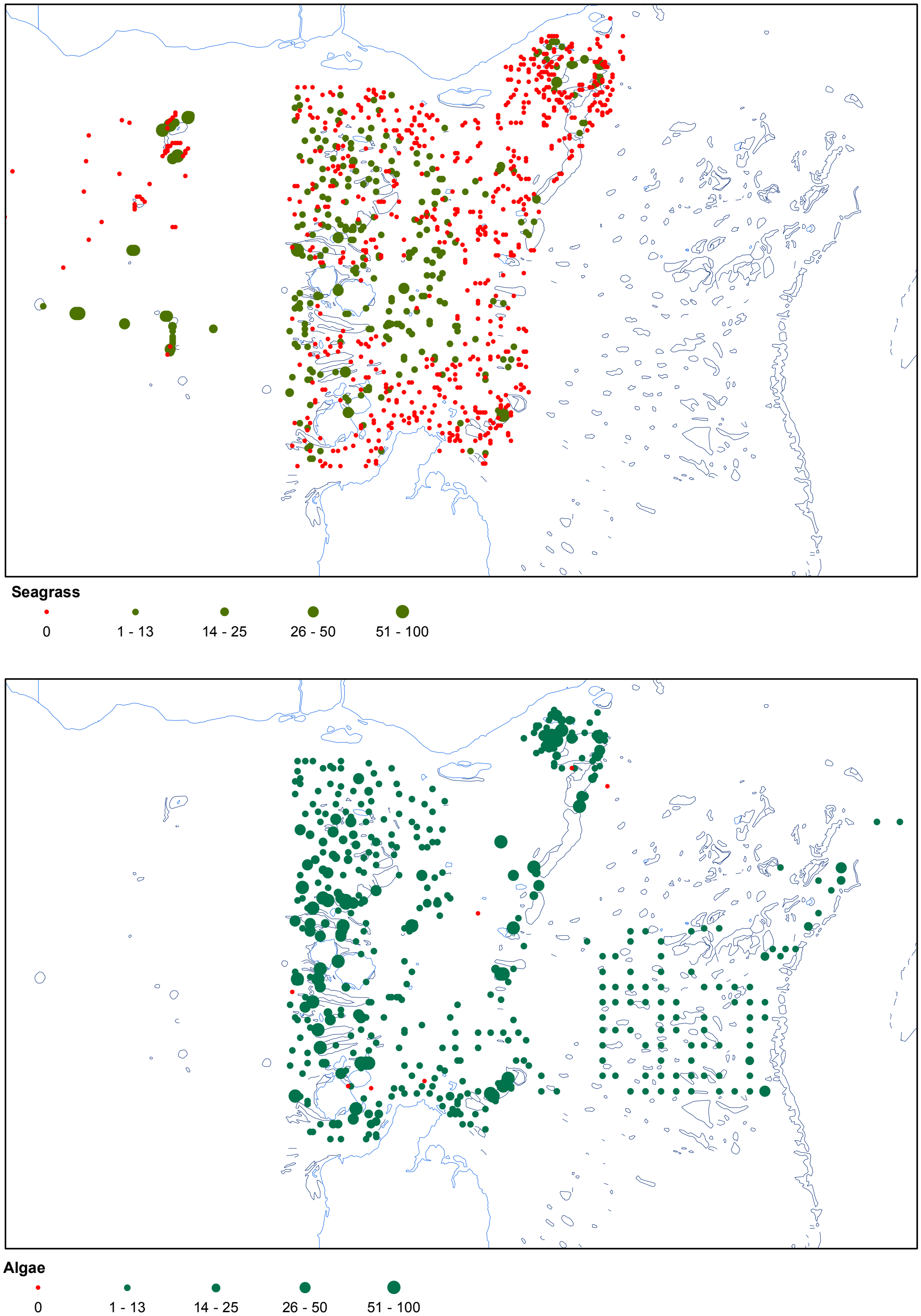


Figure 2.4-6 Maps of seagrass percent cover and algal percent cover of 500 m transects, from biological survey sites (sources, see section 2.1.1).

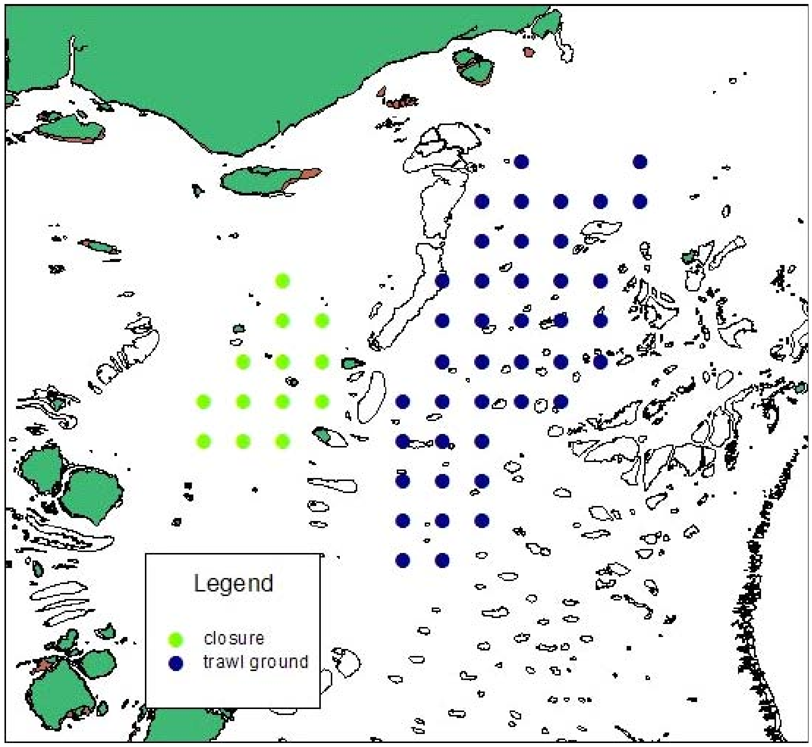


Figure 2.4-7 Map of central Torres Strait showing the sites sampled by the *FV Jacqueline-D* during the "Effects-of-Trawling" surveys of 1985-6 (sources, see section 2.1.1).

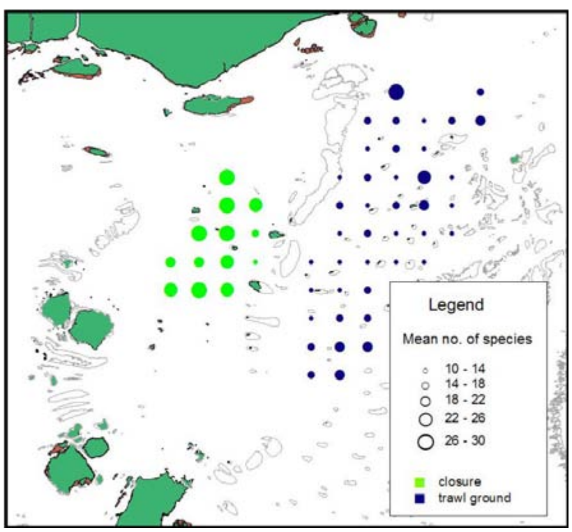


Figure 2.4-8 Map of central Torres Strait showing the mean number of fish species sampled by the *FV Jacqueline-D* during the "Effects-of-Trawling" surveys of 1985-6 (sources, see section 2.1.1).

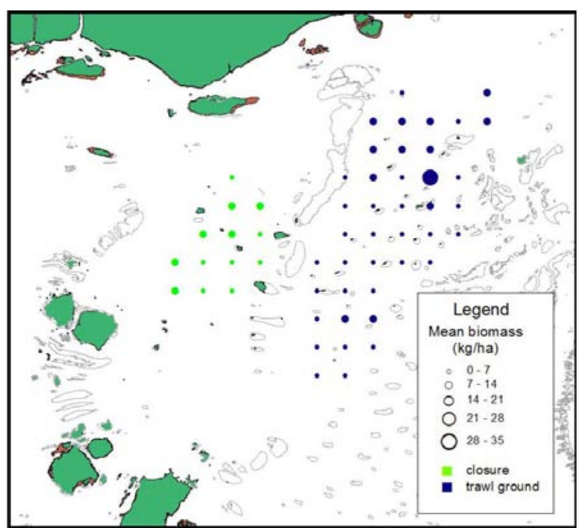


Figure 2.4-9 Map of central Torres Strait showing the mean catch rate (kg/ha) of fish bycatch sampled by the *FV Jacqueline-D* during the "Effects-of-Trawling" surveys of 1985-6 (sources, see section 2.1.1).

2.5. ACOUSTIC SURVEY DATA

Underwater acoustic sound pulses emitted from an echosounder, then reflected by the seafloor and collected by the seabed classification instrument RoxAnn provided acoustic information about the seabed in Torres Strait. This section outlines the acoustic information available to the project, and the analyses used to produce a biological and substrate stratification of the Torres Strait based on this acoustic information.

2.5.1. Data Processing

Seabed acoustic roughness and hardness data was collected continuously along with depth information (approximately twice a second), on seventeen cruises which either conducted research in Torres Strait or transited through the area collecting data. Over the 10 cruises a total of 1633625 individual RoxAnn samples were recorded. The earliest cruise was conducted on the Southern Surveyor in 1993 (ss0193), and the most recent being the 2002 annual Torres Strait rock lobster abundance survey. The vessel track for RoxAnn acoustic sampling is shown in Figure 2.5-1 below.

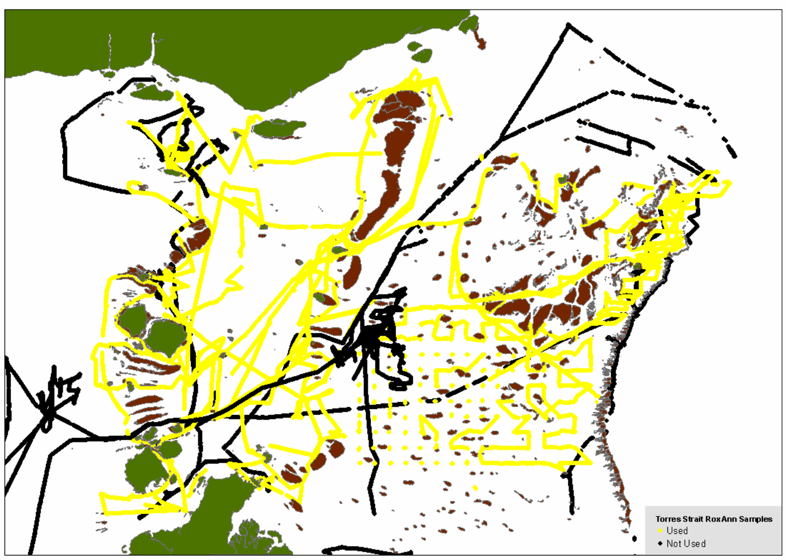


Figure 2.5-1 The vessel track for RoxAnn acoustic sampling in Torres Strait (samples used shown in yellow).

Given that there was no dedicated acoustics expertise assigned to the collection of the data during most of these cruises, and the acoustic sampling was carried out as an adjunct to existing research programs, data quality was an issue on some of these cruises. With the stratification analysis of this data being conducted post-hoc, careful data quality checking was required at a course "cruise" level to investigate systematically biased data (e.g. due to changing echosounder or RoxAnn settings over a cruise and between cruises). Initial "cruise" level data quality checks precluded some sections of acoustic data completely from further analysis. In fact six of the ten cruises (in their entirety) were not used in further analysis, as the data collected was of such poor quality, or the cruise provided little useful data. However the remaining four cruises with usable data accounted for 87% of the total available RoxAnn data and had a broad geographic coverage (see figure above, highlighting usable RoxAnn data in yellow, with unusable data in black). Therefore the "cruise level" data quality filtering of acoustic information did not significantly compromise subsequent analyses and the stratification objectives.

Even with the broad geographic coverage of usable acoustic data in Torres Strait, it should be noted that there are some information gaps. For example, there are gaps in the geographic coverage of RoxAnn data in western and north-eastern Torres Strait. These areas should be targeted for any future acoustic sampling.

The total number of RoxAnn samples collected for each of the four cruises is shown in Table 2.5-1 below. Three cruises; the 1996 Pipeline, 1997 Chevron, and 2002 Beche de Mer cruises account for the vast majority (and similar proportions) of the total available data with only a minor contribution from the earlier 1993 Rock Lobster cruise.

Table 2.5-1 Total RoxAnn samples collected.

Cruise	RoxAnn Samples
1993 – Torres Strait Rock Lobster	53381
1996 – Torres Strait Pipeline	362858
1997 – Chevron	486623
2002 – Torres Strait Beche de Mer	512211

Two cruises; the 1993 Rock Lobster, and 2002 Beche de Mer cruises covered the central area of Torres Strait from Thursday Island in the south-west to the Warrior Reefs in the north-east. The 1997 Chevron cruise covered central-eastern Torres Strait, again around the Warrior Reefs, Sassie Island and south-west to Thursday Island. While there was some data available for eastern Torres Strait from the 2002 Beche de Mer cruise (especially in the north-east); the 1996 Pipeline cruise accounted for the vast majority of data, covering the 10000 km^2 area from Kagar Reef in the south-west to Don Cay in the north-east. The geographic coverage for each of the four cruises is shown in Figure 2.5-2 below.

Careful filtering and interpretation (Pitcher et al., 1999; Kloser et al., 2001) was required to remove erroneous data (e.g. due to the effect of poor weather influencing the acoustic information, and depth outliers due to noise spikes in readings) within each cruise prior to further analysis; and take into account possible influencing relationships within the data if possible (e.g. bias effects of depth and vessel speed).

The RoxAnn instrument provided relative information on acoustic roughness and hardness. These measures (a unitless numerical index between 0 – 2048) are not absolute or easily calibratable. Also, the RoxAnn readings may be influenced by a number of variables, most importantly the echo sounder system used, including transceiver settings and transducer type. Though, once calibrated over known seabed type and with appropriate "ground truthing", these indices can provide continuous classification of basic physical seabed types for a cruise. As different echosounders and echosounder settings were used in the collection of RoxAnn data, and there was no way to calibrate the system between cruises, the data from each cruise was treated separately in subsequent analyses. On completion of these separate analyses the acoustic information was stratified into the same biological and substrate classes.

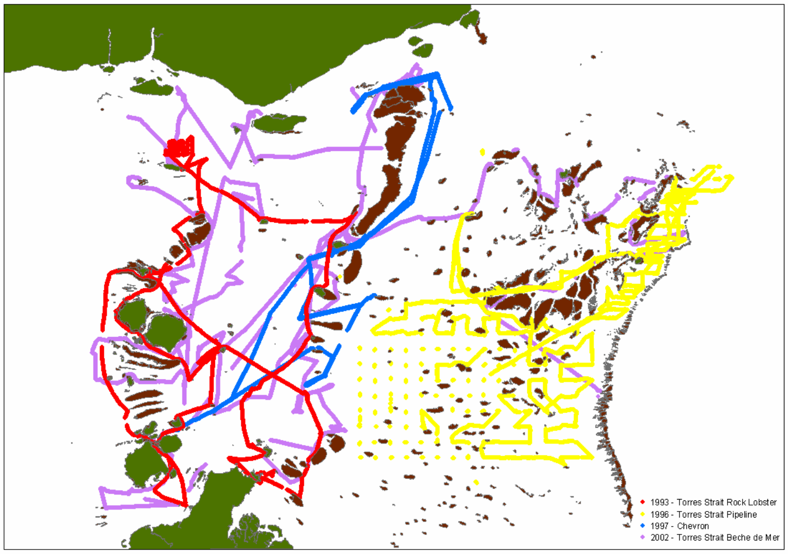


Figure 2.5-2 The geographic coverage for each of the four RoxAnn acoustic sampling cruises.

Ground truth information was available, at the scale of 0.001 (or approximately 110 m^2) decimal degrees, on the biological nature of the seabed, as well as the type of seabed substrate encountered. The biological ground truth information was categorized, on the density of biological organisms, into four classes of seabed; no benthos, very sparse benthos, sparse benthos, and dense benthos. While, the substrate ground truth information was categorized into five classes of seabed; mud (or silt), sand, rubble, sparse rock, and dense rock.

The RoxAnn acoustic depth, roughness and hardness information was aggregated on a similar grid scale (used throughout this project) of 0.01 decimal degrees (or approximately 1.1 km^2). The acoustic information within each 0.01 decimal degree grid cell was aggregated and five basic statistics generated. Statistics of minimum, maximum, median, mean and standard deviation were produced for each acoustic measure (depth, roughness, and hardness), for each of the four cruises.

A diagrammatic summary of the data used to predict seabed strata is shown in Figure 2.5-3 below.

Maps of the raw acoustic index statistics available at the 0.01 degree grid scale are shown in Figure 2.5-4 below; mean depth is rendered in blue (min = 3 m, maximum = 150 m), mean acoustic roughness is rendered in red (minimum = 83, maximum = 2379), and mean acoustic hardness is rendered in green (minimum = 57, maximum = 1412). Though the information is presented here on the same scale it is important to note that the roughness and hardness statistics for the four different cruises are not directly comparable.

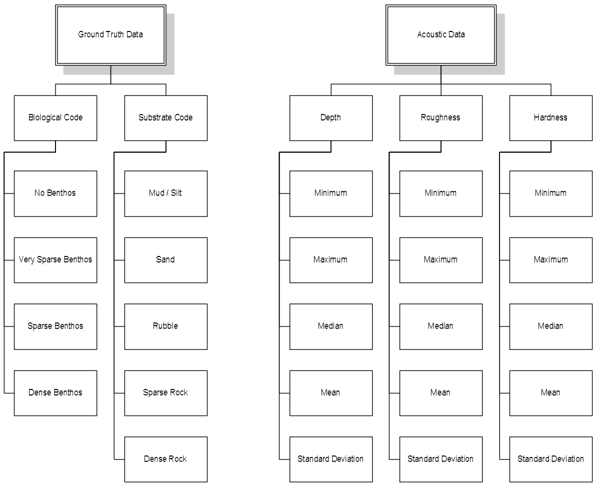


Figure 2.5-3: Schematic of Ground Truth classes and RoxAnn Acoustic predictor variables available (at the 0.01 degree grid scale), and used in the stratification analysis for each cruise.

2.5.2. Classification of Acoustics Data

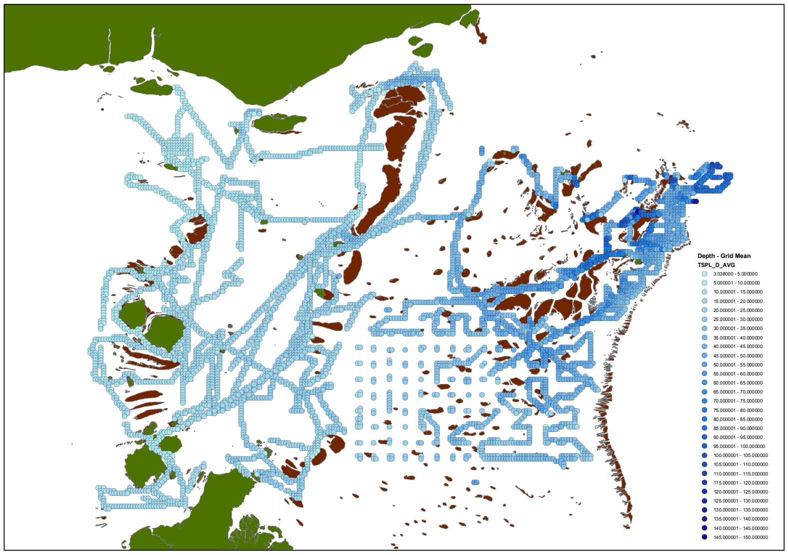
To predict seabed stratum type from the grided acoustic information, the acoustic data were first merged by location with the benthic biological code and substrate code ground truth data (see data schematic Figure 2.5-3 above). This matched up areas where both acoustic and ground truth data were available. For each of the four cruises, the acoustic grided depth, roughness and hardness and roughness statistics were then classified, based on the benthic biological code and, in a separate analysis, substrate code ground truth data, using a linear discriminant function analysis that maximally separated the biological and substrate types based on the acoustic statistics (i.e. a total of eight discriminant function analyses were conducted). The linear discriminant function analysis, which was based on the four point biological and five point substrate classification systems, indicated that the acoustic data were able to discriminate satisfactorily among the seabed types.

The performance of the linear discriminant analysis was assessed and compared between cruises using the cross-validated error rate. The cross-validated error rate was calculated using the leave-one-out method of evaluation and gives the overall posterior probability of strata membership. The error rate performance is shown in the Table 2.5-2 below for each cruise (and graphically in Figure 2.5-5 below). Overall it was noted that the error rates were consistent across the cruises analysed; ranging from 13.6% to 33.9% for biological strata and from 21.1% to 41.6% for substrate strata.

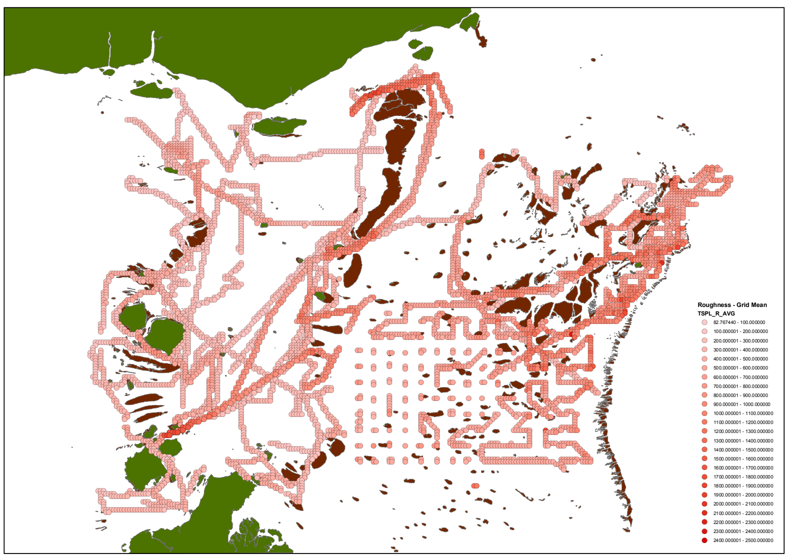
Table 2.5-2 Cross-validated classification error rates by cruise (Note - expressed as % incorrectly classified).

Cruise (Individual DF Analysis)	Overall Error Rate	
	Biological	**Substrate**
1993 - Torres Strait Rock Lobster	35.3%	41.6%
1996 - Torres Strait Pipeline	13.6%	35.3%
1997 - Chevron	28.5%	21.1%
2002 - Torres Strait Beche de Mer	33.9%	40.4%

Mean Depth



Mean Roughness



Mean Hardness

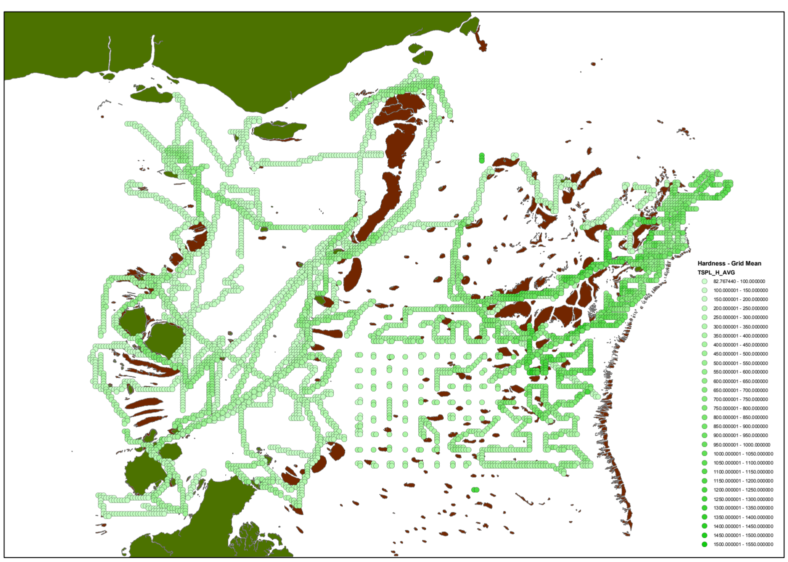


Figure 2.5-4 Maps of acoustic index statistics available at the 0.01 degree grid scale.

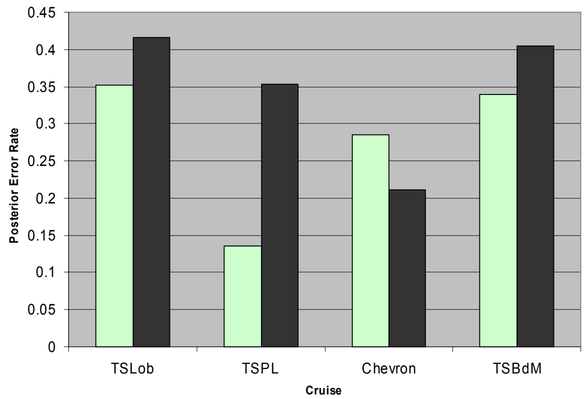


Figure 2.5-5 Discriminant function classification error rate performance for Sub_Code (■) and Biocode (□), by cruise.

An indication of which classes were the best described by the acoustic information, is given in Figure 2.5-6 below, comparing the mean classification performance across the cruises for each seabed strata. Overall, biological type strata were better described by the acoustics than substrate type strata were. This result was due to the Torres Strait Pipeline cruise biological classification performing significantly better than any of the other seven discriminant function analyses. Of the biological strata, the very sparse benthos and dense benthos classes were poorly described by the discriminant function; while sparse benthos was the best described with an error rate of 14.0%. Of the substrate strata, the sand and rubble substrate classes had poor error rates compared with the sparse rock class which was the best described class with an error rate of 11.1%.

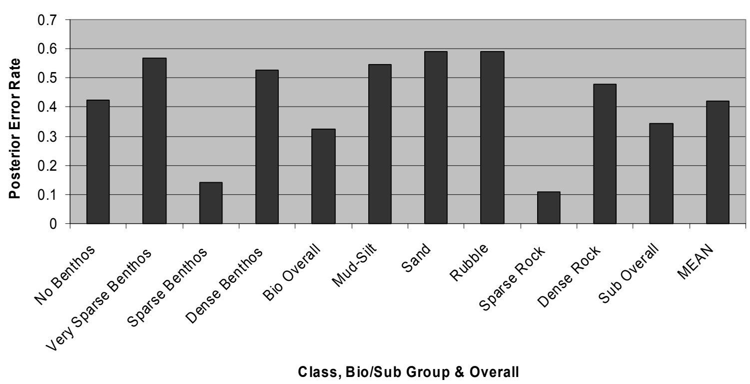


Figure 2.5-6 Mean discriminant function error rates by classification stratum class.

Table 2.5-3 and Table 2.5-4 (below) show the average classification characteristics for the linear discriminant functions across the four cruises. The tables show how the data was assigned from one classification strata into another. The values in the table are expressed as a percentage of the total data assigned to each particular stratum. The proportions of correctly classified data are shown on the diagonal of the table – where one class of data is correctly classified into that same class by the discriminant function. These classification tables highlight the performance of individual classes in the discriminant function analysis. It shows the classes where incorrectly classified data is assigned to, and particularly, where data is not assigned to adjoining strata, but, to classes that are dissimilar.

Erroneous classification results from discriminant function analysis may take either of two forms: The data may be wrongly assigned from the correct class into another incorrect class, this is knows as “miss-classification”; or the data may assigned from an incorrect class into the correct class, this is known as “false-classification”. The proportions of data which have been miss-classified are shown as the first (%) value in the table and may be read across the rows of the table (totalling 100%); while the proportions of data which have been falsely-classified are shown as the second (%) value in the table and may be read down the columns of the table (totalling 100%).

Table 2.5-3 Detailed biological classification table. Shows average proportion of data classified from one strata into another across all cruises, and highlights whether data was miss-classified (reading % totals across table rows) or falsely classified (reading % totals down table columns).

Miss Classification		**To**			
	False Classification	**No Benthos**	**V Sparse Benthos**	**Sparse Benthos**	**Dense Benthos**
From	**No Benthos**	50.3% 79.4%	0.0% 0.0%	47.9% 6.2%	0.0% 0.0%
	V Sparse Benthos	1.3% 3.9%	35.8% 66.5%	51.3% 11.9%	11.6% 5.6%
	Sparse Benthos	1.1% 10.6%	4.6% 28.9%	82.9% 65.3%	11.5% 18.9%
	Dense Benthos	0.9% 6.0%	1.1% 4.6%	30.8% 16.6%	67.3% 75.5%

Table 2.5-4 Detailed substrate classification table. Shows average proportion of data classified from one strata into another across all cruises, and highlights whether data was miss-classified (reading % totals across table rows) or falsely classified (reading % totals down table columns).

Miss Classification		**To**				
	False Classification	**Mud / Silt**	**Sand**	**Rubble**	**Sparse Rock**	**Dense Rock**
From	**Mud / Silt**	40.4% 76.1%	16.0% 9.4%	11.4% 8.3%	32.2% 3.7%	0.0% 0.0%
	Sand	1.5% 6.5%	45.9% 61.5%	5.5% 9.1%	47.0% 12.4%	0.0% 0.0%
	Rubble	1.0% 4.3%	0.0% 0.0%	34.7% 55.8%	61.8% 16.0%	2.5% 1.8%
	Sparse Rock	1.0% 13.1%	6.1% 25.7%	4.4% 22.8%	71.9% 59.7%	16.7% 37.4%
	Dense Rock	0.0% 0.0%	2.1% 3.3%	2.1% 4.1%	25.8% 8.3%	70.1% 60.8%

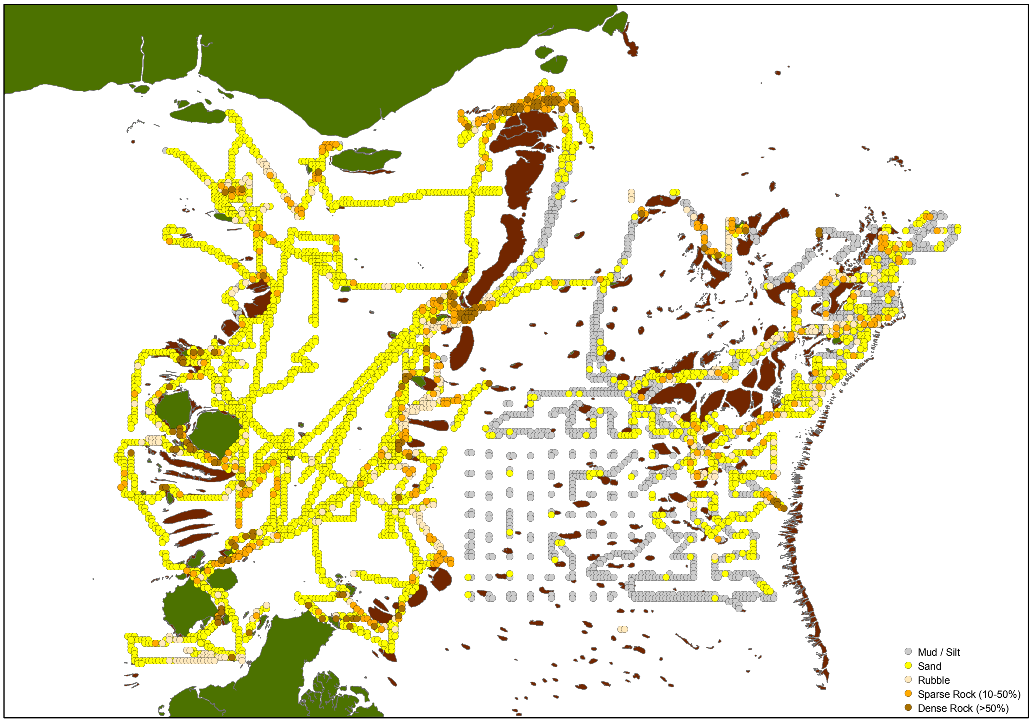
Classification results from the biological stratification of Torres Strait based on acoustic information (see Table 2.5-3 above) show that in terms of miss-classification performance the “No Benthos” and “Very Sparse Benthos” classes are poorly described, with 49.7% of the data from the “No Benthos” class being miss-classified as “Sparse Benthos”, and more “Very Sparse Benthos” being miss-classified as “Sparse Benthos” (51.3%) than “Very Sparse Benthos”. The “Sparse Benthos” and “Dense Benthos” classes have acceptable miss-classification performance with the majority of data being classified correctly (82.9% and 67.3% respectively) and a large proportion of miss-classified data falling into adjacent strata. When looking at false-classification performance all of the biological classes performed well with correct classification rates at 65% or above, with other data being falsely classified into adjacent strata. The classification table also highlights that the “Sparse Benthos” class is by far the most likely class resulting from any type of incorrect classification. The classification table for substrate stratification (see Table 2.5-4 above) shows that both of the “Rock” classes (Sparse and Dense) are well described in terms of miss-classification error with less than 30% of data being miss-classified. The other three substrate classes do not have good miss-classification error rates with less than 50% of data correctly classified. Of these three classes most of the data was miss-classified as “Sparse Rock”; with 32.2% of “Mud/Silt”, 47.0% of “Sand”, and 61.8% of Rubble miss-classified into the “Sparse Rock” strata. As was the case with the biological classification, the substrate classes are better performed in terms of false-classification, with all classes having more than 55% of their data correctly assigned.

For each cruise the acoustic depth, roughness and hardness statistics were classified using the discriminant function to assign seabed biological and seabed substrate stratum types to the remaining unclassified. This procedure classified the seabed in areas where only acoustic data existed, producing a map of seabed strata for the full length of each cruise track. With the predicted seabed strata as the basis, the analyses conducted for each of the four cruises were fused to produce a single seabed type stratification for all Torres Strait cruise tracks based on acoustic information.

Maps of the resulting biological and substrate stratification of the Torres Strait based on acoustic information are shown in Figure 2.5-7 below. The substrate stratification map highlights large areas of sand in central western Torres Strait with the presence of rock and rubble area in the channels between reefs (such as to the north and south of Warrior Reef). There is also an extensive area of south eastern Torres Strait which is predominated by mud and silt substrate graduating to sand, rubble and rock substrates in the north east (also through many inter-reefal channels). The biological stratification map outlines large areas of very sparse benthos in central western Torres Strait with increasing density in channels between the Warrior Reef complex. There is a large section of south eastern Torres Strait which has been classified as having no benthos, which corresponds to the large mud strata shown in the substrate map.

The acoustic stratification is consistent with the results of analyses based on other physical covariates. This confirms that acoustic information is a useful surrogate for seabed biological and substrate types and should be a data collection objective for any forthcoming fieldwork.

Substratum



Epi-benthos

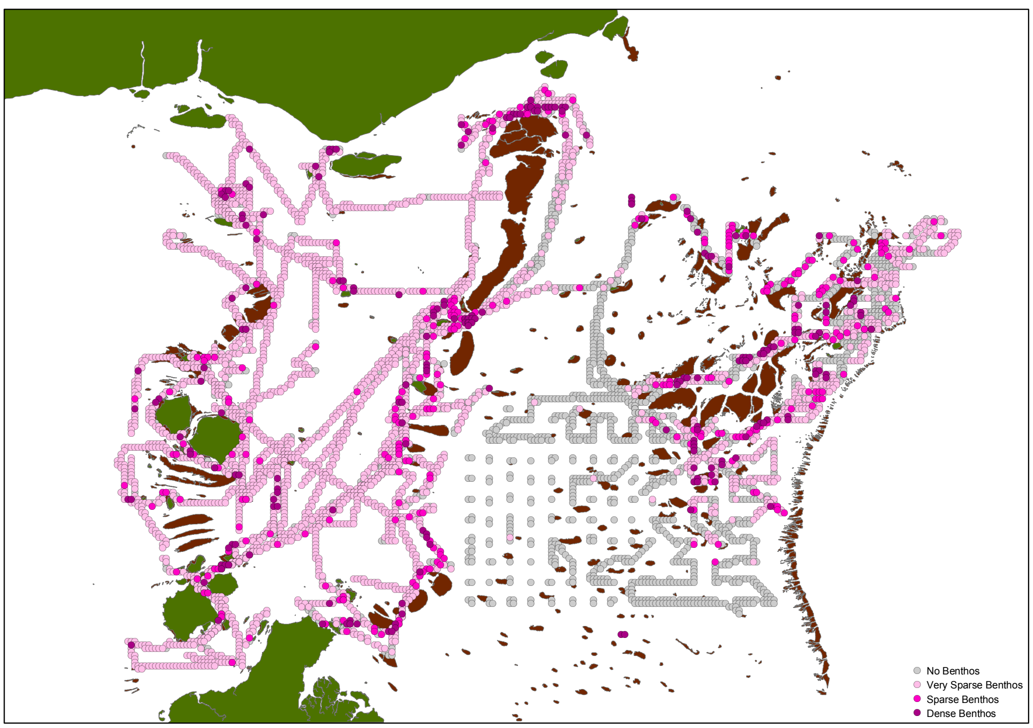


Figure 2.5-7 Maps of classified seabed substrate and biological strata for Torres Strait based on acoustic information.

2.6. SPATIAL AUTO-CORRELATION

A major consideration in determining the density of sampling in seabed surveys is the distance over which the seabed assemblages remain homogeneous, within any defined habitat type. While it is inefficient to place sample sites too close together, information on seabed patterns will be inadequately sampled if sites are too far apart. Further, the same process means that spatial prediction of assemblages cannot be applied reliably at great distances from sampled sites, even if physical environment variables are measured and taken into account. This process may be quantified by several indices; one is the spatial auto-correlation distance.

Analyses of spatial auto-correlation distance were conducted to establish over which geographic distance the similarity between the species composition per site is maintained to some degree. This distance then would give an indication of the minimum inter-sample distance needed to represent the biota adequately, as well as, the maximum distance over which prediction would be feasible. Two measures were use to estimate the order of the auto-correlation distances for the Torres Strait datasets.

The semi-variogram was used for seabed physical data, such as depth and sediment attributes, and for the biological data, such as Bio_Code and seagrass presence/absence and algae p/a the Bray-Curtis dissimilarity metric was used as a function of the inter-site distances. The between station Bray-Curtis dissimilarities were plotted against the geographic distance between the corresponding stations. If correlation had been used as the metric, such plots are called ‘correlograms’ — with the Bray-Curtis dissimilarity metric, these plots may be termed ‘Bray-Curtis-ogram’s.

2.6.1. Physical Co-Variates

Most of the semi-variograms of the seabed physical data, (percentage mud, sand, gravel and rock, degree of grain sorting and grain size) indicated that the similarity between stations decreased rapidly as the distance between the stations increased. Beyond an inter-station distance of between about 0.05 and 0.1 decimal degrees (approx 6–11 km) any consistency between the stations had degenerated (Figure 2.6-1). However, for sediment carbonate content and depth the drop in similarity between stations with increased distance was more linear.

2.6.2. Biological Survey Data

The Bray-Curtis-ograms of the low-level and medium level biological survey datasets indicates that the dissimilarity between stations increased rapidly the further apart stations were (Figure 2.6-2 and Figure 2.6-3) and after a distance greater than about 0.05–0.1 decimal degrees (about 5–11 km) the coherence is lost. Although directional relationships were not investigated, it is possible that this may be slightly better for some directions and worse for others. This distance needs to be considered in the context of the whole region. Stratification of the region based on other (physico-chemical) co-variates can lead to better local results, as was observed in a similar nearby area of the Great Barrier Reef where stratified species assemblage data did have Bray-Curtis ranges of up to perhaps 20 km within physical strata (Pitcher et al 2002).

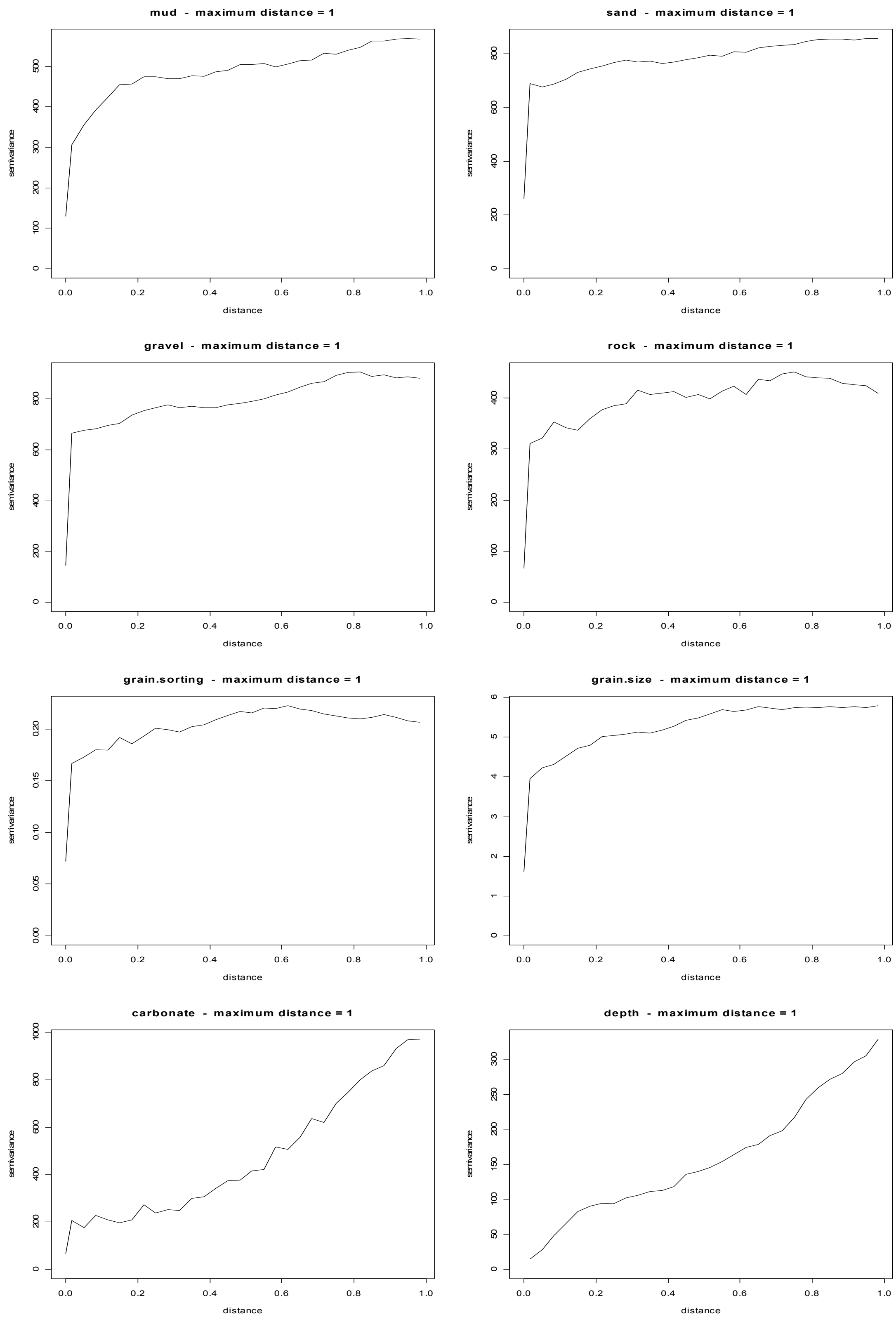


Figure 2.6-1 Semivariograms of the seabed physical data (percent of mud, sand, gravel and rock, degree of sorting, grain size, carbonate and depth) showing spatial autocorrelation between the survey sites, with distance (in decimal degrees).

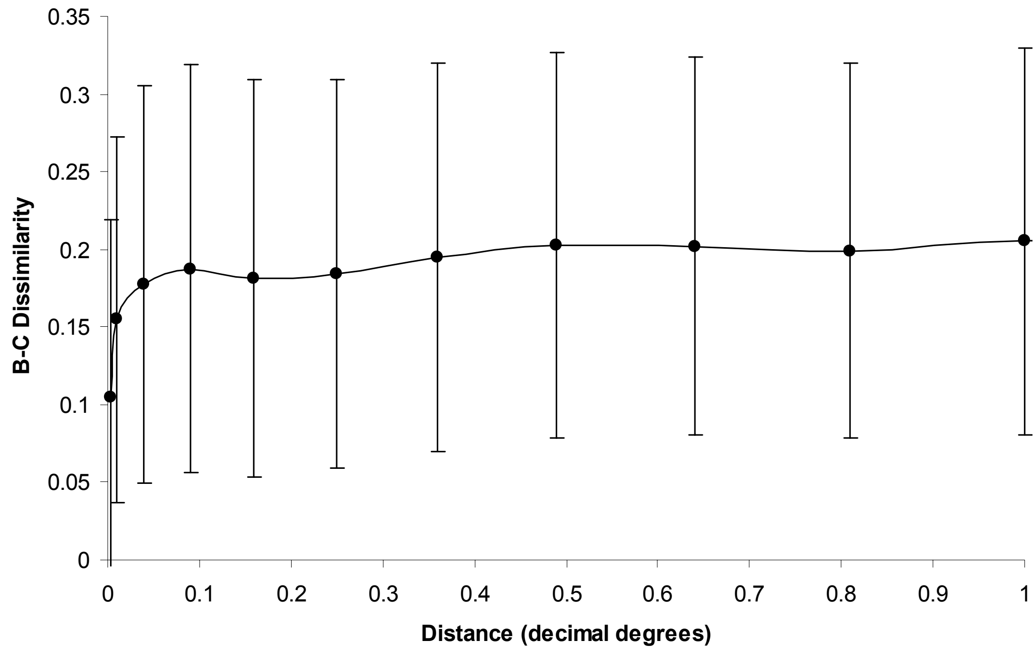


Figure 2.6-2 Bray-Curtis-ogram of the 'low-level' dataset (Table 2.4-1A), showing spatial autocorrelation for biological survey sites.

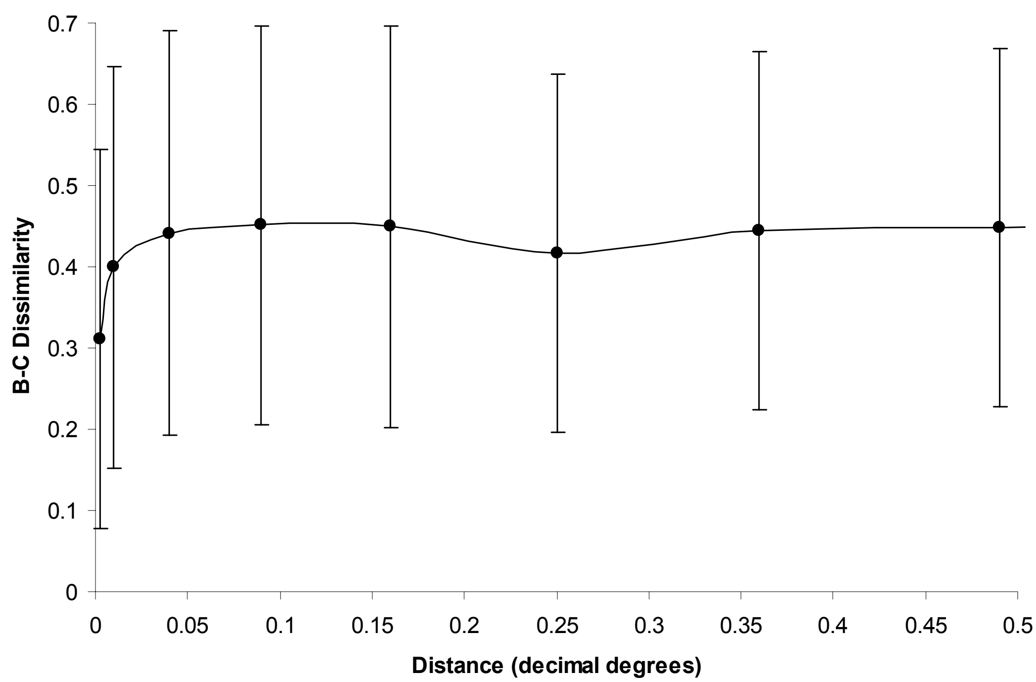


Figure 2.6-3 Bray-Curtis-ogram of the 'medium-level' dataset (Table 2.4-1B), showing spatial autocorrelation for biological survey sites. Note different x-axis scale cf Figure 2.6-2.

In summary, it is clear that the omnidirectional large scale distance that still carries some level of auto-correlation is in the order of ~5 km generally, and even within habitat strata, the spatial auto-correlation distance may extend to only ~10 – 20 km maximum. Consequently, if we do not have reliable co-variates with proven predictive capacity and relatively large homogeneous (geographic) areas, the inter-sample distance would need to be in the order of 10 km to enable some dependable spatial prediction capability. Spatial prediction mapping within these distance ranges is credible, but mapping seabed assemblages by extrapolating far beyond these distances would be highly unreliable and uncertain.

2.7. STATISTICAL CHARACTERISATION & BIO-PHYSICAL MODELING

Statistical approaches (including clustering) were used to characterise the mixtures of habitat facies in the biological survey datasets, and the relationships between these facies or clusters and the collated physical covariates was also examined. Based on any biophysical relationships and the physical covariate values at the ~50¸000 grid cells, the predicted cluster membership of each grid cell was mapped for a full coverage of Torres Strait. The biophysical relationships also indicate the relative importance of each physical covariate with respect to patterns in the biological data and provide weightings for each covariate in developing the stratification, which may be regarded as a biophysical characterisation of Torres Strait. These methods follow a similar approach as in Pitcher et al (2002).

In the case of the 'low-level' bio-survey dataset (Table 2.4-1A), a simple biophysical model was developed separately for each attribute (Sub_Code, Bio_Code, and Seagrass and Algae presence/absence), using the statistical method Linear Discriminant Analysis (LDA), which provides the best set of linear functions on the physical covariates that allocate the survey sites to the specified categories of the bio-survey attributes. The reliability of these functions (uncertainty of prediction) was also estimated by examining jack-knifed cross-validated error rates. Cross validation treats n-1 out of n training observations as a training set. It determines the discriminant functions based on these n-1 observations and then applies them to classify the one observation left out. This is done for each of the n training observations. The misclassification rate for each group is the proportion of sample observations in that group that are misclassified. This method achieves a nearly unbiased estimate but with a relatively large variance. The LDA functions were then applied to the Torres Strait wide coverage of physical covariates to map the estimated distribution of the facies of each bio-survey attribute for Torres Strait. The LDA also identifies which of the physical covariates were of use in allocating the sites to categories. Additionally, a simple cluster analysis was applied to characterise the combined attributes of the 'low-level' bio-survey dataset, and LDA was again used to develop a simple biophysical model for the clusters, provide cross-validated error rates, and to map these for an estimate of the full coverage for Torres Strait.

In the case of the 'medium-level' dataset (Table 2.4-1B), cluster analysis was applied to characterise the mixtures of these attributes, and the relationship between these clusters and the collated physical covariates, and their uncertainty, was examined and mapped using LDA, as above.

A similar approach was followed with the Torres Strait effects of trawling dataset that included detailed species-biomass distribution and abundance information from samples of seabed fish from research trawls. Cluster analysis was applied to these data to provide a number of relatively homogenous mixtures of fish species, and, as above, LDA was used to develop biophysical models and map the fish assemblages to characterise the vicinity of the sampling.

2.7.1. 'Low-level' bio-survey dataset

2.7.1.1. Substrate Classification (SUB_CODE):

The substrate classification (SUB_CODE) data, widely available from Torres Strait seabed surveys, was defined as:

1 = >50% hard substratum,
2 = 10-50% hard substratum,
3 = rubble (<10% hard & >15% rubble),
4 = sand (muddy sand & sand),
5 = mud (mud & sandy mud)

The results from forward stepwise LDA on SUB_CODE categories alone showed the following. Based on the physical covariate data that was associated with the survey sites, categories 1 & 2, and 2 & 3 were most similar and categories 4 & 5 were quite different to the others, with category 3 intermediate (Table 2.7-1). The most important covariates for discriminating SUB_CODE categories were seabed current stress, CARS Silicate standard deviation, CARS Salinity average, and CARS Oxygen sd (Table 2.7-2). Sediment covariates were moderately important. The overall jack-knifed discrimination accuracy was 51% correct and ranged from 27% correct for category 2 to 80% correct for category 5 (Table 2.7-3) — in most cases the LDA confuses the most similar (and adjacent) category, so performs somewhat better than the raw error rates indicate.

The mapping of predicted SUB_CODE over Torres Strait (Figure 2.7-1) closely reflects the available data (Figure 2.4-2) and can be considered a reasonably reliable representation of the distribution of the gross substratum types.

Table 2.7-1. Between groups F-matrix for the physical covariate data associated with the SUB_CODE categories at sites (df = 9, 836). Small F-values indicate that the covariates of respective categories are more similar; large values indicate that the covariates of categories are more different.

	SUB_CODE Category				
	1	2	3	4	5
1	0.000				
2	3.800	0.000			
3	9.071	7.443	0.000		
4	21.225	25.293	23.337	0.000	
5	36.898	43.053	46.391	25.475	0.000

Table 2.7-2. F-values for physical covariates assessed by LDA for SUB_CODE. Increasing F-values indicate covariates that were more important for discriminating SUB_CODE categories. Covariates in the left columns were included in the final model; those on the right were excluded.

Covariate included in model		**Covariate excluded from model**	
Variable	F-value	Variable	F-value
M_BSTRESS	69.26	SW_CHLA_SD	3.39
CARS_SI_SD	21.18	DEM_SLOPE	2.99
CARS_S_AV	12.03	AGSO_DEM	2.68
CARS_O2_SD	11.23	SW_D_B_IRRAD	2.39
auDB_ROCK	8.29	SW_K490_SD	2.30
auDB_MUD	7.75	SW_CHLA_Y_AV	2.28
CARS_PO4_S	5.24	SW_K490_Y_AV	1.99
auDB_GRNSZ	4.65	DEM_ASPEC	1.55
auDB_GRAVEL	4.49	CARS_T_SD	1.55
		auDB_SAND	1.13
		auDB_CRBNT	1.12
		SW_K490_Y_SD	1.12
		CARS_NO3_S	0.97
		CARS_S_SD	0.94
		CARS_O2_AV	0.68
		SW_CHLA_Y_SD	0.68
		auDB_GRNSRT	0.53
		CARS_NO3_A	0.46
		CARS_SI_AV	0.46
		CARS_T_AV	0.45
		CARS_PO4_A	0.17

Table 2.7-3. Jackknifed classification matrix for discriminating SUB_CODE categories on physical covariates. Cases in row categories classified into columns. % correct shows number of cases in each row classified into the correct column. Diagnostic statistics: Wilks' lambda=0.439, approx.F=21.421, df=36, 3134.

	SUB_CODE Category					
	1	**2**	**3**	**4**	**5**	**%correct**
1	29	12	4	2	3	**58**
2	45	43	32	25	12	**27**
3	35	27	79	43	3	**42**
4	20	25	45	202	63	**57**
5	0	2	3	15	80	**80**
Total	**129**	**109**	**163**	**287**	**161**	**51**

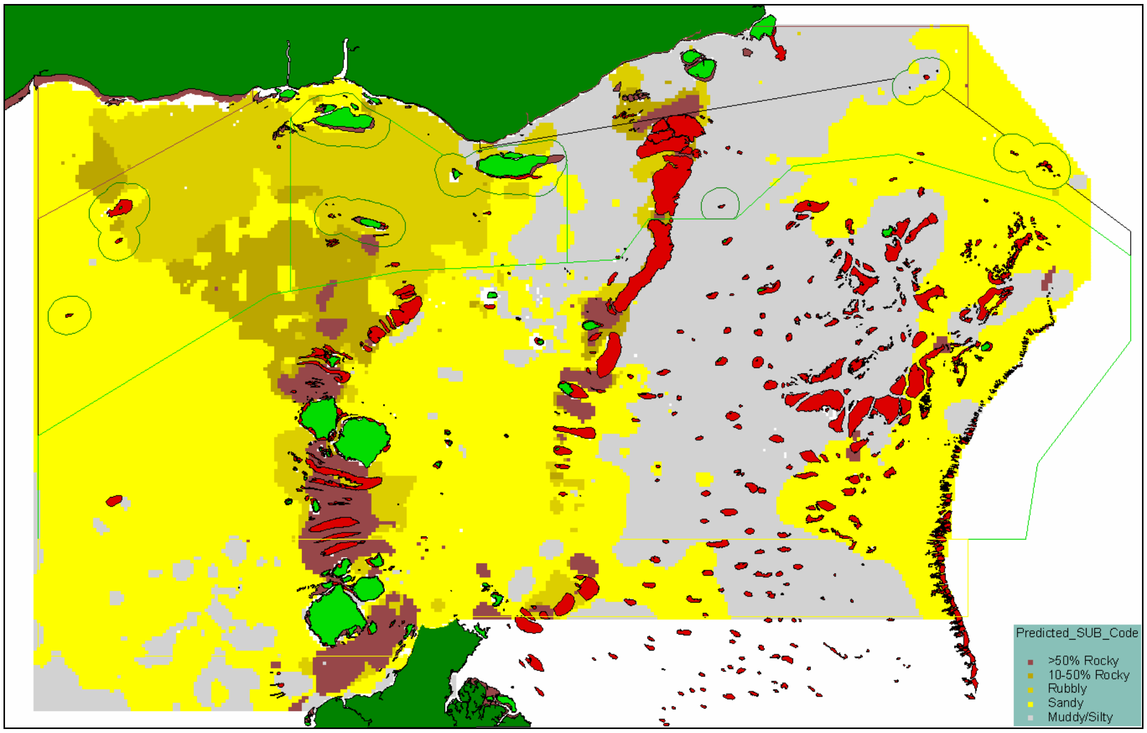


Figure 2.7-1. Mapping of predicted SUB_CODE over Torres Strait based on LDA functions applied to the gridded physical covariate dataset.

2.7.1.2. Epibenthos Classification (BIO_CODE):

The epibenthos classification (BIO_CODE), widely available from Torres Strait seabed surveys, was defined as (see section 2.1.1 for more details):

1 = dense fauna,
2 = sparse fauna,
3 = very sparse fauna,
4 = no fauna

The results from forward stepwise LDA on BIO_CODE categories alone showed the following. Based on the physical covariate data that was associated with the survey sites, categories 1 & 2, and 2 & 3 were most similar and category 4 was quite different to the others (Table 2.7-4). The most important covariates for discriminating BIO_CODE categories were seabed current stress, SeaWiFS average annual attenuation at 490 nm (turbidity), CARS nitrate standard deviation, and sediment gravel fraction (Table 2.7-5). The overall jack-knifed discrimination accuracy was 44% correct and ranged from 33% correct for category 2 to 67% correct for category 4 (Table 2.7-6) — in most cases the LDA confuses the most similar (and adjacent) category, so performs somewhat better than the raw error rates indicate.

The mapping of predicted BIO_CODE over Torres Strait (Figure 2.7-2) closely reflects the available data (Figure 2.4-2) and can be considered a reasonably reliable representation of the distribution of the gross epibenthos facies. Nevertheless, it is unlikely that the area in the vicinity of the outer barrier has the predicted extensive gardens of dense epibenthos.

Table 2.7-4. Between groups F-matrix for the physical covariate data associated with the BIO_CODE categories at sites (df = 6, 840). Small F-values indicate that the covariates of respective categories are more similar; large values indicate that the covariates of categories are more different.

	BIO_CODE Category			
	1	2	3	4
1	0.000			
2	5.395	0.000		
3	15.777	6.259	0.000	
4	42.306	27.732	32.097	0.000

Table 2.7-5. F-values for physical covariates assessed by LDA for BIO_CODE. Increasing F-values indicate covariates that were more important for discriminating BIO_CODE categories. Covariates in the left columns were included in the final model; those on the right were excluded.

Covariate included in model		Covariate excluded from model	
Variable	F-value	Variable	F-value
M_BSTRESS	31.82	auDB_CRBNT	3.08
SW_K490_Y_AV	10.99	DEM_SLOPE	2.64
CARS_NO3_SD	10.31	CARS_O2_AV	2.6
auDB_GRAVEL	10.13	SW_CHLA_SD	2.46
auDB_ROCK	6.11	CARS_T_AV	2.28
CARS_NO3_AV	4.06	CARS_PO4_AV	2.17
		SW_K490_Y_SD	2.16
		SW_K490_SD	1.8
		CARS_T_SD	1.73
		CARS_S_AV	1.46
		SW_CHLA_Y_SD	1.44
		CARS_O2_SD	1.37
		auDB_GRNSZ	1.16
		CARS_S_SD	1.04
		auDB_GRNSRT	0.99
		CARS_SI_SD	0.79
		SW_CHLA_Y_AV	0.46
		CARS_PO4_S	0.35
		CARS_SI_AV	0.2
		AGSO_DEM	0.16
		DEM_ASPEC	0.13
		auDB_SAND	0.12
		auDB_MUD	0.06
		SW_D_B_IRRAD	0.06

Table 2.7-6. Jackknifed classification matrix for discriminating BIO_CODE categories on physical covariates. Cases in row categories classified into columns. % correct shows number of cases in each row classified into the correct column. Diagnostic statistics: Wilks' lambda=0.670, approx.F= 20.114, df= 18, 2376.

	SUB_CODE Category				
	1	**2**	**3**	**4**	**%correct**
1	44	28	13	5	**49**
2	37	42	31	17	**33**
3	96	76	146	107	**34**
4	14	10	45	138	**67**
Total	**191**	**156**	**235**	**267**	**44**

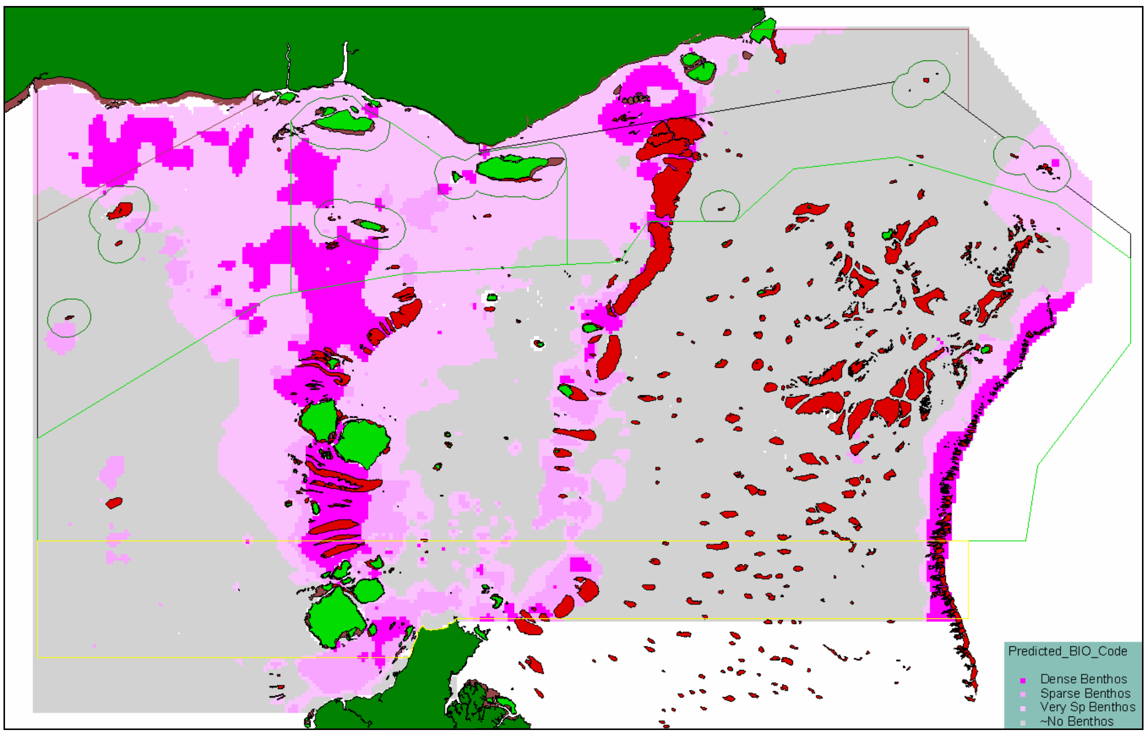


Figure 2.7-2. Mapping of predicted epibenthos BIO_CODE over Torres Strait based on LDA functions applied to the gridded physical covariate dataset.

2.7.1.3. Algae Presence/Absence (ALGAE_PA):

The Algae presence/absence data was widely available from Torres Strait seabed surveys, and was defined as: 0= absent, 1= present. The results from forward stepwise LDA on Algae_PA alone showed the following. The physical covariate data that was associated with the survey sites for presence and absence was quite different (F=22.847, df=5, 843). The most important covariates for discriminating Algae_PA were CARS Oxygen average, CARS Salinity average (Table 2.7-7). The overall jack-knifed discrimination accuracy was 73% correct and was similar for both presence and absence (Table 2.7-8).

The mapping of predicted Algae_PA over Torres Strait (Figure 2.7-3) closely reflects the available data (Figure 2.4-5), indicating that the numerous algal species are widely distributed through much of Torres Strait, and can be considered a reasonably reliable representation of the distribution of the gross algal distribution.

Table 2.7-7. F-values for physical covariates assessed by LDA for ALGAE_PA. Increasing F-values indicate covariates that were more important for discriminating ALGAE_PA categories. Covariates in the left columns were included in the final model; those on the right were excluded.

Covariate included in model		Covariate excluded from model	
Variable	F-value	Variable	F-value
CARS_O2_AV	69.53	CARS_PO4_S	3.13
CARS_S_AV	29.68	AGSO_DEM	1.24
auDB_ROCK	6.04	CARS_PO4_AV	1.14
CARS_NO3_AV	5.82	auDB_GRNSZ	0.91
auDB_CRBNT	5.26	SW_CHLA_Y_AV	0.76
		SW_CHLA_SD	0.76
		SW_CHLA_Y_SD	0.71
		auDB_MUD	0.64
		CARS_T_AV	0.61
		CARS_SI_AV	0.42
		M_BSTRESS	0.41
		SW_K490_Y_AV	0.36
		CARS_S_SD	0.32
		CARS_T_SD	0.25
		auDB_GRAVEL	0.18
		SW_D_B_IRRAD	0.18
		SW_K490_Y_SD	0.17
		DEM_SLOPE	0.13
		CARS_NO3_S	0.08
		CARS_O2_SD	0.06
		auDB_GRNSRT	0.06
		CARS_SI_SD	0.05
		auDB_SAND	0.02
		DEM_ASPEC	0.01
		SW_K490_SD	0.01

Table 2.7-8. Jackknifed classification matrix for discriminating ALGAE_PA categories on physical covariates. Cases in row categories classified into columns. % correct shows number of cases in each row classified into the correct column. Diagnostic statistics: Wilks' lambda=0.881, approx.F= 22.847, df= 5, 843.

	Algae P/A		
	0	**1**	**%correct**
0	57	22	**72**
1	205	565	**73**
Total	**262**	**587**	**73**

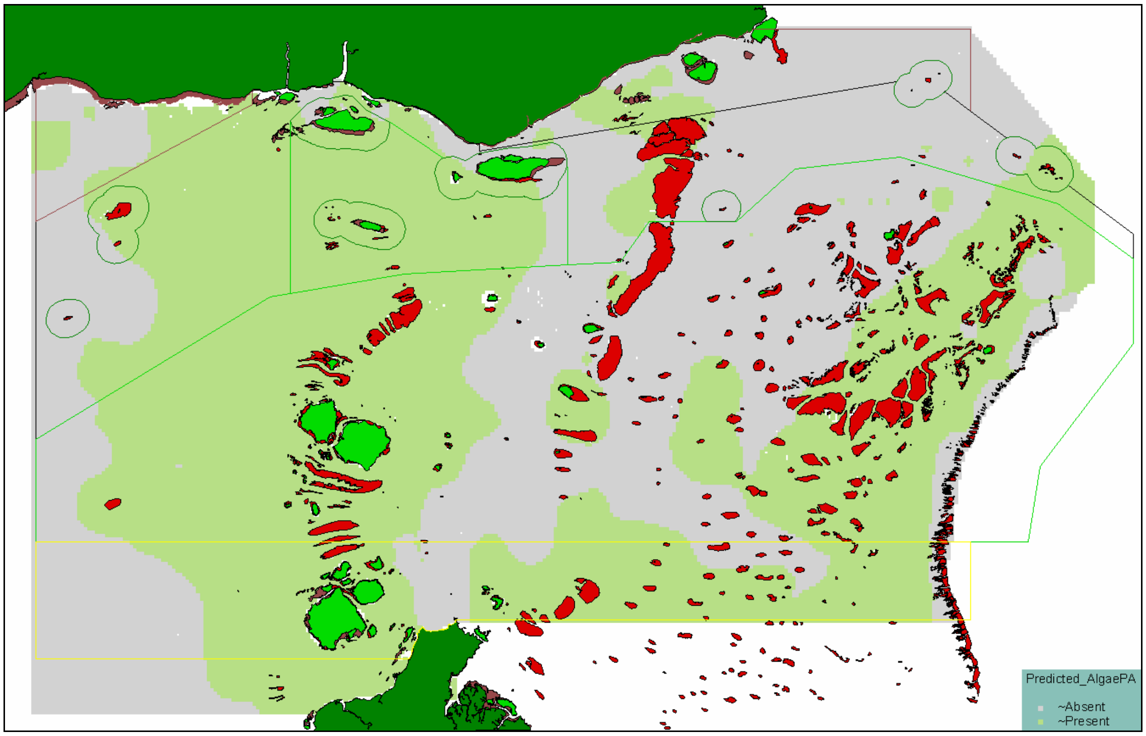


Figure 2.7-3. Mapping of predicted Algae P/A over Torres Strait based on LDA functions applied to the gridded physical covariate dataset.

2.7.1.4. Seagrass Presence/Absence (SEAGRASS_PA):

The Seagrass presence/absence data was widely available from Torres Strait seabed surveys, and was defined as: 0= absent, 1= present. The results from forward stepwise LDA on Seagrass_PA alone showed the following. The physical covariate data that was associated with the survey sites for presence and absence was quite different (F=28.132, df=6, 842). The most important covariates for discriminating Seagrass_PA were CARS Phosphate standard deviation, seabed current stress, CARS Nitrate standard deviation, sediment carbonate composition, and seabed irradiance estimated from SeaWiFS attenuation (Table 2.7-9). The overall jack-knifed discrimination accuracy was 68% correct and was slightly better for presence than absence (Table 2.7-10).

The mapping of predicted Seagrass_PA over Torres Strait (Figure 2.7-4) closely reflects the available data (Figure 2.4-5), indicating that the several seagrass species are mainly distributed through central western Torres Strait, and can be considered a reasonably reliable representation of the distribution of the gross seagrass distribution. Nevertheless, it is unlikely that the area in the vicinity of the outer barrier has the predicted seagrass presence.

Table 2.7-9. F-values for physical covariates assessed by LDA for SEAGRASS_PA. Increasing F-values indicate covariates that were more important for discriminating SEAGRASS_PA categories. Covariates in the left columns were included in the final model; those on the right were excluded.

Covariate included in model		**Covariate excluded from model**	
Variable	F-value	Variable	F-value
CARS_PO4_SD	31.55	SW_K490_SD	3.42
M_BSTRESS	29.04	CARS_T_AV	2.88
CARS_NO3_SD	25.46	SW_CHLA_SD	1.97
auDB_CRBNT	17.41	auDB_MUD	1.63
SW_D_B_IRRAD	12.48	CARS_T_SD	1.30
auDB_ROCK	5.89	auDB_GRNSRT	1.20
		SW_K490_Y_SD	1.17
		CARS_S_SD	1.00
		auDB_SAND	0.89
		SW_CHLA_Y_AV	0.77
		CARS_PO4_A	0.33
		CARS_SI_AV	0.21
		AGSO_DEM	0.19
		DEM_SLOPE	0.08
		auDB_GRAVEL	0.05
		SW_K490_Y_AV	0.04
		DEM_ASPEC	0.03
		CARS_SI_SD	0.03
		CARS_S_AV	0.03
		CARS_O2_AV	0.02
		CARS_NO3_A	0.01
		CARS_O2_SD	0.00
		auDB_GRNSZ	0.00
		SW_CHLA_Y_SD	0.00

Table 2.7-10. Jackknifed classification matrix for discriminating SEAGRASS_PA categories on physical covariates. Cases in row categories classified into columns. % correct shows number of cases in each row classified into the correct column. Diagnostic statistics: Wilks' lambda=0.833, approx.F= 28.132, df= 6, 842.

	Algae P/A		
	0	**1**	**%correct**
0	338	196	**63**
1	77	238	**76**
Total	**415**	**434**	**68**

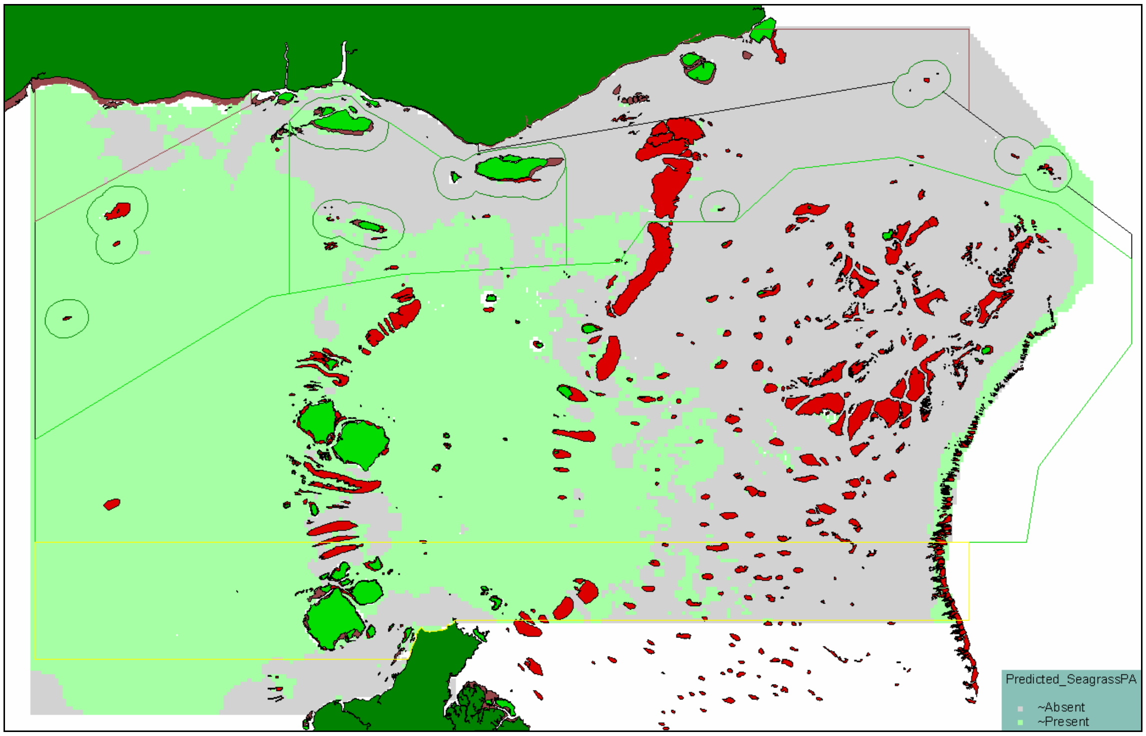


Figure 2.7-4. Mapping of predicted Seagrass P/A over Torres Strait based on LDA functions applied to the gridded physical covariate dataset.

2.7.1.5. Clustered Sub_Code, Bio_Code, Algae & Seagrass P/A:

The broad coverage 'low-level' bio-survey attributes were clustered to characterize the predominant mixtures of these habitat facies. Initially, the dataset was clustered into 4, 6 and 9 groups, using two algorithms for the K-means method (Euclidean & Sums-of-Squares), then the LDA jack-knifed classification performance of each was estimated to examine the trade-off between information detail (number of clusters) and potential mapping reliability (classification accuracy). In this case, 6 clusters appeared to be a reasonable compromise between information content and bio-physical classification success, and the Euclidean algorithm performed slightly better than SS.

From the statistics associated with each of the 6 clusters, it is possible to characterise them broadly as follows (Figure 2.7-5):

1: muddy/sandy, very sparse/no benthos, ~90% likelihood algae, ~40% likelihood seagrass
2: rubbly/some hard sub, dense/sparse benthos, ~95% likelihood algae, ~25% likelihood seagrass
3: sandy, sparse/very sparse benthos, ~65% likelihood algae, ~20% likelihood seagrass
4: rubble/some hard sub, very sparse benthos, ~95% likelihood algae, ~55% likelihood seagrass
5: mostly hard sub, dense/sparse benthos, ~100% likelihood algae, ~10% likelihood seagrass
6: some hard sub, very sparse benthos, ~95% likelihood algae, ~5% likelihood seagrass

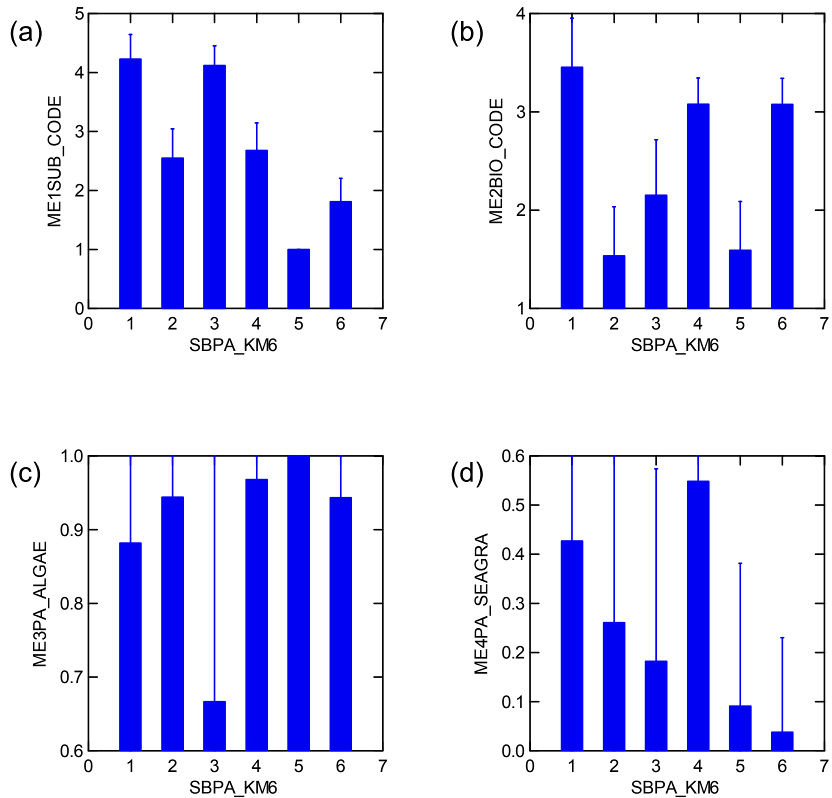


Figure 2.7-5. Mean and standard deviation of attributes for six K-means clusters of the 'low-level' habitat dataset: (a) seabed Sub_Code, (b) epibenthos Bio-Code, (c) Algae presence/absence, and (d) Seagrass presence/absence.

The results from forward stepwise LDA on the six Clusters showed the following. Based on the physical covariate data that was associated with the survey sites, cluster types 1 & 3, and 2 & 4 were most similar, with cluster 2 most different from 1 and clusters 4, 5, 6 intermediate from 1 (Table 2.7-11). The most important covariates for discriminating cluster types were seabed current stress, SeaWiFS estimated Chlorophyll a, with four other covariates moderately important (Table 2.7-12). The overall jack-knifed discrimination accuracy was 45% correct and ranged from 9% correct for cluster 3, which was fewest in number, to 65% correct for cluster 1, which was the most numerous type (Table 2.7-13).

The mapping of predicted cluster membership over Torres Strait (Figure 2.7-6) closely reflects the patterns for Sub_Code (Figure 2.7-1) and Bio_Code (Figure 2.7-2), which is not unexpected as these attributes dominated the cluster analysis. The map can be considered a reasonably reliable representation of the distribution of the broad habitat types in Torres Strait.

Table 2.7-11. Between groups F-matrix for the physical covariate data associated with the habitat cluster types at sites (df = 19, 853). Small F-values indicate that the covariates of respective categories are more similar; large values indicate that the covariates of categories are more different.

	Category					
	1	2	3	4	5	6
1	0					
2	46.916	0				
3	1.374	7.296	0			
4	33.746	2.555	5.184	0		
5	32.547	11.067	9.784	10.196	0	
6	23.899	5.506	5.919	4.864	4.117	0

Table 2.7-12. F-values for physical covariates assessed by LDA for habitat cluster types. Increasing F-values indicate covariates that were more important for discriminating cluster types. Covariates in the left columns were included in the final model; those on the right were excluded.

Covariate included in model		**Covariate excluded from model**	
Variable	F-value	Variable	F-value
M_BSTRESS	44.75	auDB_MUD	2.77
SW_CHLA_AV	14.56	CARS_T_SD	1.84
CARS_NO3_AV	8.10	auDB_ROCK	1.81
SW_K490_SD	7.62	CARS_SI_AV	1.32
auDB_GRNSZ	4.91	CARS_SI_SD	1.17
CARS_NO3_SD	4.06	CARS_S_SD	1.11
		auDB_CRBNT	0.98
		SW_K490_AV	0.96
		auDB_GRAVEL	0.92
		CARS_O2_AV	0.90
		CARS_O2_SD	0.89
		AGSO_DEM	0.88
		SW_CHLA_SD	0.88
		CARS_PO4_AV	0.79
		auDB_SAND	0.72
		CARS_T_AV	0.71
		CARS_S_AV	0.65
		CARS_PO4_SD	0.54
		auDB_GRNSRT	0.52

Table 2.7-13. Jackknifed classification matrix for discriminating habitat cluster types on physical covariates. Cases in row categories classified into columns. % correct shows number of cases in each row classified into the correct column. Diagnostic statistics: Wilks' lambda=0.570, approx.F=17.414, df=30, 3466.

	Cluster						
	1	2	3	4	5	6	%correct
1	278	36	36	51	16	14	65
2	18	43	17	32	19	32	27
3	18	3	3	2	2	5	9
4	18	33	17	35	18	34	23
5	2	1	5	6	20	10	45
6	4	11	7	6	11	14	26
Total	**338**	**127**	**85**	**132**	**86**	**109**	**45**

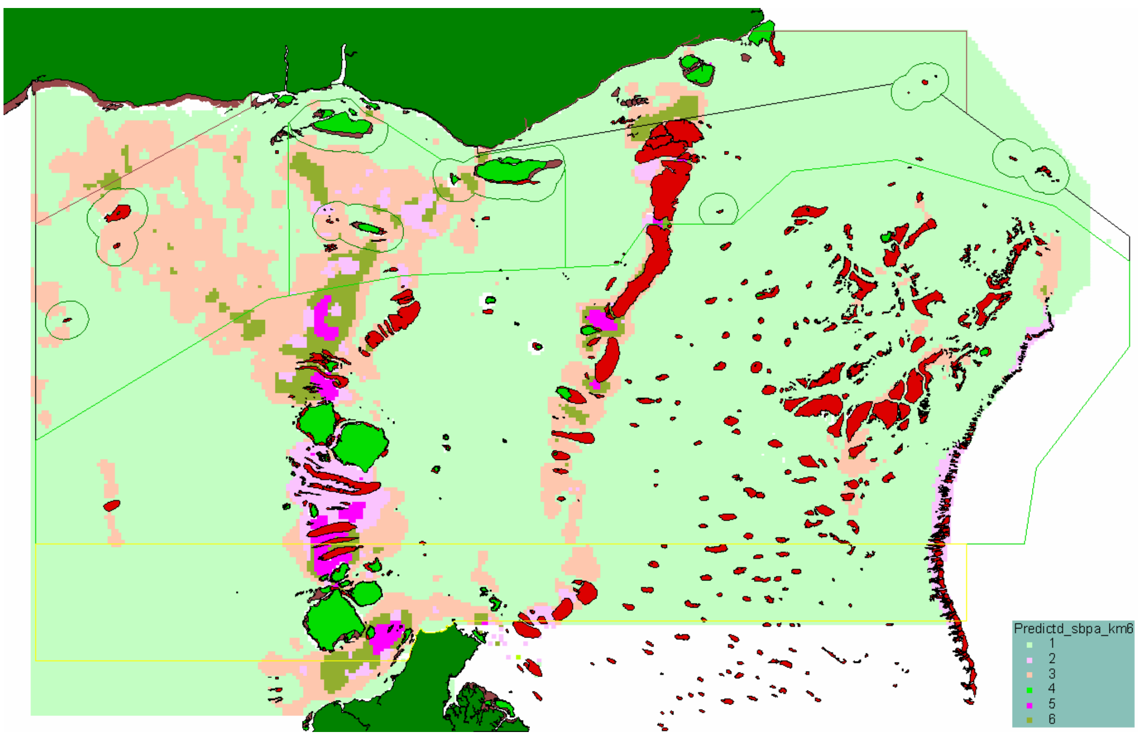


Figure 2.7-6. Mapping of predicted habitat cluster membership over Torres Strait based on LDA functions applied to the gridded physical covariate dataset. Habitat clusters were characterized as follows:

1: muddy/sandy, very sparse/no benthos, ~90% likelihood algae, ~40% likelihood seagrass
2: rubbly/some hard substrate, dense/sparse benthos, ~95% likelihood algae, ~25% likelihood seagrass
3: sandy, sparse/very sparse benthos, ~65% likelihood algae, ~20% likelihood seagrass
4: rubble/some hard substrate, very sparse benthos, ~95% likelihood algae, ~55% likelihood seagrass
5: mostly hard substrate, dense/sparse benthos, ~100% likelihood algae, ~10% likelihood seagrass
6: some hard substrate, very sparse benthos, ~95% likelihood algae, ~5% likelihood seagrass

2.7.2. 'Medium-level' bio-survey dataset

A slightly higher level of seabed habitat information was available at a reduced number of sites (Table 2.4-1B). The higher level biological survey data included the ordinal scale of epibenthos density (BIO_CODE), the percentage cover of seagrass and algae over the survey transects (PCT_TOT_SGRS and PCT_TOT_ALG respectively), and the estimated percentage cover of a number of sediment classes: mud-silt (PCT_MUD_SILT), sand-gravel (PCT_SAND_GRV), rubble (PCT_RUBBLE), consolidated rubble (PCT_CONS_RUB) and hard, rock pavement (PCT_SUB_HARD). Individually, these attributes were available from between 655 and 1196 sites in Torres Strait; however, there were only 421 sampling sites where records for all 8 of these attributes were recorded (Figure 2.7-7). These data were restricted to the central part of Torres Strait with most data collected from around the northern sections of the Warrior Reefs and around the island chain between Cape York and Papua New Guinea.

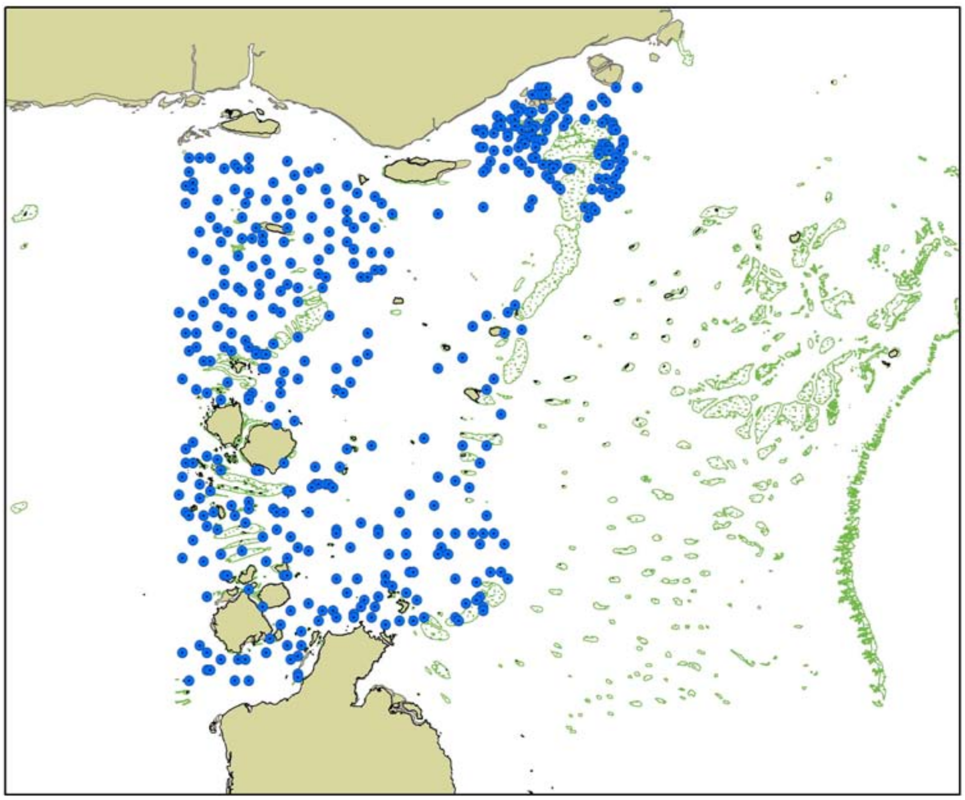


Figure 2.7-7. Map of Torres Strait showing the 421 sampling sites where records for all 8 "medium" level survey data were recorded.

In order to characterise the habitat facies represented by these data and examine the relationships between these data and the physical co-variate data, the habitat data were clustered and then linear discriminant function analysis was applied to determine how well the clusters could be described by the physical co-variate data. Only sites that had data for all 8 survey variables were included in these analyses.

The distribution of all survey variables was examined to determine whether transformation was necessary prior to analysis (Figure 2.7-8). BIO_CODE was originally recorded on an ordinal scale (see section 2.1.1 for more details). In an effort to prevent any particular variable having a disproportionate effect on the clustering process, we re-coded the values for BIO_CODE to match approximately the coverage given in the definitions (section 2.1.1) i.e. 1 = 60%, 2 = 30%, 3 = 5% and 4 = 0%. Similarly, the five sediment variables were rescaled to range between 0 and 20 rather than zero and 100. All variables were then $\log_e(x+1)$ transformed because of the highly right-skewed nature of most of their distributions (Figure 2.7-8).

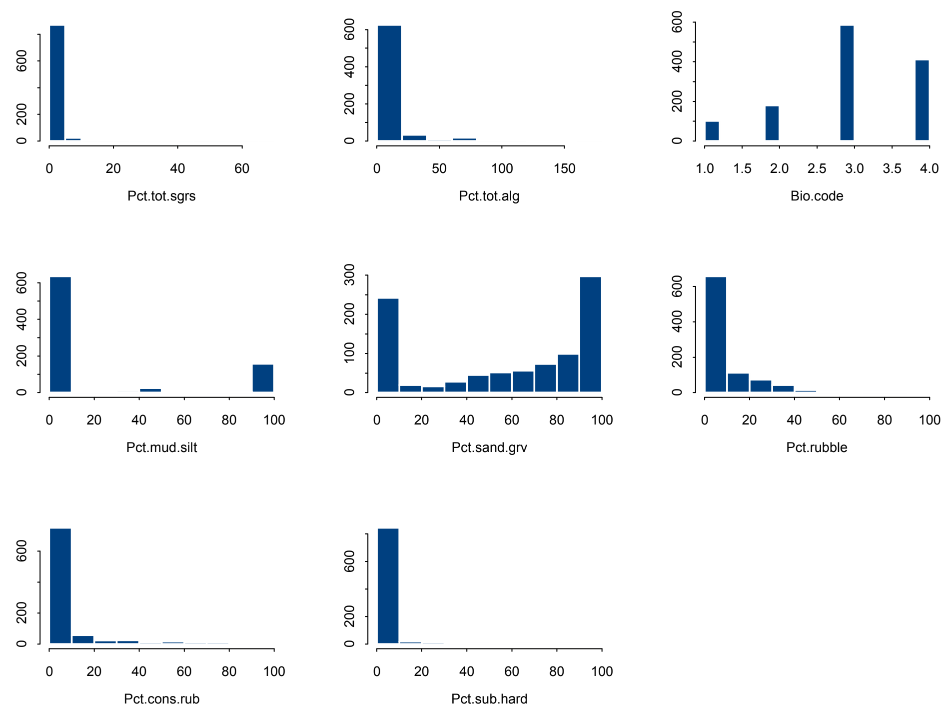


Figure 2.7-8. Histograms of the untransformed distributions of all “high” level survey data.

The Bray-Curtis dissimilarity distance metric was used to estimate the bio-survey data distance between all sites, then multidimensional scaling (MDS) with ordinal scaling in 4 dimensions was applied to reduce the dimensionality of the B-C matrix. The MDS of the “medium” level survey data revealed little in the way of discrete groupings. There was one central cloud of points containing the majority of observations surrounded by several disparate groupings containing several observations each (Figure 2.7-9).

The 4 dimensional MDS coordinates for each survey sites were then clustered using K-Means algorithm. As with the low-level dataset, a range of numbers of clusters was tried (4, 6 and 9), and as before, choosing an appropriate number of clusters was a compromise between the information content of the cluster characterisation and the biophysical classification success. The cluster membership of each site was then joined to the matching data for the collated physical co-variates (sections: 2.1.1 and 2.3) and a linear discriminant function analysis (LDA) was again used as above to develop a simple biophysical model for the clusters, provide cross-validated error rates, and to predict and map the cluster membership of each 0.01 degree cell for an estimate of the full coverage for Torres Strait.

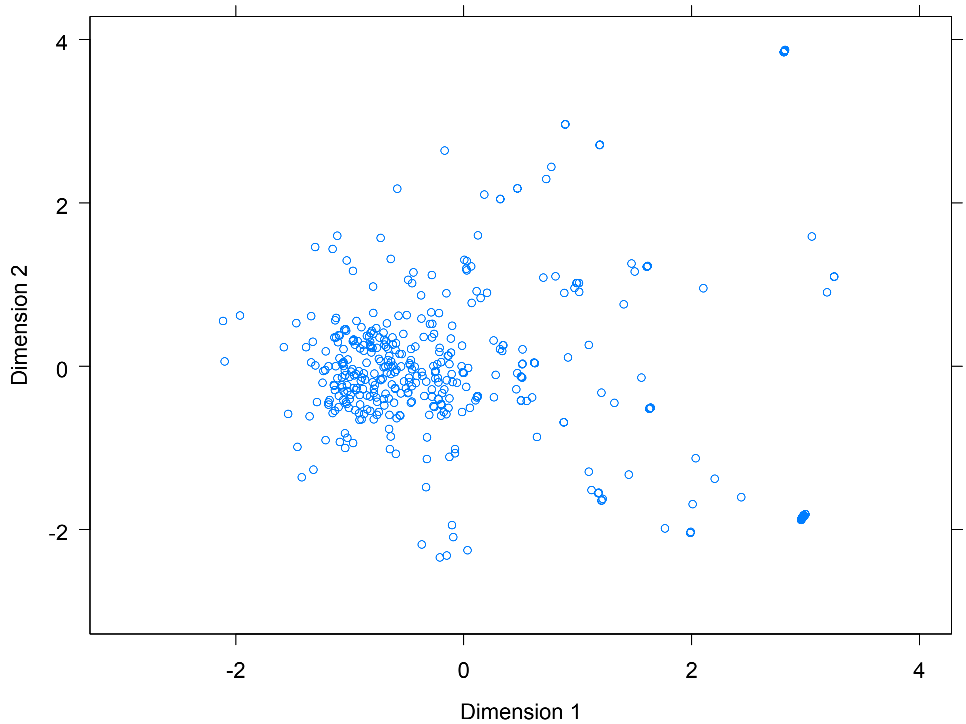


Figure 2.7-9. Ordination of the first and second dimensions of a multi dimensional scaling of the Bray-Curtis Dissimilarities of the "high" level Torres Strait survey data.

The results of the relative biophysical classification performance are shown in Table 2.7-14. The smaller the number of clusters, the better the performance of the discriminant functions in terms of correctly assigning a site to a cluster based on the values of the co-variates at that site, but the less information available about each cluster because their characteristics become more generalised as the total number of clusters decreases. In terms of the diagnostic statistics, 4 clusters gives the best performance (Table 2.7-14); however, 6 clusters were chosen to maximise the amount of biophysical information while retaining acceptable biophysical classification performance.

Table 2.7-14. Cluster diagnostics for K-Means clustering of the "high" level data.

	Wilks' Lambda	F value	Pr > F	Cross-validation: % correct
4 clusters	0.462	7.38	< 0.0001	51.2
6 clusters	0.384	4.98	< 0.0001	38.7
9 clusters	0.299	4.72	< 0.0001	22.2

Based on the statistics of the survey attributes (Figure 2.7-10), the composition of each cluster may be characterised as follows:

Cluster 1 – Sparse to dense epibenthos, sparse to medium algal cover, sandy with some rubble and consolidated rubble and pavement; no seagrass.

Cluster 2 – Sandy-muddy areas with little epibenthos, algae or seagrass.

Cluster 3 – Similar to cluster 1; sparse to dense epibenthos, sandy with some rubble and consolidated rubble. Some areas of pavement, although less than cluster 1. Less algal cover than cluster 1, very sparse seagrass cover.

Cluster 4 – Barren muddy areas having no epibenthos, algae or seagrass.

Cluster 5 – Very sparse epibenthos, sparse to medium cover of algae, sparse cover of seagrass. Generally sandy with some rubble, consolidated rubble and pavement.

Cluster 6 – Barren sandy areas with very little epibenthos and virtually no seagrass or algae.

Of the surveyed sites, the largest cluster was cluster 5 with 225 sites, cluster 4 was the smallest having only 12 sites, and clusters 1 2 3 and 6 were intermediate with between 39 to 50 sites.

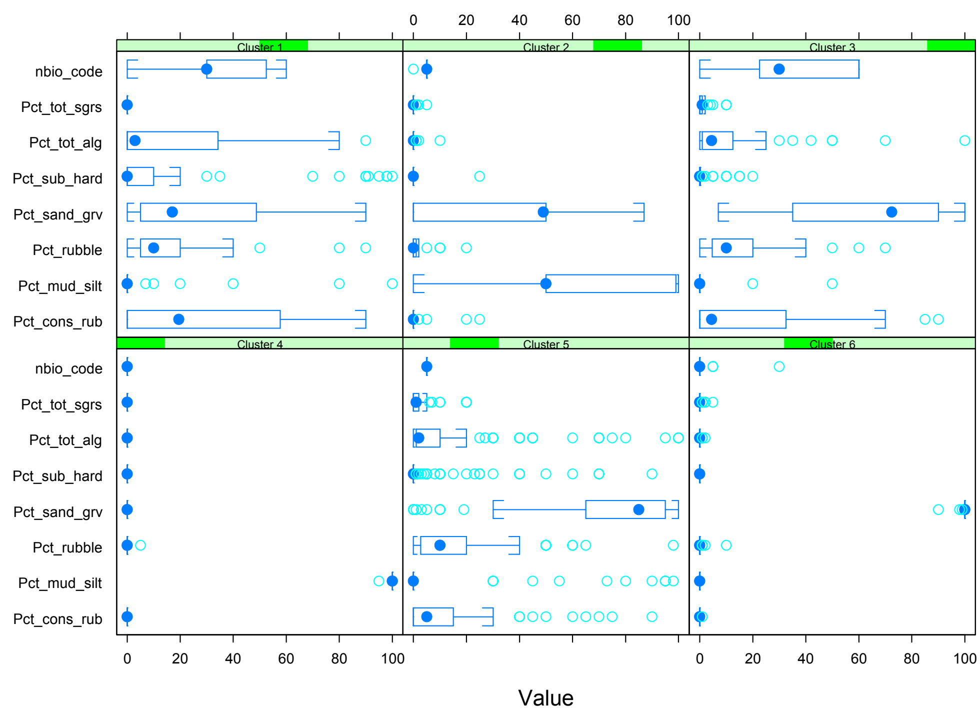


Figure 2.7-10. Box and whisker plots of distribution of the “high” level survey data among the 6 clusters. Median (dark blue closed circles), inter-quartile ranges (dark blue open boxes), 1.5 times the inter-quartile range (outer fences) except the outliers (light blue open circles).

The results of the LDA indicated that 17 of the physical co-variates appeared to be important in influencing the distribution of the clusters, and were included in the discriminant functions (Table 2.7-15). The two co-variates that had the highest partial r^2 and F values, and therefore were most important in the discriminant functions were the modelled seabed current shear stress (M_BSTRESS) and the standard deviation of the annual SeaWiFS chlorophyll *a* estimate (SW_CHLA_Y_SD). The overall jack-knifed discrimination accuracy was ~39% correct and ranged from 0% correct for cluster 4, which had only 12 sites, to 56.4% correct for cluster 2 (Table 2.7-16).

Table 2.7-15. LDA statistics of co-variates included in the discriminant functions of the "medium" level survey data.

Co-Variate	Partial r^2	F-value	Pr > F
M_BSTRESS	0.1364	12.61	<.0001
SW_CHLA_Y_SD	0.0833	7.25	<.0001
CARS_O2_AV	0.0595	5.05	0.0002
CARS_S_AV	0.0557	4.71	0.0003
DEM_BATHY	0.0535	4.51	0.0005
CARS_PO4_S	0.0483	4.05	0.0014
CARS_O2_SD	0.0466	3.90	0.0018
auDB_ROCK	0.0448	3.74	0.0025
WTD_TRWL_E	0.0415	3.46	0.0045
SW_K490_Y_SD	0.0375	3.11	0.0091
auDB_GRAVEL	0.0353	2.92	0.0133
DEM_ASPECT	0.0307	2.52	0.0288
auDB_CRBNT	0.0305	2.51	0.0296
SW_CHLA_Y_AV	0.0292	2.40	0.0368
SW_K490_Y_AV	0.0281	2.30	0.0441
auDB_GRNSZ	0.0266	2.18	0.0560
CARS_PO4_AV	0.0218	1.78	0.1168

Table 2.7-16. Cross-validation summary for the performance of the discriminant functions in assigning survey sites to clusters. The number in the top of each cell is the number of sites; the number in the bottom is the percent of the total number of observations for that cluster. The shaded diagonal cells show the numbers of sites that were correctly assigned to each cluster (columns, from rows).

CLUSTER	1	2	3	4	5	6	Total
1	20	1	12	0	14	3	**50**
	40.0	2.0	24.0	0.0	28.0	6.0	**100.0**
2	2	22	3	0	5	7	**39**
	5.1	56.4	7.7	0.0	2.8	18.0	**100.0**
3	7	1	16	0	14	6	**44**
	15.9	2.3	36.4	0.0	31.8	13.6	**100.0**
4	1	4	1	0	2	4	**12**
	8.3	33.3	8.3	0.0	16.7	33.3	**100.0**
5	35	20	53	0	88	33	**229**
	15.3	8.7	23.1	0.0	38.4	14.4	**100.0**
6	1	7	10	0	12	17	**47**
	2.1	14.9	21.3	0.0	25.5	36.2	**100.0**
Total	**66**	**55**	**95**	**0**	**135**	**70**	**421**
	15.7	**13.1**	**22.6**	**0.0**	**32.1**	**16.6**	**100.0**

The discriminant functions were used to predict the cluster membership of every other 0.01 degree cell in the entire Torres Strait area, based on the values of the co-variates at each cell. This prediction was then mapped to show the distribution of predicted habitat clusters (Figure 2.7-11). The mapping suggests there are areas of sparse to dense epibenthos with some algal cover and a substrate consisting of sand and rubble (clusters 1 and 3) in the areas between many of the reefs and islands in the central Torres Strait. There are extensive sandy barren (cluster 6) areas in the south-west, central and eastern Torres Strait. Barren sandy-mud and muddy (clusters 2 and 4) areas extend across the northeast, to the east of the Warrior Reef complex and amongst the mid-shelf eastern reefs. The central area to the north of Cape York and parts of the north-western Torres Straits are characterised by a very sparse coverage of epibenthos, sparse algal and seagrass cover and a substrate consisting mainly of sand with patches of rubble, consolidated rubble and pavement (cluster 5).

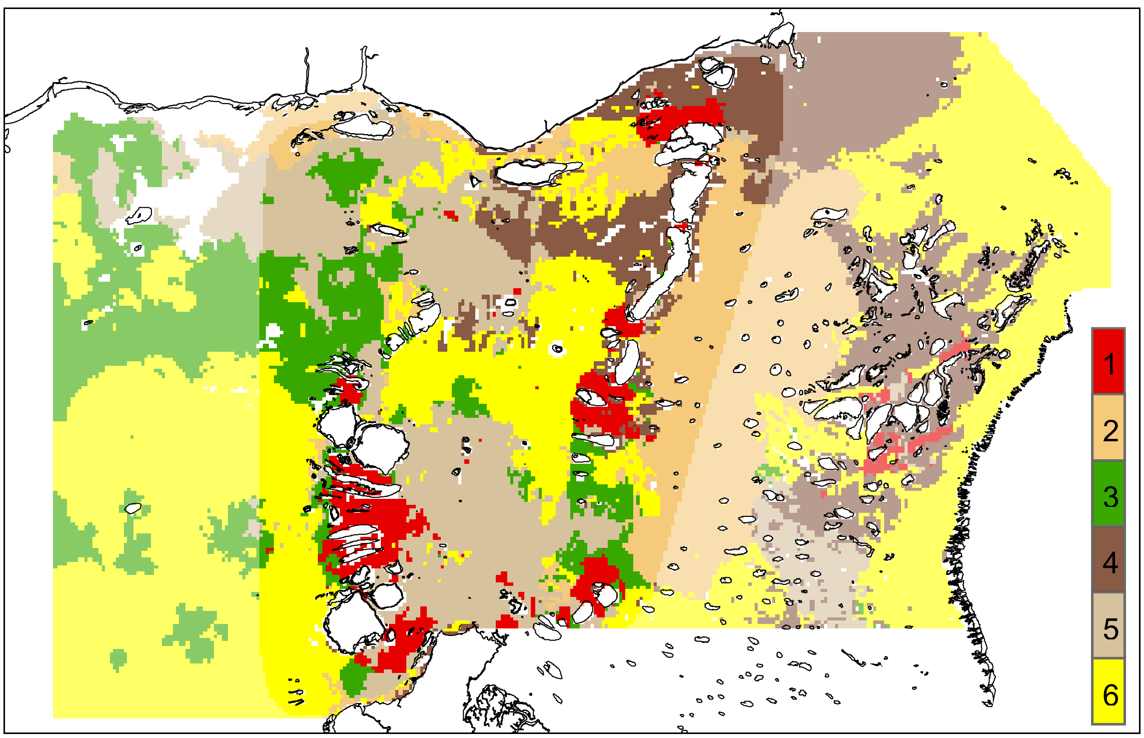


Figure 2.7-11. Mapping of predicted cluster membership (K-Means; 6 clusters) of the 'medium'-level survey data for the 0.01 degree gridded physical environment data, using linear discriminant function analysis. The darker shaded mapping represents the predictions within the area where survey data was collected; the lighter shaded mapping is outside the survey area.

2.7.3. Seabed Fish Dataset

The seabed fish dataset from the mid 1980's effects of trawling series of cruises, described in section 2.4.2.4 is characterised statistically here in a similar way as the seabed habitat data.

The standardised species sample weight data, averaged across all voyages, at each site was joined to the matching collated physical co-variate data. The mean catch rates were loge + 1 transformed and a matrix of Bray-Curtis dissimilarity metrics was calculated. The matrix was then reduced by hierarchical agglomerative clustering (with group-average linking) and by MDS. The first 4 dimensions of the MDS were then clustered using the K-Means algorithm in a similar manner to the “medium” level survey data. Linear discriminant function analysis was then performed to determine how well the co-variate data could be used to predict and map the fish bycatch assemblage.

The cluster and MDS analyses of the mean catch rates showed fairly clear groupings based on whether the sample site was within the commercial trawl grounds or within a closure (Figure 2.7-12, Figure 2.7-13).

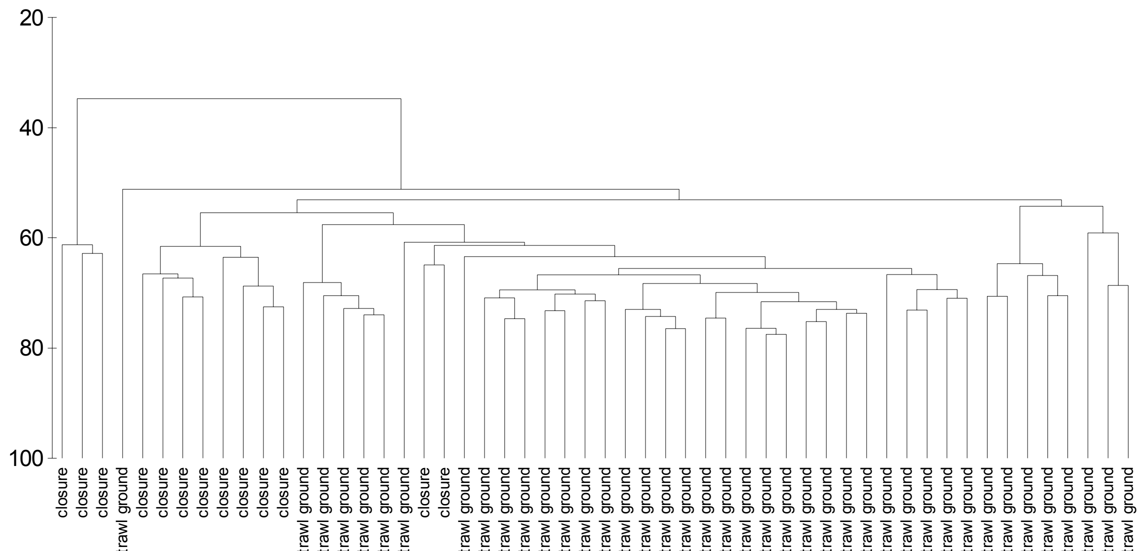


Figure 2.7-12. Hierarchical agglomerative clustering (group-average linking) of the mean catch rates of fish bycatch from prawn trawls done in the eastern Torres Strait during 1984-5. Each site is labeled based on whether it was located within the commercial trawl grounds or an area closed to fishing.

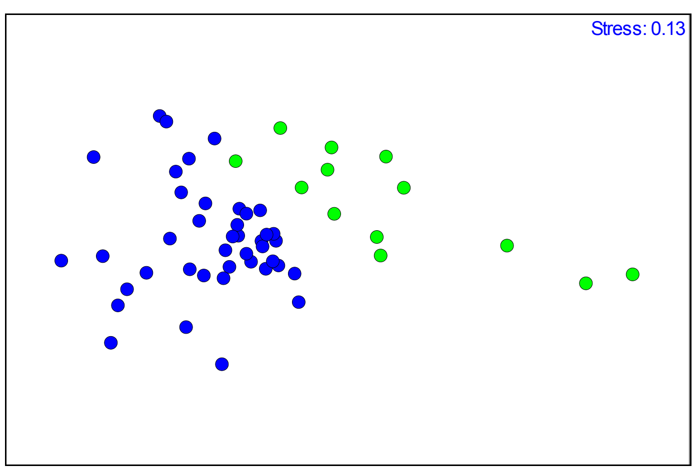


Figure 2.7-13. MDS ordination of the mean catch rates of fish bycatch from prawn trawl sampling conducted in the eastern Torres Strait during 1984-5. Each site is labelled based on whether it was located within the commercial trawl grounds (●) or an area closed to fishing (●).

Table 2.7-17. Mean catch rates of the 10 most abundant (biomass) fish in each cluster caught in the bycatch study in the central Torres Strait during the mid-1980s.

Cluster	Species	Family	Kg/ha	Habitat
1	*Lethrinus genivittatus*	Lethrinidae	7345.0	seagrass & weed beds
1	*Choerodon cephalotes*	Labridae	1815.6	coral reefs & nearby seagrass beds
1	*Pentapodus setosus*	Nemipteridae	1516.6	sand-rubble fringe of coral reefs
1	*Siganus canaliculatus*	Siganidae	1102.5	sand-weed areas
1	*Upeneus luzonius*	Mullidae	1066.0	Muddy bottoms
1	*Lethrinus laticaudis*	Lethrinidae	650.9	juveniles on seagrass; adults on coral reefs
1	*Gnathanodon speciosus*	Carangidae	572.3	coastal and offshore waters, usually near reefs
1	*Tragulichthys jaculiferus*	Diodontidae	507.6	soft bottom and trawl grounds
1	*Pseudomonacanthus elongatus*	Monacanthidae	471.7	?
1	*Psammoperca waigensis*	Centropomidae	428.7	rocky or coral reefs, frequently in weedy areas
2	Dasyatinae	Dasyatidae	3203.6	Sandy areas
2	*Priacanthus tayenus*	Priacanthidae	2389.3	coral reefs and rocky bottoms
2	*Saurida undosquamis*	Synodontidae	2057.4	sandy or muddy bottoms
2	*Nemipterus peroni*	Nemipteridae	1928.7	trawling grounds
2	*Saurida micropectoralis*	Synodontidae	1772.8	muddy bottoms
2	*Platycephalidae*	Platycephalidae	1549.0	sandy or muddy bottoms
2	*Nemipterus hexodon*	Nemipteridae	1264.0	trawling grounds
2	*Paramonacanthus japonicus*	Monacanthdae	967.1	weedy and sandy areas of coastal reefs
2	*Pomadsys maculatum*	Haemulidae	923.3	sandy or muddy bottoms
2	*Nemipterus furcosus*	Nemipteridae	883.3	trawling grounds
3	*Sphyrna mokarran*	Sphyrnidae	13034.8	coastal-pelagic, in passes and lagoons
3	*Stegastoma fasciatum*	Stegastomidae	4431.8	coastal & offshore waters in the vicinity of coral reefs
3	Dasyatinae	Dasyatidae	3910.4	sandy areas
3	*Rhina ancylostoma*	Rhyncobatidae	3798.7	coastal waters on mud or sand bottoms
3	*Nemipterus furcosus*	Nemipteridae	1540.8	trawling grounds
3	*Scolopsis taeniopterus*	Nemipteridae	1497.2	Sandy areas in the vicinity of coral reefs
3	*Priacanthus tayenus*	Priacanthidae	1336.7	coral reefs and rocky bottoms
3	*Myliobatus australis*	Myliobatidae	1266.2	?
3	*Nemipterus peroni*	Nemipteridae	1000.3	trawling grounds
3	*Pastinachus sephen*	Dasyatidae	977.6	flat sand or mud bottoms
4	*Saurida undosquamis*	Synodontidae	7307.5	sandy or muddy bottoms
4	*Carangoides talamparoides*	Carangidae	3895.8	coastal waters
4	*Leiognathus splendens*	Leiognathidae	3086.0	coastal waters, commonly on trawl grounds
4	*Priacanthus tayenus*	Priacanthidae	2946.3	coral reefs and rocky bottoms
4	*Paramonacanthus japonicus*	Monacanthdae	2923.1	weedy and sandy areas of coastal reefs
4	*Nemipterus peroni*	Nemipteridae	2239.5	trawling grounds
4	*Nemipterus furcosus*	Nemipteridae	2221.0	trawling grounds
4	*Scolopsis taeniopterus*	Nemipteridae	1938.6	sandy areas in the vicinity of coral reefs
4	*Leiognathus fasciatus*	Leiognathidae	1844.5	coastal waters, commonly on trawl grounds
4	*Saurida micropectoralis*	Synodontidae	1761.1	muddy bottoms

The K-Means clustering of the first 4 dimensions of the MDS ordination was restricted to 4 clusters because there was little evidence of finer clustering within the hierarchical dendrogram (Figure 2.7-12) of the only 54 sample sites in this dataset.

The K-Means clustering resulted in two small clusters and two larger clusters:

Cluster 1 – 3 sites, 84 species of fish – was dominated by species commonly associated with coral reefs

Cluster 2 – 8 sites, 173 species of fish – was dominated by species commonly found on trawl grounds

Cluster 3 – 20 sites, 225 species of fish – included some pelagic species as well as those found on trawl grounds

Cluster 4 – 23 sites, 271 species of fish – common trawl grounds species, but the dominant species were mostly different to those of cluster 3 and the pelagic species were not present.

Information on the most abundant species of fishes characterising these clusters is provided in Table 2.7-17.

The results of the LDA indicated that only 7 of the physical co-variates appeared to be important in influencing the distribution of the fish assemblage clusters, and were included in the discriminant functions (Table 2.7-18). The three co-variates that had the highest partial r^2 and F values, and therefore were most important in the discriminant functions were standard deviation of salinity (CARS_S_SD), the standard deviation of temperature (CARS_T_SD) and the amount of sand (auDB_SAND). The overall jack-knifed discrimination accuracy was ~79% correct and ranged from 75% correct for cluster 3, to 100% correct for cluster 1, which had only 3sites (Table 2.7-19).

Table 2.7-18. LDA statistics of co-variates included in the discriminant functions of the fish bycatch data.

Co-Variate	Partial r^2	F-value	Pr > F
CARS_S_SD	0.6959	38.13	<.0001
CARS_T_SD	0.4445	13.07	<.0001
auBD_SAND	0.3103	7.20	0.0004
CARS_PO4_AV	0.2632	5.24	0.0035
SW_CHLA_Y_SD	0.1731	3.28	0.0290
SW_D_B_IRRAD	0.1669	3.07	0.0369
CARS_O2_AV	0.1192	2.03	0.1231

As was done with the seabed habitat survey data, the LDA functions were used to predict the cluster membership of each 0.01 cell in the entire Torres Strait study area, based on the values of the co-variates at each cell. These predictions were then mapped to show the estimated distribution of benthic fish assemblages (Figure 9). The prediction beyond the trawl fish sampling area has unknown certainty and is likely to be unreliable.

Cluster 1 which was predominantly reef associated fish is within the area closed to the fishery. Outside the area sampled, the discriminant functions have allocated virtually the whole of the western Torres Strait to this cluster; something which is plainly incorrect. Interestingly though, the areas just to the east of the outer Barrier reef have also been allocated to cluster 1 (Figure 2.7-14). Cluster 2 (predominantly trawl ground fish) is concentrated in the north east of the study area and to the eastern Torres Strait outside the study area. Clusters 3 & 4 extend north-south in the central part of the Strait.

Table 2.7-19. Cross-validation summary for the performance of the discriminant functions in assigning bycatch sites to clusters. The number in the top of each cell is the number of sites; the number in the bottom is the percent of the total number of observations for that cluster. The shaded diagonal cells show the numbers of sites that were correctly assigned to each cluster (columns, from rows).

CLUSTER	1	2	3	4	Total
1	3 100.0	0 0.0	0 0.0	0 0.0	**3** **100.0**
2	0 0.0	7 87.5	0 0.0	1 12.5	**8** **100.0**
3	1 5.0	1 5.0	15 75.0	3 15.0	**20** **100.0**
4	0 0.0	1 4.4	4 17.4	18 78.3	**23** **100.0**
Total	**4** **7.4**	**9** **16.7**	**19** **35.2**	**22** **0.7**	**54** **100.0**

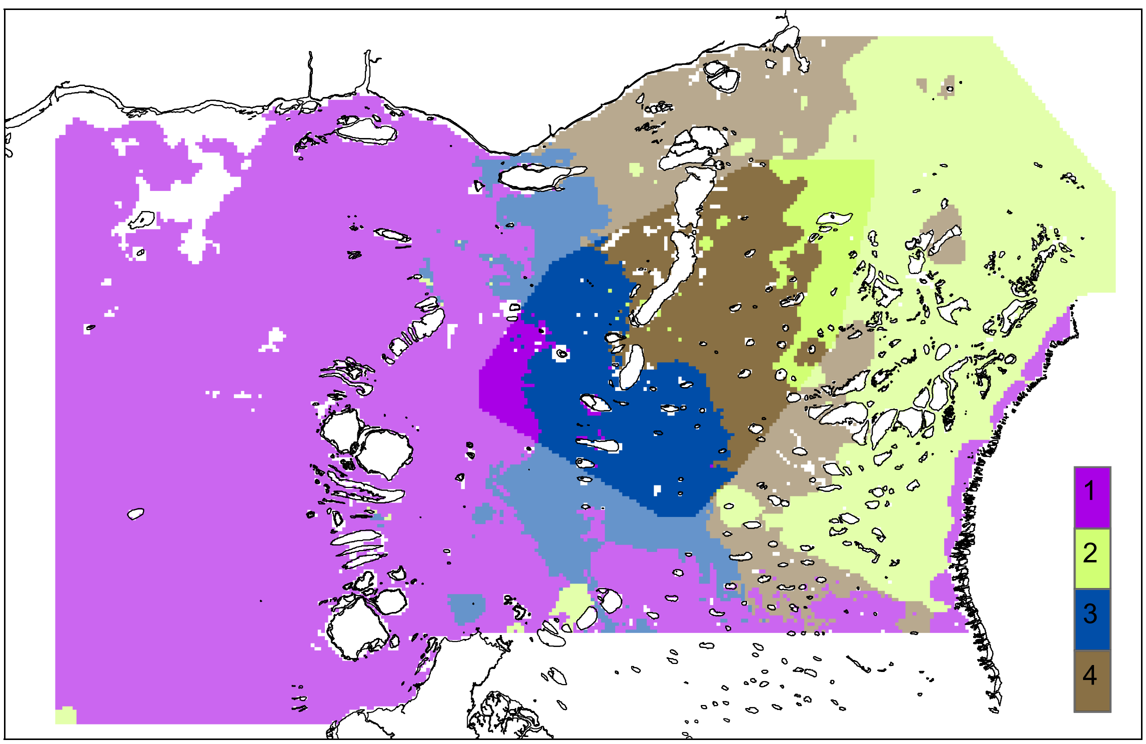


Figure 2.7-14. Mapping of predicted cluster membership (K-Means; 4 clusters) of the fish bycatch data for the 0.01 degree gridded physical environment data, using linear discriminant function analysis. The darker shaded mapping represents the predictions within the area where bycatch data was collected; the lighter shaded mapping is outside the survey area.

2.8. STRATIFICATION AND SAMPLING DESIGN

The future sampling for seabed biodiversity mapping in Torres Strait, as part of the CRC-TS, requires an optimal strategy for the selection of survey sites. The primary purpose of the survey itself is to obtain data on the spatial distribution of benthic biota, so that subsequent bio-physical modelling can make use of the physical environment co-variates to interpolate and map. Given that the number of sites that can be sampled is limited, it is obviously important to place the samples in a way that yields as much information as possible overall. This requires that the environment space, or multi-dimensional covariate space, rather than the 2-dimensional space must be sampled representatively and the approach to achieve this is stratification. Further, the stratification must be relevant to the benthic biotic, so it must be informed by measures of the biological importance of each covariate. This approach will optimally ensure that the biodiversity and physical attributes of as many different habitats types as possible, given the available resources, would be characterised. The physical variables collated as part of this project, which are known in advance of the survey, will be used to guide the stratification. Biological information will be taken into account by weighting the physical variables based on their relative importance in bio-physical relationships — variables of greater influence on biological patterns having a larger weighting and influence in the stratification.

From an earlier study (Pitcher et al, 2002) we have measures of the “importance” of these co-variates with respect to correlations with the abundance of many benthic species in a detailed survey of an adjacent area of the far northern Great Barrier Reef. Conceptually, important variables are those for which benthic composition changes significantly along a gradient of the variable. The survey should be designed to ensure that such important variables are sampled finely, so that the expected benthic diversity is reliably captured. That is, we should stratify our design with respect to the important variables.

Further, the sampling strategy should also consider the spatial resolution required for management utility. A scale of several 10s km was considered appropriate for broad scale characterisation. The implications of the spatial auto-correlation distance (section 2.6) and considerations of the benefit-cost of logistics (at about 1 site per hour) also indicate a sampling density of approximately 10 km average separation. In this approach, approximately 20-50 primary strata with similar physical characteristics will be identified from importance weighted physical covariates of almost fifty thousand 0.01° grid cells covering the shelf area of Torres Strait. The size (area) of strata will vary depending on the number of grid cells having particular similar physical characteristics. ‘Replicate’ future sampling sites, about 10-20, will be assigned to each primary stratum.

2.8.1. Stratification

The potential survey area in Torres Strait, after excluding reefs and other areas that were too shallow, included 41,285 cells of side 0.01° (~1.11 km), each square being a candidate sample site. Given the spatial autocorrelation distances, the average distance between sites should not exceed about 0.1° (~11.1 km) indicating that not less than about 400 of these squares should be sampled. A 10% margin was added to this lower limit, thus the design provided for 440 sites, although the resources of the future mapping project would allow only about two thirds of these to be sampled during 2003/04 to 2005/06. At the centre of each 0.01 degree cell, the values of 28 physical variables were collated or interpolated and represent the Torres Strait region as a cloud of 41,285 points within a 28-dimensional physical-variable space. Ultimately, this space was to be partitioned into 440 relatively homogeneous regions (or *strata*), such that the expected benthic biodiversity would be homogeneous within each stratum but heterogeneous among strata. A sampling site would then be selected from each stratum to produce a set of 440 sites. This section describes the methods for achieving this partitioning or stratification of physical-variable space.

2.8.1.1. Principles of Partitioning

The basic principle behind the partitioning can be illustrated with the following simple two-dimensional example. Consider two physical variables *x* and *y* for which we have values at 1000 sites, and suppose that these sites sample the covariate space roughly uniformly (Figure 2.8-1(a)). We wish to partition the covariate space into 20 strata. If the two variables were equally important, then the partitioning in Figure 2.8-1(b) would be adequate, since the strata are roughly the same width in *x* and *y*. This partitioning was achieved using the "partitioning around medoids" (PAM) algorithm (Kaufman and Rousseeuw, 1990) (see below).

However, suppose the *x* variable is known to be 4 times more important than the *y* variable. Then we would prefer a partitioning more like that in Figure 2.8-1(c), where the strata are roughly 4 times narrower in the *x* direction than in the *y*. This is very simply achieved by first scaling the *x* variable 4-fold and then applying PAM to the scaled covariates, as in Figure 2.8-1(d).

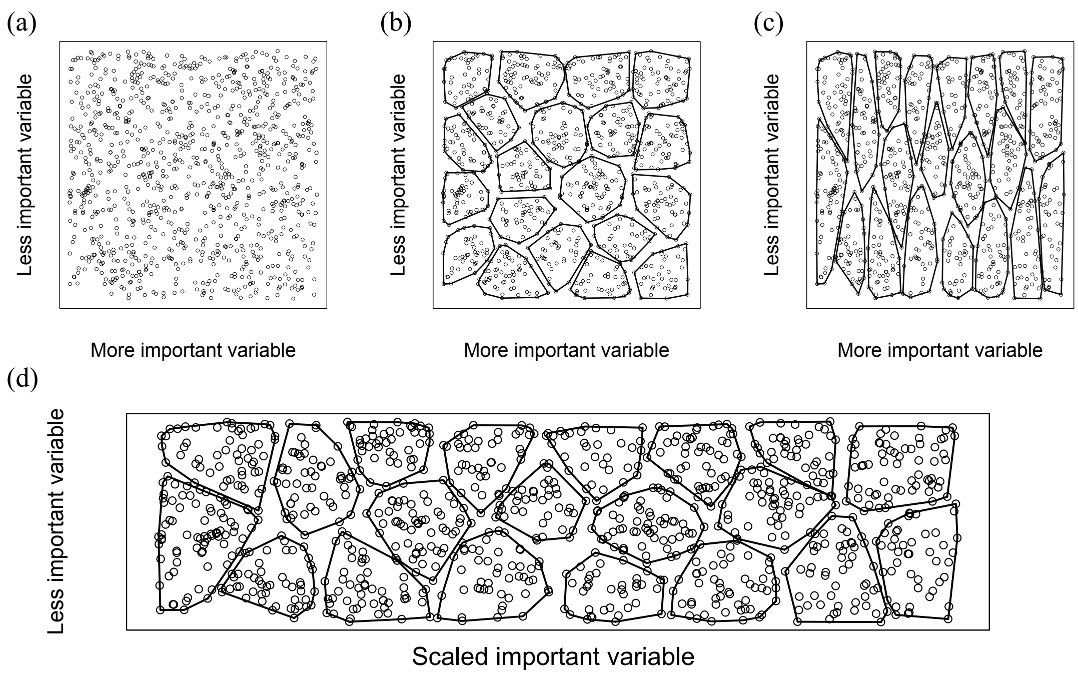


Figure 2.8-1. Partitioning covariate space in two dimensions: (a) 1000 points randomly sampled from the square covariate space. (b) a partitioning into 20 clusters using PAM; (c) a preferred partitioning that accounts for the relative importance of the variables; (d) the partitioning in (c) is achieved using PAM on the scaled covariate space.

The partitioning of the Torres Strait grid cells was an analogous procedure in 28 dimensions. Each variable was scaled so that its 'range' was proportional to its importance. However, unlike in the example, the physical variables were not uniformly distributed across their range and may have extreme outlying values. To guard against the distorting influence of such values, the ‘range’ was taken as that of the middle 95 percentiles. The term “*I95R*” is used here for this range, in acknowledgment of the inter-quartile range, *IQR*, of which this is a generalization. Formally,

$$I95R(v) = v_{(97.5\%)} - v_{(2.5\%)},$$

where $v_{(i\,\%)}$ is the i-th percentile of variable v.

2.8.1.2. Variable Importance

The collated physical variables were quantified on various disparate measurement scales that were unlikely to have any direct relevance to their biological importance. To scale the variables appropriately to inform the stratification, it was necessary to derive an importance value for each variable. The primary component was the biotic importance, but it was also necessary to include a study area adjustment and a reliability adjustment. The biotic importance quantifies the link between the biota and the physical variables and was developed from the detailed species data sampled in the adjacent GBR, but was checked for consistency against the analyses conducted in section 2.7. The study area adjustment was a refinement to the biotic importance to account for potential differences in the range of the physical variables between the Torres Strait and the GBR. The reliability adjustment was a further refinement to reduce the influence of variables that are spatially poorly resolved. These are described in detail below.

Biotic importance I_{bio}

In a previous study, Pitcher et al (2002) performed univariate analyses of 30 benthic statistical assemblages (comprising ~800 species) and 90 single species analyses on 306 sites using a similar suite of physical covariates as explanatory variables. They derived tree models for abundance, logistic regression models for presence/absence data and lognormal regression models for abundance conditional on presence. Their method used model selection to arrive at parsimonious models with some explanatory power and lead to the derivation of a measure of importance for each variable. For each species the relative amount of variation explained by each variable was computed, i.e. the contribution of the variable to the overall R^2. The average of this quantity over all species was defined to be the importance for that variable.

Clearly, the actual dependence of biota on the physical variables is multivariate and highly complex. Moreover, the explanatory power of the physical variables is fairly low, averaging about 30%. Nevertheless, this definition of importance captured the broad pattern over a fairly diverse range of biota. Also it allowed for variation in explanatory power, since species that had low R^2 contributed less to the importance.

The three types of models considered by Pitcher et al were in broad agreement over the ranking of the variables. However, as the tree model approach was most readily cross-validated, these results are reproduced here; the importances are shown in Figure 2.8-2(a).

An alternative but similar approach called random forests (Breiman, 2001) was also considered. In this procedure a bootstrap sample (with replacement) of all 306 sites is taken and a full tree model is fit without pruning. The method for selecting the splitting variable at each node differs from standard trees, where all variables are considered for splitting. In contrast, for random forests, a reduced set of *m* candidate variables, chosen at random, are considered for splitting, and the candidate with the best split is selected as usual. This bootstrap procedure is repeated 500 times to produce a 'forest' of tree models. Predictions can be made from the forest by taking the average prediction from the individual trees. For each sample, roughly 37% of sites are not selected. These 'out-of-bag' sites provide a test data set for estimating (without bias) the prediction error of the forest as a whole. As *m* increases two effects occur: the prediction error of individual trees improves, and the correlation among trees increases. The first acts to reduce overall prediction error but the second acts to increase it. There is therefore an optimal value for *m*, which Breiman has shown to be close to the square root of the total number of variables. Given that 28 co-variates were available for Torres Strait, $m = 5$ was chosen.

The out-of-bag sites also provide a means of defining importance. The importance of variable *v* is the percent rise in the out-of-bag mean sum-of-squared errors when the values of *v* are randomly permuted. This is a relative measure that can be averaged over species. The results are shown in Figure 2.8-2(b).

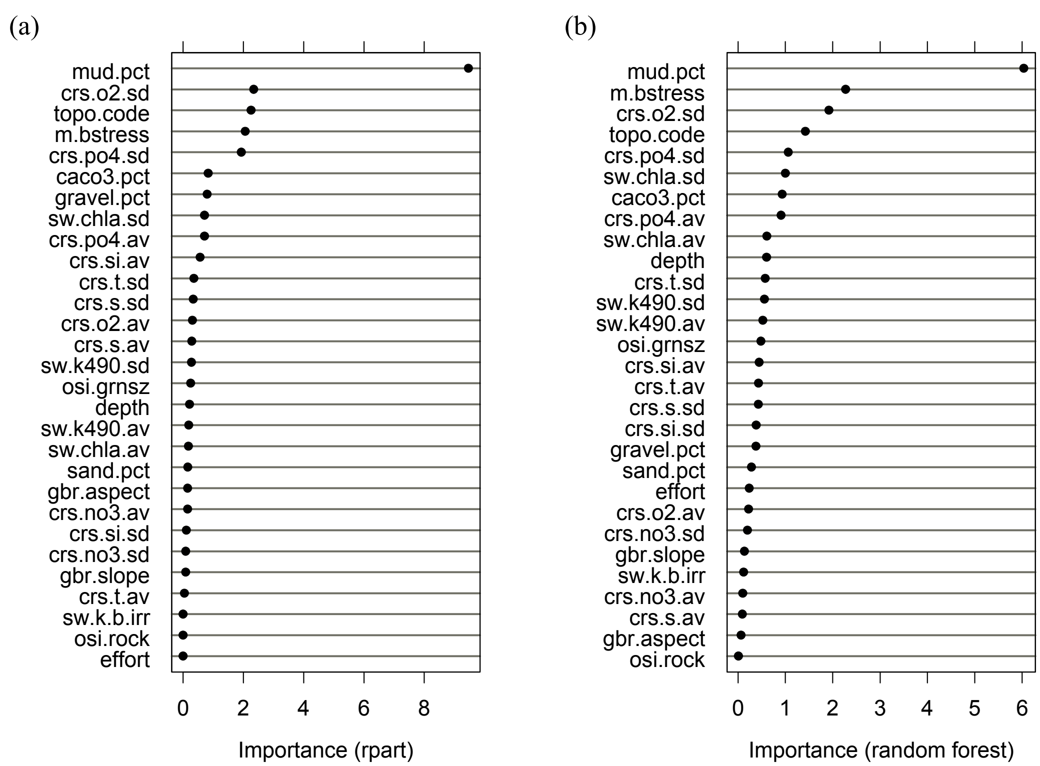


Figure 2.8-2. Variable importance computed by (a) cross-validated trees and (b) random forests.

The results for random forests were qualitatively similar to those for the tree models with slight adjustments to the rankings. The decay in importance with ranking was somewhat smoother for the random forests. Also, because of the use of random candidate variables, the random forests procedure tended to overcome the potential of some variables to dominate other closely correlated variables in the fitting; each variable gets a 'fair go'. Thus, the random forest importances were considered more robust and were used in the stratification approach.

Two variables in the Torres Strait dataset, slope and topographic code, were considered too unreliable and were excluded from the stratification. A third variable, % rock from the OSI auSeabed database that was considered unreliable and found to have very low importance in the in the northern GBR, nevertheless proved moderately important in the Torres Strait from the limited biotic information available (section 2.7). Over several analyses, on average it ranked about the same importance as other OSI sediment attributes, and, as we did not have an importance measure for this variable, we arbitrarily assigned the same importance to OSI rock.

Table 2.8-1. Calculation of adjusted importance I_{adj}: d_{err} is error distance in degrees, I_{bio} is the random forests biotic importance, reliability is $R = (d_{err})^{-½}$, and $I_{adj} = (I_{bio}QR)^{0.6}$.

Variable	Distance d_{err} (°)	Biotic imp. I_{bio}	I95R ratio Q	Reliability R	Adjusted imp. I_{adj}
m.bstress	0.008	2.3	2.6	11.5	12.6
sw.k490.av	0.004	0.5	3.2	16.1	7.3
osi.mud	0.045	6.0	0.9	4.7	7.2
sw.k490.sd	0.004	0.6	2.1	16.1	5.8
sw.chla.sd	0.004	1.0	1.0	16.1	5.3
sw.chla.av	0.004	0.6	0.7	16.1	3.1
agso.dem2	0.037	0.6	1.9	5.2	2.9
cars.o2.sd	0.352	1.9	1.5	1.7	2.6
cars.po4.av	0.912	0.9	3.5	1.0	2.1
osi.crbnt	0.110	0.9	1.0	3.0	1.8
osi.grnsz	0.045	0.5	1.1	4.7	1.7
osi.gravel	0.045	0.4	1.4	4.7	1.7
cars.s.sd	0.312	0.4	3.0	1.8	1.7
cars.po4.sd	0.912	1.1	2.0	1.0	1.6
cars.s.av	0.312	0.1	11.6	1.8	1.5
effort	0.039	0.2	1.4	5.1	1.4
osi.sand	0.045	0.3	1.1	4.7	1.2
sw.d.ben.irr	0.020	0.1	1.4	7.0	1.1
cars.si.av	0.362	0.4	1.5	1.7	1.1
cars.t.sd	0.312	0.6	1.1	1.8	1.1
cars.si.sd	0.362	0.4	1.1	1.7	0.8
dem.aspect	0.037	0.1	1.3	5.2	0.6
cars.no3.sd	0.347	0.2	1.3	1.7	0.6
cars.no3.av	0.347	0.1	1.0	1.7	0.3
cars.o2.av	0.352	0.2	0.4	1.7	0.3
cars.t.av	0.312	0.4	0.1	1.8	0.3
dem.slope	0.037	0.1	0.0	5.2	0.0

Study area adjustment *Q*

The raw importance values from the GBR needed to be adjusted to take into account that the Torres Strait study area is different. Some variables, in particular bottom stress, have a larger range over Torres Strait than over the far northern GBR survey area. Such variables may therefore be more important in Torres Strait. Thus, importances were rescaled in proportion to the ratio of *I95R* between the two regions; the scale factor *Q* (see Table 2.8-1).

The derived importances were also checked by comparisons with analyses of biotic data from the Torres Strait (section 2.7). It was not possible to perform an importance analysis for Torres Strait in the same detail as for the northern GBR study, because the Torres Strait datasets largely consisted of very generalized habitat characterisation, or for the single species-biomass dataset (trawl fishes) very limited coverage. However, a guide to relative covariate importance was available from *F*-values from stepwise discriminant analysis on clusters defined by substrate type or habitat type (section 2.7). The selected variables were in broad agreement with the adjusted importances here. In particular, bottom stress was identified as important for differentiating both substrate and habitat.

Reliability adjustment *R*

The third consideration was that the physical variables had widely differing reliability that needed to be taken into account in the calculation of importance. All the physical variables were available on the design grid of 0.01° cells. However, most variables were interpolated onto this grid based on sample data at a coarser resolution. Therefore, an error distance d_{err} was defined to quantify this spatial imprecision (see Table 2.8-1).

The CARS data were interpolated from a rather limited number of CTD casts, and, for example, in the case of silicate, the average density of casts with silicate data was approximately 1 in 1,300 km², corresponding to an average distance d_{err} of 0.36 degrees between casts. For the effort data, which came from logbooks reporting effort at 6-minute resolution, d_{err} was set to be the average distance from the design grid cell to the centre of the 6-minute effort cell. For the OSI data, d_{err} was set to be the average distance to a sample point from each design grid cell. The SeaWiFS data in their raw form were already specified at the same scale as the design grid; in this case d_{err} was set to be the average distance to the grid cell centre within a grid cell.

The ratio of largest to smallest d_{err} was about 230 (refer Table 2.8-1). It was considered that rescaling over such a large range would be too severe an adjustment and would effectively eliminate the CARS variables from influencing the stratification. Thus, the square root of d_{err} was taken and its reciprocal was defined as the reliability scaling factor *R*.

Adjusted biotic importance I_{adj}

To incorporate reliability, initially the product $I_{bio}QR$ was considered and compared with the study area-adjusted importance $I_{bio}Q$. First, the two adjusted importances were normalized to sum to 1 and sorted in descending importance, as in Figure 2.8-3. The reliability-adjusted importance has much stronger contrast between low-ranked and high-ranked variables, a distortion which was considered unacceptable. Therefore, the reliability-adjusted importance was ‘tuned’ by raising to a power γ. The value of γ was chosen to make the tuned importance match the study area-adjusted importance as closely as possible: $\gamma = 0.6$ gave the minimum sum-of-square differences (compare the blue and green lines in Figure 3):

$$I_{\text{adj}} = (I_{\text{bio}}QR)^{0.6}.$$

Finally, for each physical variable v, the scaled version v_{scaled} that was used in the stratification was defined thus:

$$v_{\text{scaled}} = [v \div I95R(v)] \times I_{\text{adj}}(v)$$

This scaling ensures the *I95R*'s of the scaled variables are proportional to the adjusted importances.

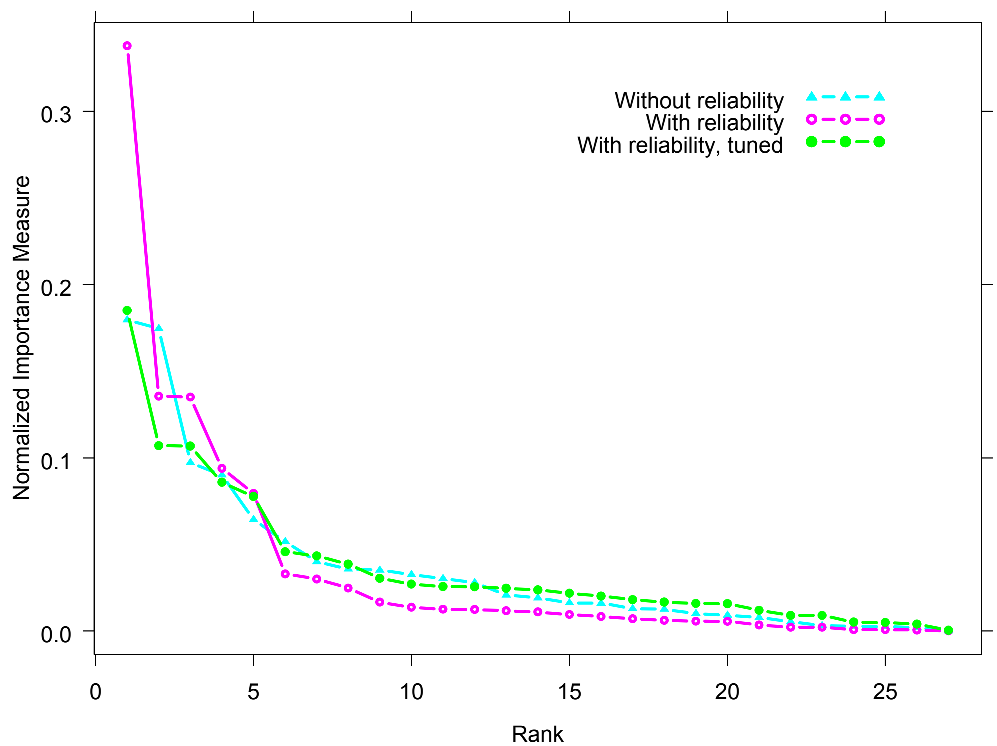


Figure 2.8-3. Importance measures excluding reliability ($I_{\text{bio}}Q$), including reliability ($I_{\text{bio}}QR$), and including reliability, but tuned to match the shape without reliability $(I_{\text{bio}}QR)^{0.6}$. Each version is normalized to sum to 1. The orders of the variables with and without reliability are different.

2.8.1.3. The Clustering Process

Having achieved a biologically informed scaling of the physical variables, the next step was partitioning. However, before proceeding, it was necessary to reduce the dataset for computational manageability and to provide an orthogonal coordinate space for clustering.

There was a certain degree of redundancy among the physical variables. For instance, some variables (phosphate, silicate, chlorophyll A, K490) had a high correlation (>80%) between their average value and standard deviation. There was strong correlation (>77%) among all SeaWiFS chlorophyll A and

K490 measurements, and there were also some negative correlations, e.g. between temperature and silicate standard deviations (–85%). Hence, there was an opportunity to apply data reduction techniques to make the data set more manageable and, importantly, orthogonal prior to clustering.

Table 2.8-2. Variable loadings for the first 7 principal components. Absolute loadings greater than 0.5 are highlighted in yellow, and absolute loadings between 0.3 and 0.5 are highlighted in green. The variables are ordered by adjusted importance. Relative variance is the fraction of the total variance explained by the principal component.

Loadings	**Principal Component**						
Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**
m.bstress	**+0.80**	+0.41	–0.41	+0.02	+0.10	–0.04	–0.04
sw.k490.av	+0.31	–0.47	+0.29	–0.12	**+0.61**	–0.12	+0.19
osi.mud	–0.14	–0.49	**–0.79**	–0.28	–0.02	+0.01	+0.07
sw.k490.sd	+0.27	–0.41	+0.13	+0.19	+0.02	+0.30	**–0.52**
sw.chla.sd	+0.32	–0.36	+0.09	+0.28	**–0.67**	+0.08	+0.16
sw.chla.av	+0.16	–0.19	+0.09	–0.08	–0.19	–0.37	+0.34
agso.dem	–0.08	+0.04	–0.12	+0.35	+0.23	+0.44	+0.36
cars.o2.sd	+0.11	+0.00	+0.12	–0.38	–0.05	+0.49	+0.27
cars.po4.av	+0.05	+0.09	+0.12	–0.44	–0.10	+0.07	–0.06
osi.crbnt	+0.03	+0.06	+0.06	–0.16	–0.07	–0.26	+0.04
osi.rock	+0.04	–0.01	+0.01	+0.04	–0.03	+0.00	+0.04
osi.grnsz	–0.02	–0.06	–0.06	–0.12	+0.03	+0.08	–0.20
osi.gravel	+0.02	+0.06	+0.04	+0.12	–0.02	–0.12	+0.38
cars.s.sd	+0.07	–0.03	+0.08	–0.28	–0.08	–0.07	–0.01
cars.po4.sd	+0.03	+0.06	+0.08	–0.29	–0.14	+0.07	–0.07
cars.s.av	–0.05	+0.08	–0.05	+0.11	–0.13	+0.17	+0.00
effort	–0.03	–0.01	–0.05	+0.00	–0.02	–0.08	–0.14
osi.sand	+0.00	+0.01	+0.06	–0.05	+0.01	+0.07	–0.27
sw.d.ben.irr	+0.01	+0.01	+0.03	–0.15	–0.09	–0.08	–0.11
cars.si.av	–0.02	–0.05	–0.03	+0.10	+0.03	–0.21	–0.13
cars.t.sd	+0.03	+0.05	+0.06	–0.20	–0.09	+0.14	+0.05
cars.si.sd	–0.02	–0.04	–0.03	+0.13	+0.03	–0.23	–0.08
dem.aspect	+0.00	+0.00	+0.00	–0.01	+0.00	–0.01	+0.00
cars.no3.sd	+0.01	+0.01	+0.01	–0.10	+0.01	+0.21	+0.11
cars.no3.av	+0.01	–0.01	–0.01	–0.01	+0.01	+0.05	+0.04
cars.o2.av	+0.01	–0.01	+0.01	–0.01	+0.01	+0.01	+0.04
cars.t.av	+0.01	+0.00	+0.01	–0.03	–0.02	+0.00	–0.02
Relative Variance	**0.52**	**0.21**	**0.12**	**0.04**	**0.03**	**0.01**	**0.01**

Data reduction

Singular value decomposition (SVD) was used to separate the data into principal components, from which we retained the most important components accounting for 99% of the variance in the data. This was contained in the first 14 components, and in fact the first 7 components contained 95% of the variance. SVD decomposed the 41,285 × 28 data matrix X of scaled physical variables into a product of matrices UDV^T, where U was the 41,285 × 28 score matrix, D was the 28 × 28 diagonal matrix of singular values, and V was the 28 × 28 orthogonal loadings matrix. To project the data into a smaller dimensional space, but retain the relative distances of the data, a new data set was defined as UD^* where D^* (28 × 18) consists of the first 18 columns of D. This data is equivalent to rotating the scaled data by V (i.e. XV) and projecting into the 18-dimensional subspace spanned by the first 18 columns.

The effect of this transformation was observed by examining the variable loadings V. The rows of V correspond to the original variables and the columns to the principal components. Large values (on the scale 0 to 1) indicate alignment of the variable with the principal component. The important variables should be expected to have high loadings on the first few principal components, and the less important variables to have higher loadings on the later principal components.

The loadings on the first seven principal components are shown in Table 2.8-2. Principal component 1 was mainly associated with bottom stress, whereas the second component was associated with various SeaWiFS measurements, as well as with bottom stress and mud. Because the 3 most important SeaWiFS variables are highly correlated with one another, they have similar loadings. The 3rd component was principally mud, the 4th introduced depth and two of the CARS variables, and the 5th component comprised mainly the difference between average chlorophyll A and K490.

Including geographic constraints

Another important consideration was whether spatial position should be included in the stratification. In the absence of covariate information, it would be usual to stratify entirely on geographical position, making each stratum simply connected. On the other hand, if we ignore geography completely, and base the stratification only on physical covariates, then the strata will tend to be fragmented in geographical space. This would not necessarily be a bad thing. However, if the fragments become very small then the quality of the stratification may become degraded by spatial uncertainty in the covariates themselves.

This issue was examined in the design of the GBR Seabed Mapping survey and the conclusion was that it would be prudent to include a small amount of geography (Pitcher et al, 2002). Using the recommendations from that study, latitude and longitude were scaled equally so that the *I95R* of the scaled latitude equalled 0.25 times the *I95R* of the first principal component of the rotated data. The scaled spatial variables were included as extra dimensions in the clustering, and their effect was generally to prevent the clusters becoming too highly fragmented in space.

The PAM and CLARA algorithms

The clustering algorithm “partitioning around medoids” (PAM) of Kaufman and Rousseeuw (1990), which is implemented in Splus, was used to cluster the physical dataset. The PAM algorithm is a robust alternative to the k-means algorithm. It uses a distance matrix and the number of clusters must be specified. Whereas K-means minimizes distances to the average for the cluster, in PAM, each cluster contains a *medoid* that is the cluster member whose summed distance to all other cluster members is a minimum. The medoid is a kind of generalized median for multiple dimensions; it is to this that the algorithm owes its robustness. The algorithm works by searching for clusters that minimize the total distance to cluster medoids.

PAM is not immediately useable for large data sets, because the size of the distance matrix becomes unmanageable. Therefore Kaufman and Rousseeuw's CLARA algorithm, which is an implementation of PAM for large data sets, was applied. This works by first selecting a random subset of the data, then applying PAM to generate a clustering, and finally assigning the remainder of the data to the nearest cluster in the subset. The procedure is repeated many times to give several candidate clusterings, from which the candidate that minimizes the total distance to cluster medoids is chosen. The algorithm can be tuned by adjusting the subset size and the number of repeats, both of which should be as large as practicable.

Further, a weighted version of CLARA was developed specifically for this project. In this implementation, each initial subset was selected with non-uniform probabilities or weights, which enabled the clustering to be influenced to some extent to seek rarer physical environment strata, as explained below.

Two-stage partitioning

The partitioning was performed in two stages. In stage 1, we generated an initial coarse partitioning of the entire data set into 50 'superclusters', or primary strata. Then in stage 2, each supercluster in turn was partitioned, generating a total of 440 subclusters.

The initial reason for having two stages was computational efficiency. For k clusters and n observations, the computation time is of order kn^2; but if $\sqrt{k}$ superclusters was computed first, and then $\sqrt{k}$ subclusters (on average), the computation time can be reduced to the order $\sqrt{k}\, n^2$. In fact stage 1 is the most computationally intensive stage, taking of order $\sqrt{k}$ times longer than stage 2. Even for 50 superclusters, which was somewhat larger than $\sqrt{440}$, the computational saving was substantial. This was an important consideration when developing a method, particularly where many subsets of the data must be run.

However, the main reason for using a two-stage method was that it allowed more control over the partitioning. This was because at stage 2, it becomes possible to choose the number of subclusters within each supercluster, subject to a total of 440. In particular, it was possible to raise the level sampling effort into uncommon and rarer areas in covariate space, that may be potentially more interesting in terms of biota, at some expense to common areas.

Choosing the number of subclusters

After stage 1, there were 50 superclusters of various sizes ranging from 62 to 2113 cells. Then it was important to determine how to optimally distribute the 440 subclusters among the 50 superclusters.

In order to answer this question, initially the following hypothesis was adopted: clusters with large numbers of cell members tend to be more homogeneous and represent commonness, compared with small clusters. Support for this hypothesis can be seen in Figure 2.8-4 for a synthetic bivariate normal data set. The larger clusters (in terms of numbers of cells) near the middle have smaller bivariate space (i.e. are more homogeneous), whereas the more heterogenous clusters around the fringes tend to have fewer points (i.e. are smaller clusters).

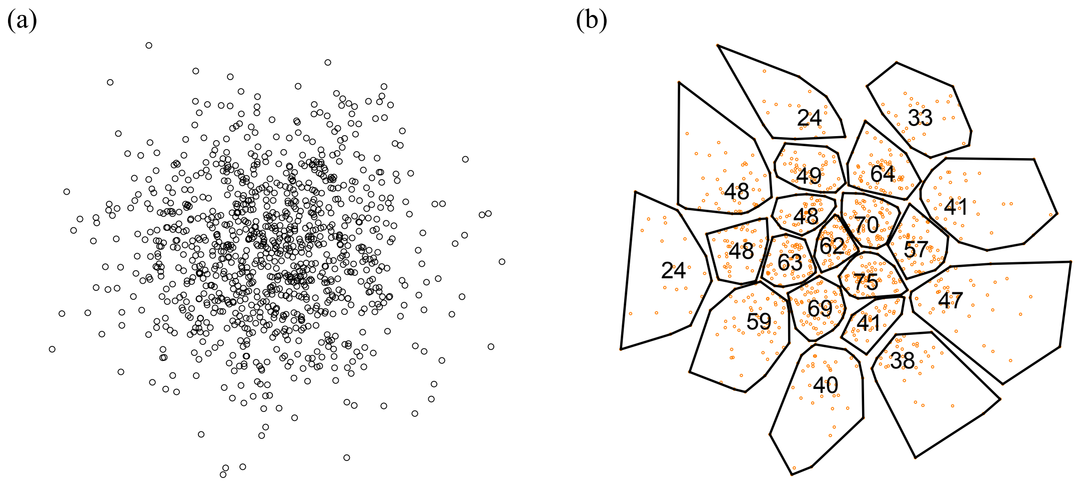


Figure 2.8-4. (a) Bivariate normal distribution of 1000 points. (b) Partitioning into 20 clusters using PAM. Each cluster is labeled by the number of points in the cluster. The more populous clusters tend to be tighter and so more homogeneous.

Therefore the stratification strategy should be such that the density of sampling should be lower for larger superclusters, i.e. the number of subclusters N_{sub} depends sub-linearly on the supercluster size S. This issue also arises in the context of species-area curves, where the number of species increases with area sampled, but less than linearly. In fact, for species-area curves a square-root relationship is sometimes used. Following this principle, the initial approach could be $N_{\text{sub}} \propto \sqrt{S}$.

This approach would attempt to bias the sampling away from common sites towards rarer, perhaps more ‘interesting’, sites so that they also can be sampled adequately. Nevertheless, the square-root approach provides a somewhat crude approximation to the amount of ‘interest’ in a supercluster, relating it simply to the size of the supercluster, without regard to its contents. A better approach would be to quantify the interest as a sum over the interest in individual sites. For this, it was necessary to define the interest at a site.

The more common sites are those lying in high-density areas of covariate space. Since common sites will be well sampled in any case, it was reasonable to define ‘interest’ as some inverse power of density. However, computing the density in more than 2 dimensions is difficult; instead the one-dimensional densities of each physical variable was considered separately. Suppose d_{vi} is the density of variable v at site i, normalized so that the total density over all sites is 1. Then we define the interest w_i at site i as the variable importance-weighted sum,

$$w_i = \sum_{v=1}^{28} I_{\text{adj}}(v) d_{vi}^{-a} \ ,$$

where $a > 0$ is a parameter to be chosen. Then define the interest of a supercluster as the total interest over sites within the supercluster, and choose the number of subclusters to be proportional to this quantity. That is, for the k^{th} supercluster $C(k)$:

$$N_{\text{sub}}(k) \propto \sum_{i \in C(k)} w_i \ .$$

The density is estimated from the 41,285 values using a gaussian kernel whose width is calculated by biased cross-validation (Scott, 1992). As an example, Figure 2.8-5 shows the true density (total area = 1) for bottom stress. The bulk of the distribution lies below 0.5; whereas previous experience has demonstrated that sites above 0.7 were of particular interest for epibenthic fauna (see section 2.7).

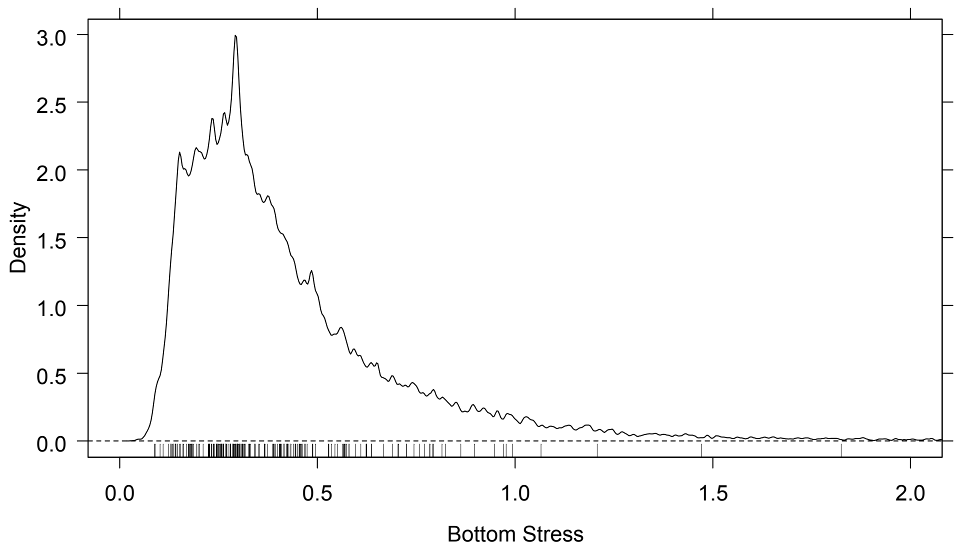


Figure 2.8-5. Density of bottom stress estimated by a gaussian kernel of width 0.018 calculated using biased cross-validation. Also shown is a 'rug' of values for 200 randomly selected sites.

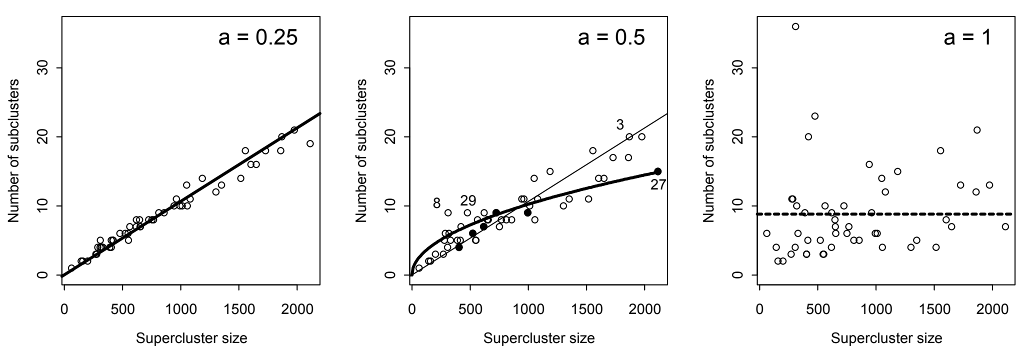


Figure 2.8-6. Number of subclusters vs supercluster size for 3 different values of the exponent *a*. The sloping line corresponds to $N_{sub} \propto S$, the curve to $N_{sub} \propto \sqrt{S}$, and the horizontal line to N_{sub} = constant. In the middle plot, four superclusters are labelled for later reference in the text, and superclusters denoted by a black dot are mapped in Figure 2.8-8.

Figure 2.8-6 shows the relationship between number of subclusters and supercluster size for *a* = (0.25, 0.5, 1). For the case *a* = 0.25, the relationship was almost linear; this was barely distinguishable from the case *a* = 0, in which all sites had equal interest. At the other extreme, case *a* = 1 flattened the relationship, making number of subclusters nearly independent of supercluster size and too sensitive to individual high-interest sites within a supercluster. The intermediate case *a* = 0.5 was close to the square-root proposal discussed earlier and provided the required increased sampling of rarer sites without unacceptable under-sampling of common sites. This value for *a* was used as it provided an improved stratification adjustment compared with the initial square-root proposal.

There was a concern that, at the superclustering stage, rarer sites might be missed in the CLARA random subset selection stage since rare sites would be unlikely to be selected in a small random subset and, as a consequence, the superclusters could be too large and homogeneous. Such superclusters, being comprised largely of common sites, would have fewer subclusters, and so there would be less chance of isolating the rarer sites into their own subclusters. Two steps were taken to reduce this risk. Firstly, we computed more superclusters than was computationally optimal (ie. 50 > √440). Thus, superclusters would be smaller, allowing for better detection of heterogeneity within a supercluster. Secondly, a weighted version of CLARA was developed, with site interest w_i as the weighting. Thus, rarer sites were more likely to have a chance at being chosen in the random sample of the algorithm, and therefore more likely to seed a separate supercluster.

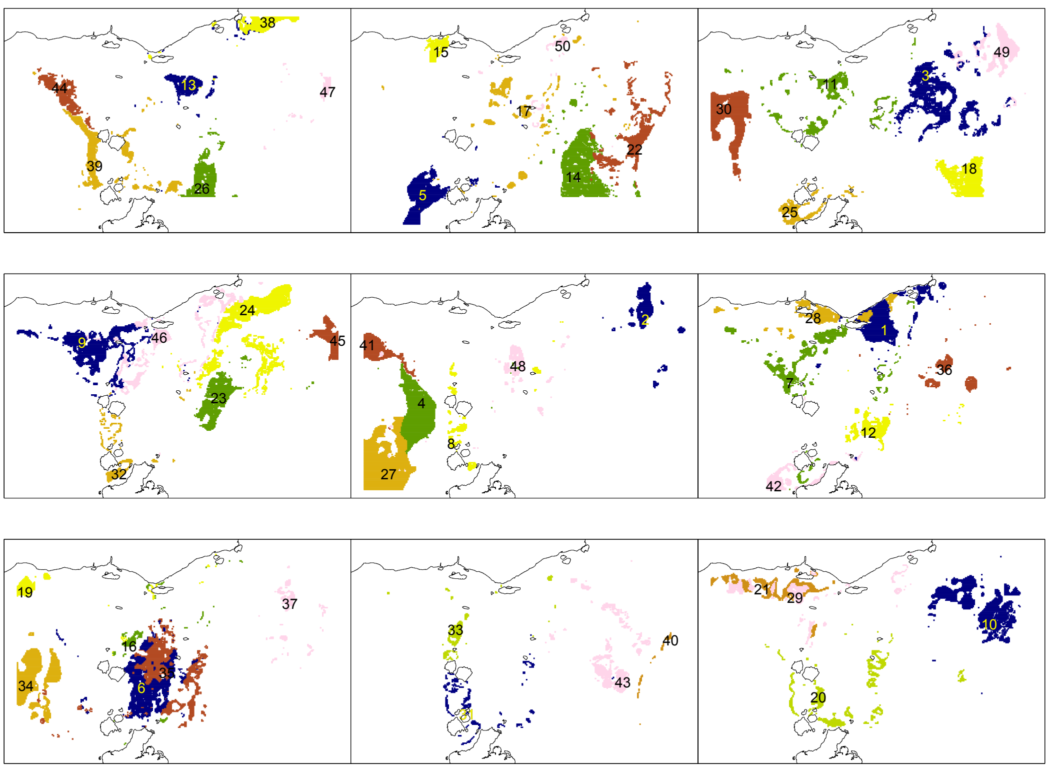


Figure 2.8-7. The 50 superclusters in geographical space. The clusters have been separated into nine panels in order to make them distinct. The largest cluster (27) is in the centre panel and the smallest (40) is in the bottom centre panel. Clusters 3 (top right) and 24 (centre left) have the largest number of subclusters (20).

Figure 2.8-7 shows maps of the resulting 50 superclusters after the first stage of clustering. Because the clustering was in covariate space, there was no guarantee that the clusters would be simply connected in geographical space, even though latitude and longitude were included as covariates. Indeed some clusters, especially those around the centre and north, were highly fragmented (e.g. 17, 11, 46 and 20). Despite their geographical appearance, these clusters' sites have similar physical characteristics. In the other hand, some clusters, especially those in the west, southwest and southeast, are fairly spatially contiguous (e.g. 14, 30, 18 and 27). Part of the reason for this is that the covariate values in these regions are based on spatial interpolation from sparse data points, and so the covariates vary smoothly in space.

Figure 2.8-8 and Figure 2.8-9 show the subclustering within some of the superclusters. In Figure 2.8-8, all the superclusters have an 'average' amount of subclusters, as indicated by the middling locations of the black dots within the vertical spread in Figure 2.8-6; the exception is supercluster 27, which has below-average sampling density. All these superclusters are spatially fairly contiguous, and the same is also true of their subclusters.

The cluster medoids (indicated by red dots) are the most central member of the cluster in covariate space. Often, but not always, the medoid is close to the geographical centre of the cluster. The medoid is a very useful by-product of the PAM/CLARA algorithm and, because it is always a cluster member, the medoid can be used as a representative of the cluster. This is a distinct advantage over the mean, which, for an irregularly shaped cluster, might not lie near any of its members.

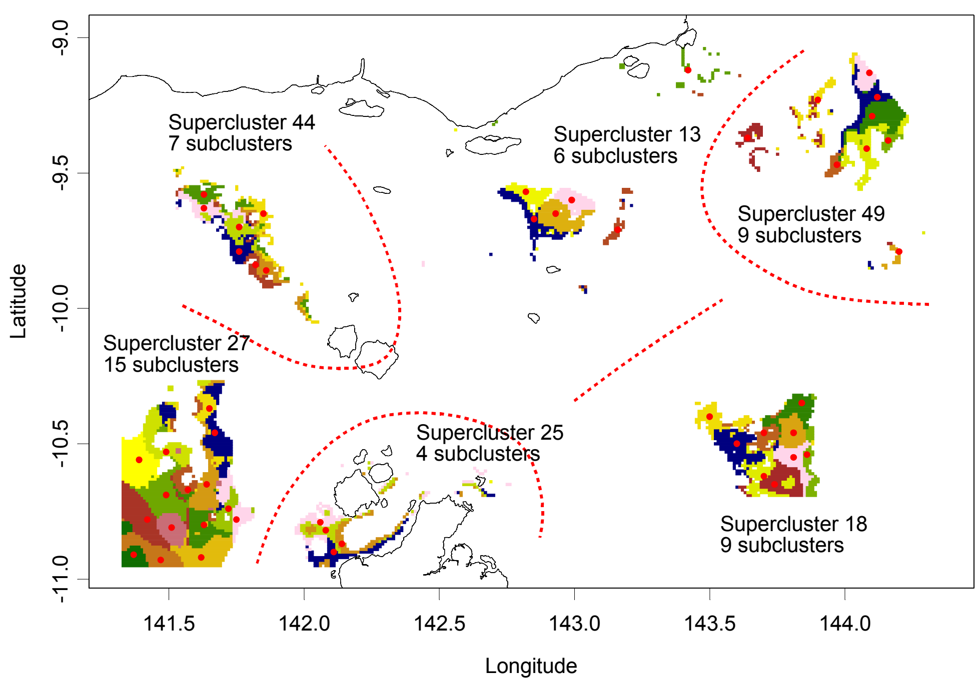


Figure 2.8-8. Six fairly compact superclusters in geographical space and the subclustering within them. The subcluster medoids are indicated by a red dot.

Three superclusters (3, 8 and 29) with above-average sampling (at top of the vertical spread in Figure 2.8-6) are mapped in Figure 2.8-9. These superclusters are much more fragmented, and some of the subclusters are also fragmented. Some of these clusters lie in the part of covariate space with high bottom stress, and, since bottom stress varies rapidly in geographical space, the clusters themselves fragment over smaller scales. In particular, for supercluster 8, the clustering has four distinct well separated geographic areas and assigned 2 to 3 subclusters in each.

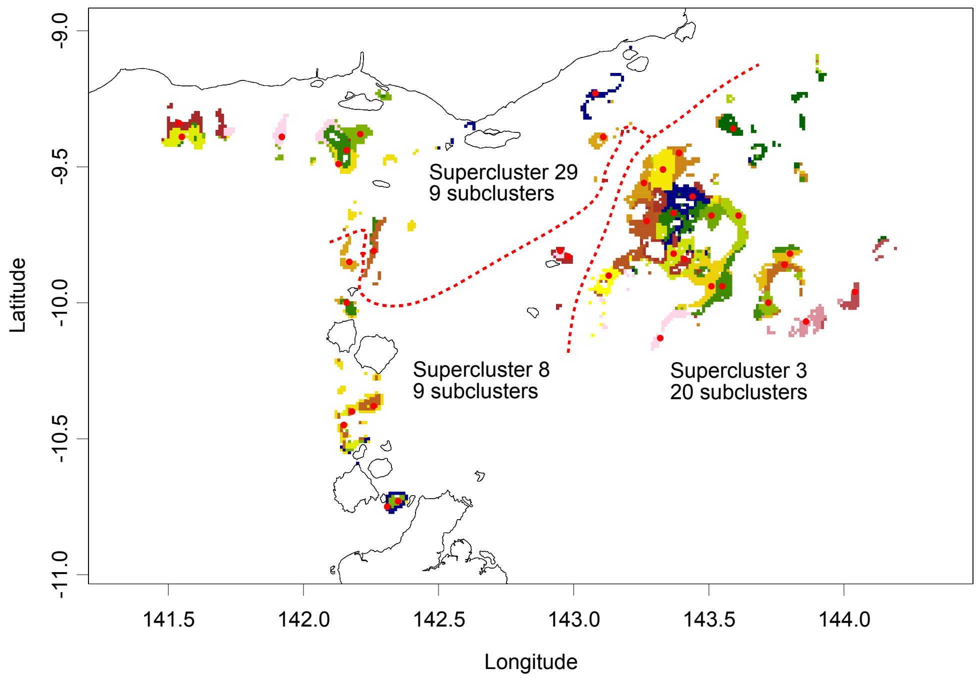


Figure 2.8-9. Three fairly fragmented superclusters in geographical space and the subclustering within them. The subcluster medoids are indicated by a red dot.

Assessing the resulting stratification

There is no unequivocally optimal approach to survey design. For instance, in the two-dimensional example of Figure 2.8-1, we could have used the k-means algorithm instead of PAM, and the resulting partitioning, which would have been different, would nevertheless have been a quite reasonable alternative. Although there is no single 'right answer', it is nevertheless necessary to establish that the resulting partitioning is reasonable. There are several ways to assess the stratification.

First, the strata were mapped. We have already partially shown this in Figure 2.8-7 to Figure 2.8-9. However, a map of all 440 strata would be rather overwhelming and very difficult to interpret. A clearer alternative was to plot the locations of the stratum medoids, since each medoid was in some sense the most typical representative of the stratum. In fact, the choice of medoids as actual survey sites would be a quite reasonable candidate sampling strategy and could be called "medoid sampling".

Figure 2.8-10 shows the medoid sites against the background of all possible sites. This would provide acceptable general coverage of the entire Torres Strait region. However, the sampling would be finer in some parts (such as the north, the east and around the longitude of Thursday Island) and coarser in other parts (the west, southwest and southeast). This was consistent with expectations and a desirable property of the stratification, which was being sought. The more heavily sampled regions are areas with either high bottom stress (around the islands and outer reef) or high chlorophyll A (the north), both highly important variables.

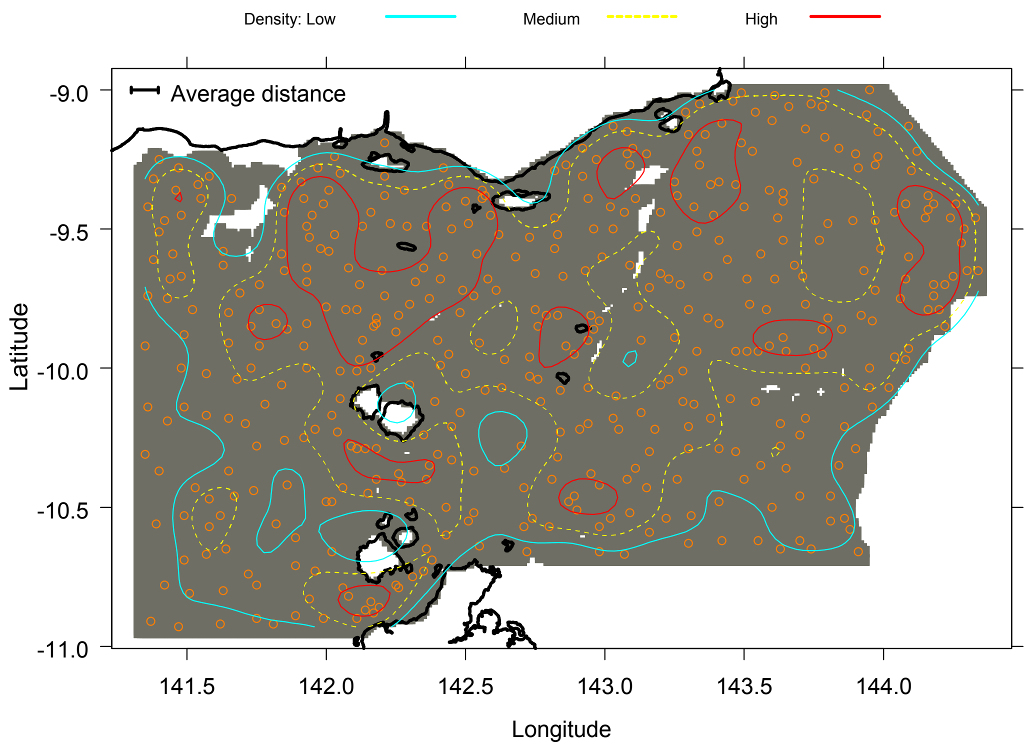


Figure 2.8-10. Medoids of the 440 strata. The bar labelled 'average distance' is the minimum spacing that would lie between 440 points if they were regularly spaced. The contour lines show the kernel density estimate of the medoids. The low, medium and high densities are, respectively, 0.66, 0.88 and 1.10 points per average distance squared. The 41,285 possible survey sites are indicated by the grey background.

The second way to assess the stratification was to examine the expected distribution of the physical covariates at the sample sites. Again, the medoid sampling can be used as a representative sampling. Figure 2.8-11 shows the density of bottom stress, average K490 and percentage mud over the stratum medoids compared to over all 41,285 sites. Transformed scales have been used, on which the distributions were roughly symmetrical, to make the comparison clearer. For completely random sampling, the density would be similar to that over the full data set. But in the medoid sampling, there was relatively less sampling in the high density (common) areas, and more sampling in the tails (rarer areas), which was the objective of the stratification. For bottom stress, more sampling is put into sites with values above 0.7, at the expense of the more common sites with values in the range 0.25–0.5. For K490, sacrifices are made in the range 0.075–0.1 to increase the sampling both above and below this range. For mud, the sampling is increased above around 40% at the expense of areas with no mud at all.

The representativeness of the medoid sampling can be checked by comparing its density with densities arising from many random samplings of the stratification. Figure 2.8-11 also shows confidence intervals for the density, which were obtained from the 5th and 95th percentiles of the pointwise densities of 20 random samples. Although there were small biases, overall the medoid-sampling density was fairly representative of the range of possible densities arising from stratified sampling.

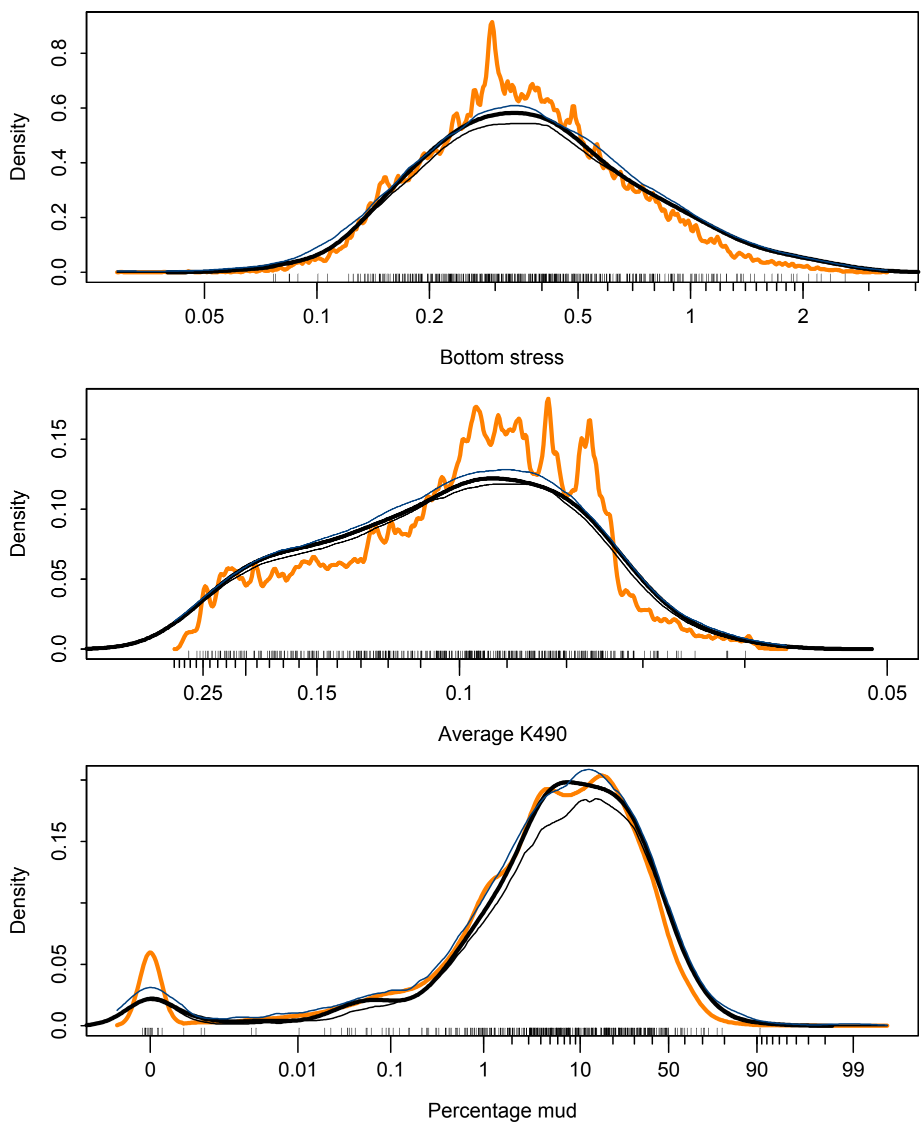


Figure 2.8-11. Distribution of the most important physical covariates on the subcluster medoids (black) and on the full Torres Strait data (orange). The thin curves are 90% confidence intervals for the density. For clarity we show covariates on a log scale for bottom stress, an inverse scale for K490 and a logit scale for mud. Also shown is a rug of the 440 medoid values (jittered around 0 for mud).

A similar approach was applied to the spatial distribution of medoid sites, by averaging the kernel density over repeated samplings (Figure 2.8-12). The results again show that the medoid sampling was fairly representative, although the medoid sampling was slightly denser around Thursday Island and just northeast of Cape York.

As a final assessment, the relationship between covariate value and the probability of selecting a site (the reciprocal of the stratum size) was examined. This is shown in Figure 2.8-13 for bottom stress, average K490 and percentage mud. Although there was considerable scatter in the probabilities, sites with rarer values in these covariates tended to have higher probability of selection. This was confirmed by the locally smooth regression (which has been applied to all 41,285 points, not just to those displayed). This also shows that the rarer physical environment combinations tended to reside in small strata (<100 cells), whereas large strata (~250 sites) hold the more common sites. This is especially evident for bottom stress.

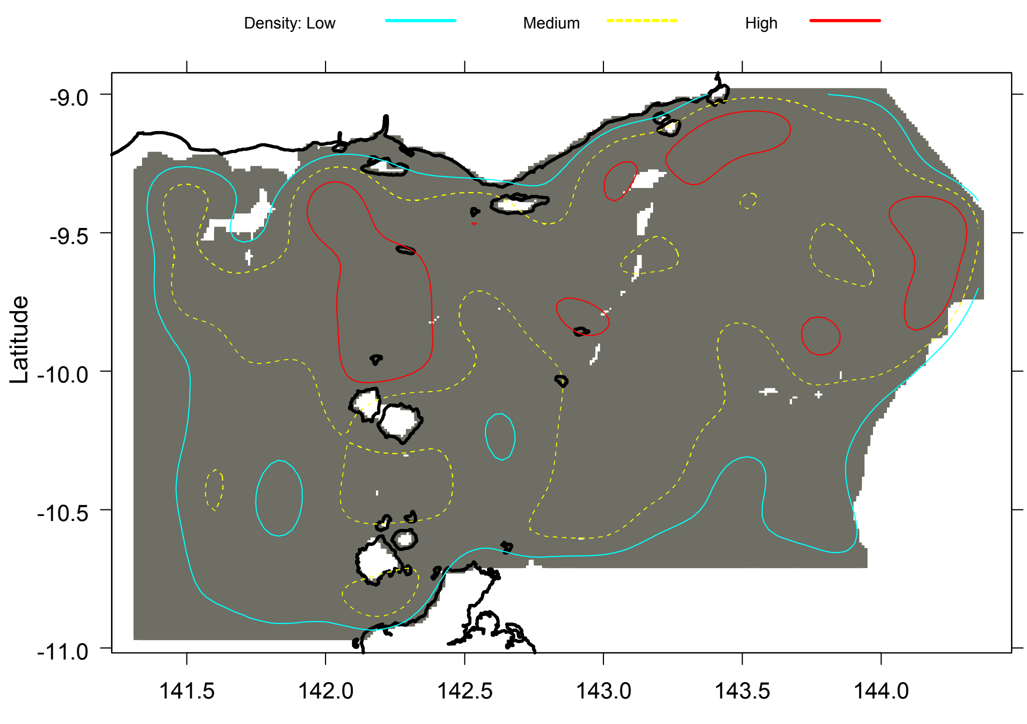


Figure 2.8-12. Average kernel density estimate of the sample sites over 20 independent random samplings. The low, medium and high densities are, respectively, 0.66, 0.88 and 1.10 points per average site separation squared. The 41,285 possible survey sites are indicated by the grey background

Defining trawl substrata

The above has described how 440 substrata were defined from which benthic sampling sites may be chosen. However, somewhat fewer of these same sites (324) were to be selected for trawl sampling and it was necessary to identify which would be the most representative. Although one method would be to simply choose the 324 sites at random, an approach that took advantage of the existing stratification was preferred, to ensure that the selection was heterogeneous. The approach taken was to go back to the superclusters and recompute the number of subclusters required per supercluster to give a total of 324, using the same methodology as before. On average the number of trawl subclusters was about three-quarters (324/440) the number of original subclusters. For instance, supercluster 3, which had 20 original strata, had 15 trawl strata. It was not feasible to try to cluster the sites into trawl subclusters, because there was no way to prevent the original strata straddling several trawl strata. Instead, it was necessary to cluster the sites such that all sites in an original stratum remain together.

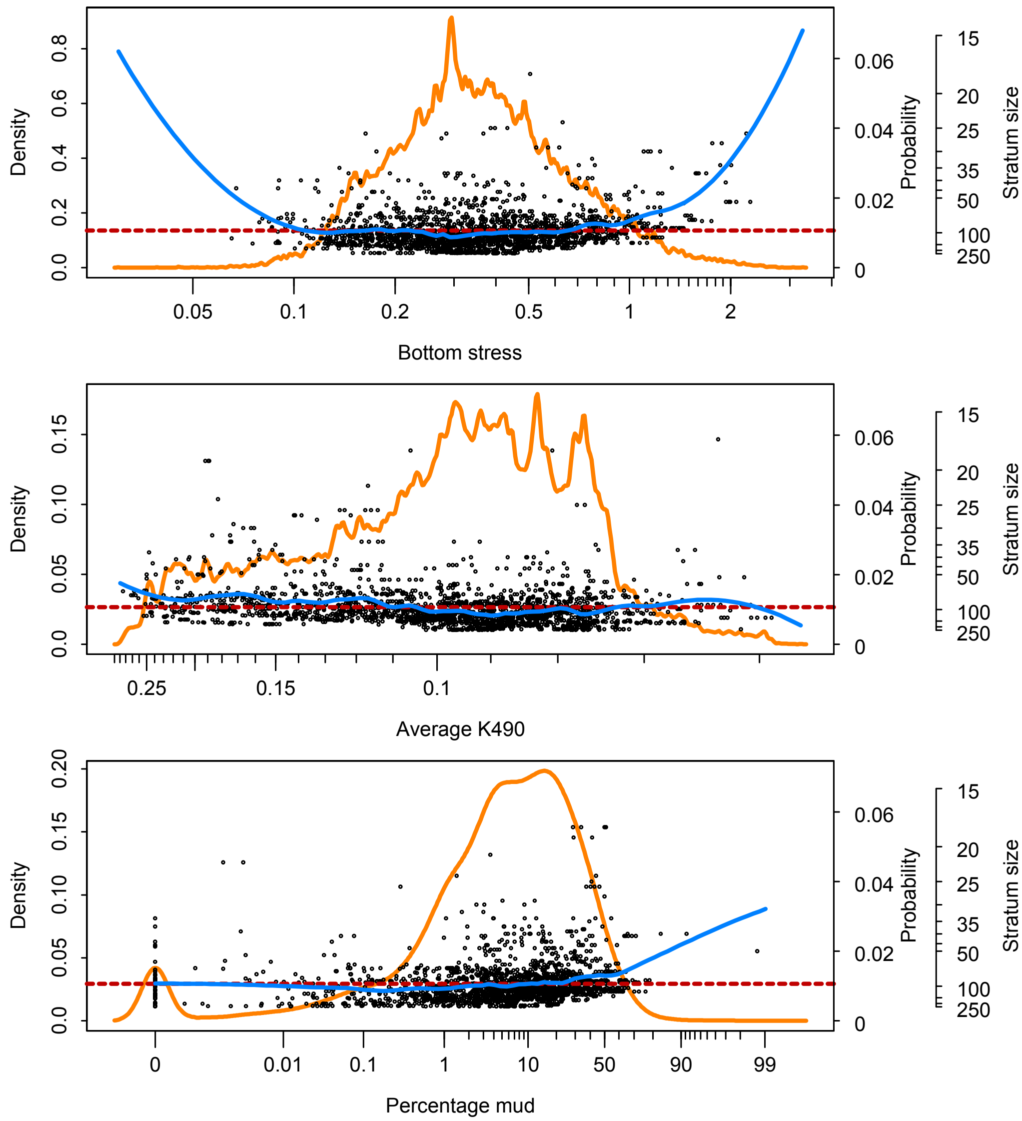


Figure 2.8-13. Local regression smooth of probability of site selection (blue line, right axis) with covariate density (orange line, left axis) for reference. Probabilities for 2000 randomly chosen points (independent of the stratification) are also shown. The dashed horizontal line is the average probability. For clarity we show covariates on a log scale for bottom stress, an inverse scale for K490 and a logit scale for mud.

The simplest way to do this was to cluster the stratum medoids. It was appropriate to use the medoid to represent its stratum as a whole because the medoid lies centrally within the stratum in co-variate space. Since there were at most 20 medoids to cluster, the calculation was computationally trivial. For example, in supercluster 3, substrata 1, 3 and 17 were amalgamated into one trawl cluster, substrata 7 and 15 into a second, substrata 4 and 10 into a third, and substrata 8 and 16 into a fourth, while the other 11 trawl clusters coincide with the original substrata.

After the medoids were clustered, each medoid's trawl substratum number was assigned to all other cells in its substratum. Thus, each cell now belongs to both a substratum and a trawl substratum. Thus for any selection of 440 benthic survey sites, the trawl sites could be selected from these by choosing one from each trawl stratum, either at random or by other objective.

2.8.2. Sample Site Selection

In the previous section, the notion of medoid sampling was raised to illustrate the stratification. Medoid sampling would be a perfectly reasonable method of selecting sites that would deliver the "most typical" cell, with respect to physical covariates, within each of the strata. A random selection of sites from within each of the strata would also be an acceptable method. However, the stratified random method has a relatively high risk of selecting cells too close together and too far apart, creating clumps and voids in the coverage of the region, when in fact a representative coverage that also takes account of the spatial autocorrelation distance was desired. Considering that strata were often fragmented into patches of varying numbers of cells, including single cells, there was also a high risk of selecting isolated cells as sites — these would be less likely to be representative of their stratum due to errors in the covariates. A site selection method that avoided these problems as much as possible was sought.

Initially, a weighted random selection was used, with weights dependent on the spatial geometry of the particular patch within each stratum that the cell belonged to. Cells with fewer neighbours of the same stratum and on the edges of patches (i.e. geographically close to a different stratum) were given less weight, whereas sites in the middle of patches were given more weight. This strategy was intended to reduce the possibility of a site being unrepresentative of its stratum due to errors in the covariates and to avoid selecting adjacent sites. Examination of several weighted random selection options indicated quite a number of adjacent cells being selected and a number of excessively large voids between selected sites. Consequently, a method that more stringently avoided selection of adjacent cells and voids was needed.

The method finally used did not include any deliberate random jittering of site selection. For each of the 440 benthic strata, first all those cells that had the maximum number of neighbours and were the maximum distance from the edge of patches were selected. For many of the strata, several cells met these criteria (total 1698). To remove duplicate cells within strata, the cell with the minimum medoid distance was selected. In about a quarter of cases, the actual medoid cell was selected. This strategy maximized the co-variate representativeness and spatial regularity of the selection, within the desired constraint of the stratification, and minimized the likelihood of clumps and voids, and adjacent, edge and isolated cells.

As described in the previous section, fewer sites could be sampled by trawl methods, so the 440 benthic medoids were clustered to provide 324 most representative options. Of these, 240 were a one to one match with their benthic strata, so no further selection was needed. However, in 84 cases, a trawl site had to be selected from 2-4 benthic site options. In these cases, the benthic site chosen to be sampled by trawl also was, to maintain spatial coverage, that which belonged to the largest patch in its cluster.

The sites selected are mapped in Figure 2.8-14. This site selection process provided a good compromise between coverage of the range of biologically important physical environments in TS and evenness of spatial coverage, given the limited number of sites that could be sampled and the inadequacies of the data available for the stratification. Such a coverage could not be achieved with regular grid sampling or completely randomised sampling.

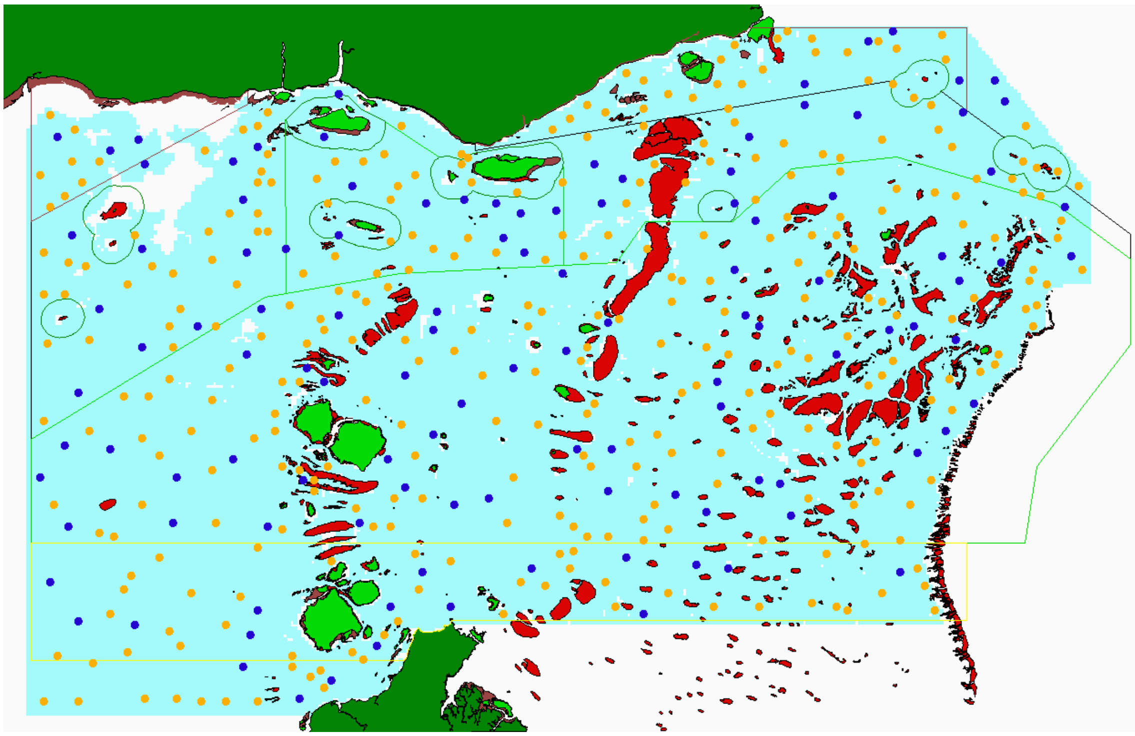


Figure 2.8-14 Map of the sites selected for sampling the seabed in Torres Strait, overlaid on a background of all cells included for possible selection (light blue). White areas were excluded as outside the study area or too shallow for navigation. ●: sites for benthic and trawl sampling, ●: sites for benthic sampling only.

2.8.3. Mapping the Physical Characterization

The biologically informed stratification developed in section 2.8.1 is a physical characterisation of Torres Strait that can be considered an *a priori* surrogate for patterns in seabed biodiversity assemblages, to be tested and improved by the future sampling to be conducted by the CRC-TS Seabed Mapping Project. Given the likely interim utility of this information, a method of representing this complex multi-variate data in a single map was sought.

2.8.3.1. The Colour Key

The objective was to produce a map of the Torres Strait with similar colours representing similar physical environments, which could be expected to have similar benthic biotic assemblages. The colour mapping should encompass as much information as possible in a reduced form — this was achieved by deriving a colour key from the first and second principal components of the biological importance weighted covariate data used in the stratification. A biplot of the principal components and physical variable vectors would provide a key to the environmental characteristics of the map. Particular directions in the biplot that corresponded to important covariates, should be coloured in an intuitive manner. Red was used to denote high bottom stress and green to denote high average chlorophyll A (which correlated with K490). Blue corresponded with depth. High density areas of the biplot (common areas) should have a neutral colour such as white or grey.

A further desirable property of the colour key is that it should cover the data space compactly, to avoid large areas of the key having no data and wasting part of the colour space. The colour key should therefore be shaped to conform to the distribution of the data in principal components (PC-)space. This was done by mapping a circular colour disk to a simply connected region enclosing the data. In order to do this, it was necessary to first define a boundary of the data in PC-space. One way to do this was to find the convex hull; however, for the Torres Strait data, this included a void in which no data existed. Instead, a more compact boundary was found by computing a two-dimensional kernel density function and delineating a contour of sufficiently low density. The boundary is partly concave.

Having defined a boundary, there were two alternative methods for mapping the colour disk to the region inside the outer density contour boundary: polynomial mapping and conformal mapping. The polynomial mapping was found to be more flexible but because of the partly concave shape there was not always a one-to-one mapping between PC-space and colour space, and it was non-trivial to invert from PC-space to colour.

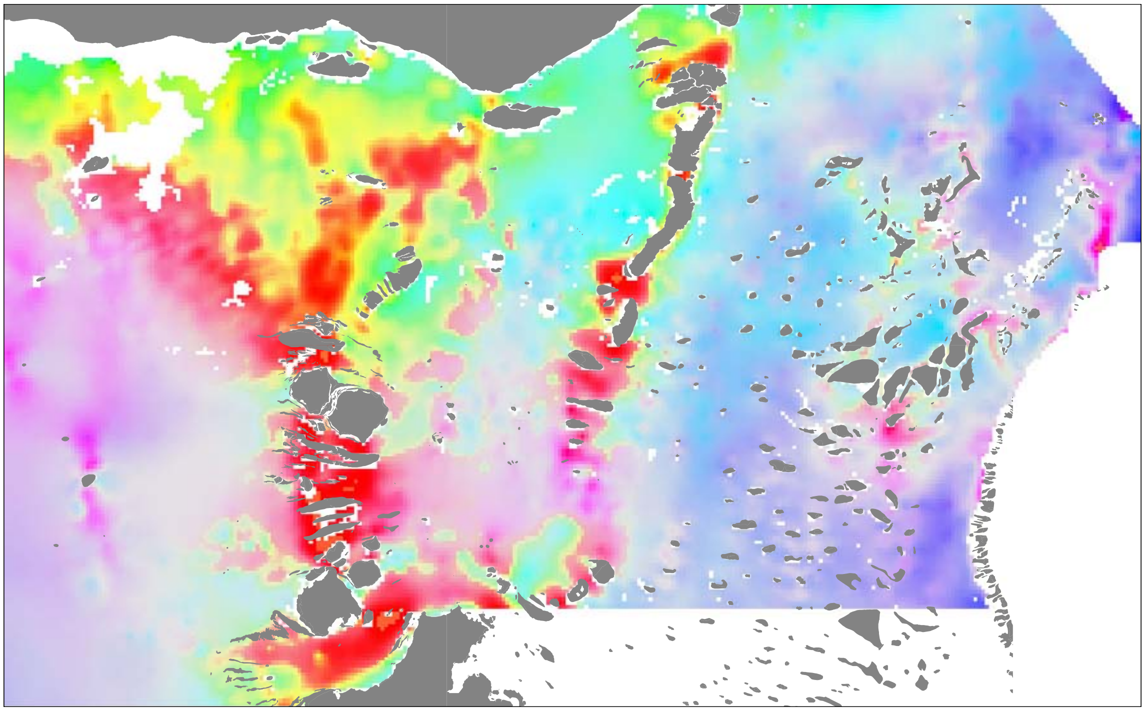
2.8.3.2. Conformal Mapping

The conformal mapping method originates from complex number theory. A mapping from the colour disk to a simply connected polygon is expressible as a complex integral, whose parameters must be estimated by a non-linear algorithm. Trefethen (1980) provided a FORTRAN program to compute this integral. An interface to this code was developed that runs in R. Conformal mappings have certain benefits (such as local preservation of angles) but most importantly they are guaranteed to map the interior of the colour disk to the interior of the polygon (i.e. the mapping will not stray outside the boundary).

As with the polynomial method, the point in PC-space that the centre of the disk was mapped to was specified. The matching of points on the edge of the disk with vertices of the polygon was done by the non-linear algorithm. In order to match intuitive colours to the desired directions in PC space, it was necessary to impose a further transformation on the colour disk, which amounted to an angular stretch and shift. This was done using a periodic piecewise linear function of the angle. To complete the physical characterisation map, each grid cell must have a colour associated with it. Hence, the colour key mapping must be inverted, so that points in PC-space become mapped to points in colour space. This inverse mapping is available in FORTRAN code (Trefethen, 1980).

The resulting physical characterisation map of Torres Strait is shown in Figure 2.8-15. High bottom stress areas were coloured red, high Chlorophyll/K490 areas green, and the mud direction was coloured cyan. Sites coloured cyan have high levels of mud. Deeper areas tend to be blue.

The colouring of a map to highlight different covariates can be highly effective at illustrating similar and different physical environements, especially when the colour space has been fully utilized. The two colour mapping techniques investigated each had advantages and disadvantages. The main disadvantage of the polynomial method was discussed above. On the other hand, the conformal mapping method tended to cover jutting-out parts of the PC-space from fairly small regions in colour space (e.g. the red area of the key in Figure 2.8-15). This would be a significant disadvantage if such an area were densely populated with data.



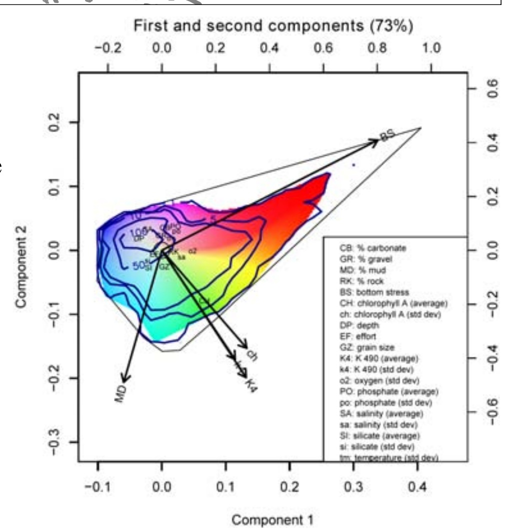


Figure 2.8-15 Map of the stratification of the Torres Strait seabed, with similar colours representing physically similar strata. *Inset right*: colour key from a bi-plot of the first and second principal components of the biologically weighted physical data. Contours (blue lines) of a kernel density estimate of all 41285 scores is also shown at levels 1, 5, 10, 50 and 100. The convex hull of these scores is shown by a thin black line. Loadings of important variables are denoted by black arrows and are labelled. The less important variables (those with smaller loadings) are denoted by small labels. These two components explain 73% of the variance. In the background is the colour key, which is a mathematically distorted colour disk mapped to the interior of the outermost contour using conformal mapping. Red has been chosen to align with high bottom stress and green with high K490 and chlorophyll A. average. Cyan aligns with mud. Common cells close to the mode of the density function have less saturated colours (grey).

3. DISCUSSION

This project has successfully collated a substantial base of knowledge of the physical environment of the seabed and water column, and of the basic seabed habitat types of Torres Strait. Nevertheless, the coverage of each dataset was either incomplete in space and/or time, or in the case of modelled datasets uncertain due to gaps in underlying data, or in the case of remote sensed data ambiguous due to confounding sources of radiance. Thus, there are notable data gaps and key information is needed to provide the knowledge required to deliver the broad objectives of the CRC Torres Strait and to assist Regional Marine Planning and evaluation of management in the region.

3.1. STATE OF KNOWLEDGE

Specific details of available data and maps are provided in section 2; the main facts of knowledge are summarized here.

3.1.1. Oceanographic Data

Tides and currents are known to dominate the physical oceanography of Torres Strait. The friction of the complex bathymetry of the region severely limits the net flow through the region and causes tidal elevations to be out of phase across the main reef tracks and strong tidal currents particularly in channel areas. Knowledge of tides and currents in the Torres Strait comes largely from the output of models, which have progressed from highly simplified channel models through 2- and 3-dimensional models that show good agreement with the (spatially limited) observations of sea levels and currents.

A spatial and seasonal mapping of basic hydrographic conditions (temperature, salinity, oxygen and nutrients — CARS) is available for Torres Strait, though the spatial and temporal coverage of the source data is sparse and seasonally biased. Nevertheless, temperature peaks broadly over the summer monsoon period and during winter, a cold-water anomaly forms over the shallow parts of Torres Strait. Salinities fall under the influence of the monsoon rains, when a freshwater feature also occurs along the PNG coastline, then salinity gradually increases again after the monsoon. The trade-winds can occasionally introduce freshwater from the Fly River plume to the northeast. Strong tidal mixing within Torres Strait generally prevents the development of vertical stratification and dissolved oxygen levels in the well-mixed waters tend to be relatively high, though the data coverage is extremely sparse. Nutrients (nitrate, phosphate, and silicate) appear to increase during the monsoon and decrease during the trade wind season, consistent with riverine inputs, but the existing data coverage is insufficient. A single hydrographic station, monitored from 1977 to 1983 near Booby Island in southwestern Torres Strait, showed temperature and salinity follow a very regular annual cycle and provided localised insight into interannual patterns suggesting variability in annual monsoonal low-salinity anomalies and nitrate peaks, and higher levels of interannual variability in dissolved oxygen due to trade winds.

Suspension of sediments is enhanced by strong spring tide cycles and wind stress, particularly in shallow areas and where riverine input occurs, such as the PNG coastline, leading to localised turbidity maxima. Turbidity estimates have also been derived from SeaWiFS satellite data but can be confounded in shallow areas like Torres Strait. Very few direct measurements have been made.

The phytoplankton (chlorophyll) in Torres Strait is not known from any direct measurements and estimates based on satellite ocean colour are likely to be confounded by turbidity and shallow water. Nevertheless, there are indications that chlorophyll levels in Torres Strait are higher during the monsoon season and decrease towards winter — a pattern opposite to that in the Gulf of Carpentaria and Coral Sea. The factors controlling primary production in Torres Strait are not known adequately, but may be related to seasonal nutrient levels and light availability.

3.1.2. Physical seabed data

Torres Strait is a shallow area of continental shelf of complex topography with numerous reefs and islands. The eastern area includes deeper water but is more complex; the northwestern area is very shallow and limits navigation. The bathymetry of the main shipping channels is well known and perhaps as much as half of the region has been mapped at low resolution, but there is very little high resolution data. The available data has been collated from a variety of sources including the Hydrographic Office, Geoscience Australia and CSIRO Marine Research.

The seabed sediment composition of Torres Strait covers the full spectrum from fine terrestrial muds near riverine inputs to coarse biogenic carbonate sands and gravels among coral reefs further from land. Overlaid on that pattern is the influence of the strong tidal currents that scour the fine sediment from narrow channels, leaving coarse gravels and rocks, and deposit them in calmer areas. The currents also create and move sizeable dunes or bedforms. These patterns in sediments and substratum have been shown to influence patterns in biological assemblages. The sediment data available with broad coverage includes mud, sand, gravel, and rock fractions, grain size and sorting, and carbonate content. The data has been collated principally by OSI, University of Sydney and CSIRO Marine Research.

Acoustic remote sensed data was available from selected CSIRO Marine Research surveys with substantial coverage in central and east/southeast Torres Strait and has provided a useful surrogate for sediment where it had not yet been collected, and continuously along the vessel track. The acoustics data was also an indicative surrogate for basic epibenthic habitat.

Prawn trawl effort data, collated from the QDPI logbooks and summarized at a fairly coarse 0.1 degree resolution, showed that trawling was largely confined to a relatively narrow strip in central eastern Torres Strait and that extremely intense effort was confined to an area of about 200 km^2. From studies elsewhere, it is known that trawling activities are often even more aggregated at finer scales. Knowledge of spatial and temporal patterns of trawl effort is required for evaluations and management of the environmental sustainability of trawling.

3.1.3. Biological seabed data

Compared with most other areas of seabed within Australia, the Torres Strait is relatively well known. However, the biological survey data having good coverage of Torres Strait provide only a relatively coarse level of habitat information. This broad characterisation of the Torres Strait seabed is available for about two-thirds of the region, at a spatial resolution in the vicinity of about 10 km, and includes a 5-point ordinal scale for substratum type, a 4-point rank scale for megabenthos gardens, and presence and absence of algae and seagrass. Higher resolution knowledge is available for some of these seabed biotic components, but with much more limited spatial coverage. For example, seagrass percent cover; seagrass species cover or biomass; percent cover of some algal genera; megabenthos biomass; seabed fish species biomass; and abundance of a few specific species eg. lobster, pearl shell and some holothurians.

Seabed current stress appears to be very important in structuring patterns in the major biological habitats of the seabed, as has been known for some time (eg. Long *et al*. 1997). This study has confirmed its importance, as well as the importance of a range of other physical covariates, including: chlorophyll (from SeaWiFS), turbidity, oxygen, phosphate, salinity, silicate, nitrate, sediment grain size, and depth.

The sediments are known to be dynamic, having been scoured from inter-reef channels they can be deposited in large dunes that move and have been observed smothering sessile megabenthos or exposing bare rocky substratum. Seagrass diebacks are known to occur in north west Torres Strait coincident with changes in sediment dynamics, but the cause and effect is uncertain in relation to sediment.

Seabed fish species assemblages in vicinity of the trawl fishery in north east Torres Strait have been shown herein to differ between areas open and closed to trawling, most likely due to environmental differences rather that effects of trawl effort. From previous work (Harris *et al*. 1994), it is known that the trawl bycatch fish assemblage overlaps little with the fish assemblages on coral reefs in the same vicinity.

3.2. KNOWLEDGE GAPS

Specific details of data gaps are provided and mapped in section 2. The major deficiencies causing critical knowledge gaps are summarized here.

3.2.1. Oceanographic Data

There are few tidal and current monitoring stations in Torres Strait to provide data to validate models. Those available have been positioned in key navigation areas or to examine the Fly river plume results. The vast majority of Torres Strait is an observation gap with respect to these measurement stations, leading to model uncertainty and lack of knowledge of the broader the low frequency circulation, or dispersion and connectivity patterns. Gaps in bathymetry also affect the reliability of hydro-dynamic models.

The CARS mapping of basic hydrographic conditions is limited by the extreme spatial and temporal sparsity of the underlying hydrographic data in the region — ie. these parameters are more unknown than known — and there is a critical need to supplement these data to gain any knowledge of basic productivity processes in Torres Strait. Time series hydrographic data was available from only a single station and ceased ~20 years ago. Other key locations and contemporary data are gaps limiting knowledge of interannual variability and environmental change in the region.

Suspended sediments have been measured in only a very few locations and knowledge of their spring-neap and seasonal cycles is inadequate, as is knowledge of the role of suspended sediments in the biogeochemical cycles of Torres Strait. Knowledge of phyto-plankton biomass is limited to confounded estimates of chlorophyll inferred from satellite ocean colour and there is no knowledge of primary production and plankton community structure in Torres Strait.

3.2.2. Physical seabed data

Apart from the major shipping channels and low resolution mapping from research voyages, the bathymetry of most of Torres Strait has not be surveyed and is poorly known. Thus, for the majority of Torres Strait, bathymetry is potentially unreliable for circulation modelling and bio-physical mapping. Knowledge gap areas of particular concern include: NE Torres Strait, which is very complex with large reef and shoal formations, deep areas and channels; and NW Torres Strait, which is very shallow with large sand ridges and shoals, mostly uncharted and difficult to navigate. High resolution bathymetry is extremely limited in spatial coverage. Thus, there is very limited knowledge of fine scale topographic complexity.

Extensive gaps in surface sediment grain size attributes include most of east/southeastern and northwestern Torres Strait, which are largely unsampled. Given the importance of sediment for biota, the use of physical surrogates for assemblage prediction in these areas is questionable. Nevertheless, the remainder of Torres Strait has been only patchily and inadequately sampled, except for the main shipping channels and an area west of the southern warrior reefs. The gaps for surface sediment carbonate data are similarly distributed but even more extensive. There is very little knowledge of the spatial distribution of surface sediment organic content, and very restricted sampling of sediment depth profiles or sediment cores.

The trawl logbook effort data coverage is relatively complete, however the resolution is coarse compared with actual trawling activities. Thus, there is a gap in knowledge of fine scale trawl effort patterns needed to assess and manage the environmental sustainability of trawling.

The overall gaps in the physical seabed data, weighted by relative importance in the stratification are shown in Figure 3.2-1. Across all the physical seabed covariates, the sampling has been inadequate (significantly greater than the spatial auto-correlation distance for these attributes) for bio-physical mapping over the majority of Torres Strait. Since the current stress model and the SeaWiFS data have relatively high resolution, these gaps are largely due to the sediment, bathymetry, and CARS data respectively, which lead to uncertainty in the existing bio-physical maps. For reliable mapping of biological assemblages from bio-physical modelling, the required spatial resolution of the physical data is finer than that needed for the assemblages themselves (at ~0.1°) — that is, the physical data need to be available at a resolution finer than ~0.05° (see section 2.6.1) to reduce uncertainty in maps that attempt to predict biological assemblages where they have not been sampled.

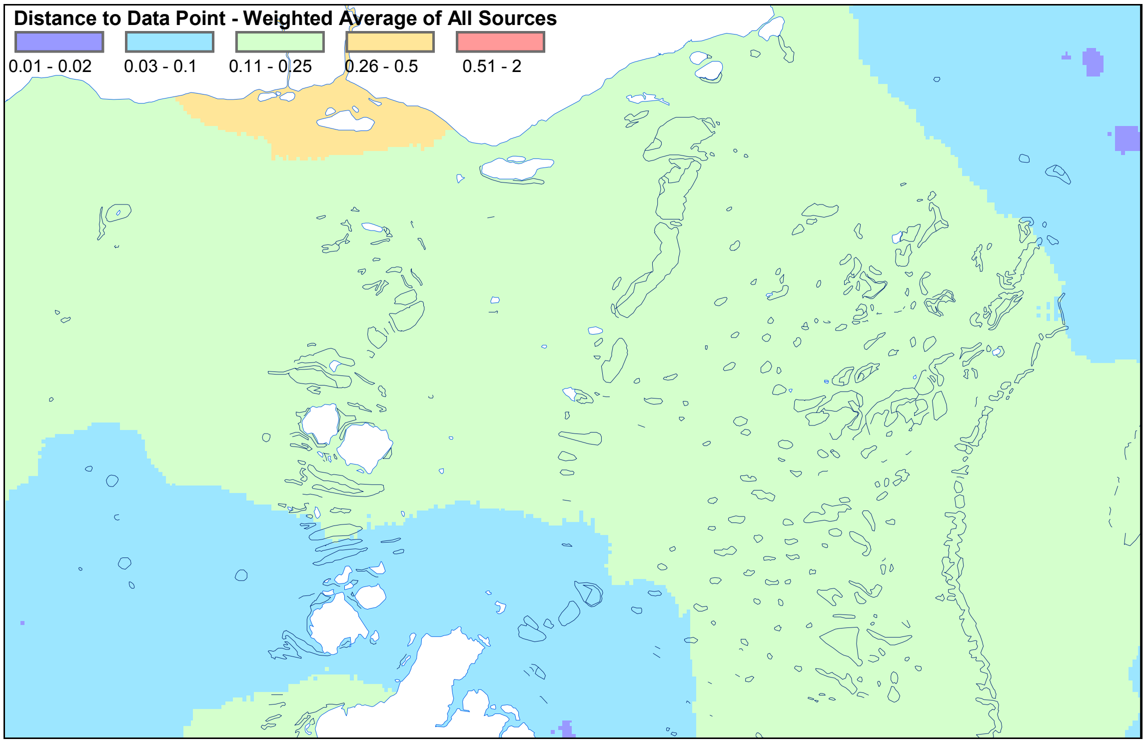


Figure 3.2-1 Map of overall gaps averaged across all seabed physical covariates, weighted by co-variate importance (section 2.8.1.2). That is, the distance-to-data of more the important co-variates has greater influence in the calculation of the average for each point.

3.2.3. Biological seabed data

The most basic of seabed habitat type information is unknown for north eastern Torres Strait and for extensive areas of western Torres Strait. While qualitatively helpful, this level of information has been shown to be inadequate for quantitative management applications where species biomass data are required. However, seabed species assemblages have not been sampled in Torres Strait, except for the trawl bycatch fish component in part of the prawn trawl grounds. Thus, other components of seabed assemblages are unknown, even in the sampled trawl grounds, and there are no species assemblage data available for the remainder of Torres Strait.

In addition to the lack of a static spatial description for seabed assemblages, there is very little knowledge of population or assemblage dynamics or of ecosystem processes. The dynamics of the lobster and prawn populations have been sampled for fisheries assessment purposes, and the annual lobster surveys over 15 years have recorded some changes in basic habitat type in a limited area. A few surveys of seagrass, mostly in central north-west Torres Strait, at irregular time intervals, have documented seagrass–dieback. Primary productivity, whether benthic or planktonic, has not been studied in Torres Strait nor have other ecosystem processes such as secondary productivity and higher trophic relationships, other inter-species inter-actions, or coupling between the benthic and pelagic ecosystems.

3.3. KEY INFORMATION NEEDS

The key information needed to advance critical knowledge of the Torres Strait marine environment, for sustainable management of the region, largely corresponds to the critical data gaps identified.

3.3.1. Oceanographic Data

Additional tidal and current monitoring stations need to be established at key locations distributed broadly across Torres Strait to provide data to validate circulation models and provide knowledge of the broader low frequency circulation, dispersion and connectivity patterns. Bathymetric data is also needed from gap areas because of its important to circulation. Hydrographic moorings are also needed at similar key locations to collect temperature, salinity, oxygen, nutrient and chlorophyll data to provide knowledge of interannual variability and environmental change in the region and to develop an understanding of productivity processes in Torres Strait. Measurements of suspended sediments are also required at critical locations across spring-neap tidal and seasonal cycles.

Complete studies are required of biogeochemical cycles, phyto-plankton and zoo-plankton assemblages and their primary and secondary productivity, again in a range of key sub-environments distributed broadly across Torres Strait.

The work program for the Torres Strait CRC will address a limited number of these data issues, and at varying levels of detail. Field measurements to be made by the “Bio-Physical” and “Seabed Mapping” Tasks and further hydrodynamic modelling will provide more detailed understanding of circulation and dispersion patterns, as well as some additional hydrographic data to improve CARS within the region. There will also be a substantial increase in the suspended sediment dataset, which will be complimented by sediment transport modelling. There may be opportunities to collect in situ chlorophyll data to assist with interpretation of ocean colour. However, the existing CRC-TS program does not include plankton sampling or studies of the biogeochemical cycles needed to understand the processes controlling primary and secondary production in Torres Strait.

3.3.2. Physical seabed data

Consistent coverage of bathymetric data over Torres Strait, at resolution less than the characteristic spatial auto-correlation distance (<<10 km), is required for circulation modelling and bio-physical mapping. Bathymetric data of sufficient resolution for navigation purposes is highly desirable, preferably acquired by airborne laser (LADS) and/or swathe mapping systems. High resolution bathymetric data is particularly important for areas like the Torres Strait where the topography is highly complex with many areas of reef intersected by narrow channels, and shallow shoaling areas.

Adequate sampling of surficial sediments, for grain size and composition attributes, is required in the extensive unsampled areas of Torres Strait, to provide a basis for understanding the important sediment processes, biogeochemical cycles and for use as surrogates for biotic assemblage prediction. Again the required scale is less than the characteristic spatial auto-correlation distance (<10 km). Measurement of sediment organic content, depth profiles and sediment cores would also contribute greatly to this knowledge.

Fine scale trawl effort data needs to be acquired from Vessel Monitoring Systems installed on Torres Strait trawlers and would provide the effort data at a resolution compatible with that needed to assess and manage the environmental sustainability of trawling.

Again, the work program for the Torres Strait CRC will address several of these data issues at varying resolution. Field sampling and measurements of physical attributes to be made by the "Seabed Mapping" Task include soundings of bathymetry along the vessel track, site samples of surface sediment for grain size and composition, and site measurements of hydrographic data (profiles of temperature, salinity, oxygen, turbidity and fluorescence) to improve CARS within the region. The resolution of the site-based sampling will be about 10-12 km — close to the upper limit of the spatial auto-correlation distance (<10 km). However, due to resource constraints, the coverage of Torres Strait at this scale will be incomplete, and gaps totalling about one third of the region will remain. In addition, the "Bio-Physical" Task will collect high-resolution swathe bathymetry and sediment data at number of small scale study sites.

3.3.3. Biological seabed data

Fundamental biological information is needed as a basis for regional marine planning in Torres Strait to contribute to the development of conservation, management and monitoring requirements. Information on the seabed habitat type is a basic need for unknown areas in north-eastern and western Torres Strait. However, it has been shown that knowledge of a broad spectrum of biota, at the species level, within seabed assemblages is required to properly characterize assemblage patterns, to develop bio-physical models and for quantitative management applications (Pitcher *et al*. 2002). For these purposes, seabed species assemblages need to be sampled in Torres Strait with multiple sampling devices, accurately identified with detailed taxonomic skill, and quantified and mapped. Careful identification of species is required because Torres Strait is significant biogeographically due to the periodic separation of east & west faunas due to changes in sea level. This biogeographic information is also essential for regional marine planning.

Successful management of multiple uses of the Torres Strait marine ecosystems also needs to take into account possible interactions between sectors and complex effects on the ecosystem that cannot be accommodated by single-sector management approaches. This requires dynamic modeling of the important components of the ecosystem, the human uses and the management (Management Strategy Evaluation — MSE), which has data-needs beyond static maps of seabed biological assemblages. The level and detail of data input required for such approaches varies substantially, but in general includes knowledge of population and assemblage dynamics and of ecosystem processes such as primary productivity (both benthic: seagrass, macro/micro-algae, and planktonic: phyto-plankton), secondary productivity (herbivory: includes dugong, turtle, fishes) and higher level trophic relationships (predation), other inter-species inter-actions (eg. competition), as well as coupling between the benthic and pelagic ecosystems. To be successful, ecosystem-based management requires this kind of ecosystem-level knowledge.

As with oceanographic and physical seabed information needs, the work program for the Torres Strait CRC will address several of these data issues to varying extents. The Torres Strait Seabed Mapping Project will provide data on seabed habitat and megabenthos from video, specimens of fauna and flora from epibenthic sled sampling and specimens of fishes, crustaceans and other bycatch species from trawl sampling. These specimens will be identified and quantified and their bio-physical relationships analysed to produce maps of assemblage composition and abundance, and develop attributes to assist planning (eg. biomass, species richness, rarity, uniqueness, condition, vulnerability). This information will be provided to another CRC TS project that will develop an MSE approach to evaluate management strategies for trawling that reduce environmental impacts. However, as noted above, the resolution of the site-based sampling will be close to the upper limit of the spatial auto-correlation distance and, due to resource constraints, the coverage of Torres Strait at this scale will not be complete. Further, the existing CRC TS program does not include studies of the biological assemblage dynamics, trophic relationships or other ecosystem processes needed to develop ecosystem-based management. The CRC TS program will, nevertheless, make considerable progress toward ecosystem-based management in the Torres Strait.

4. REFERENCES

Bode, L. and L. Mason, 1995. Tidal modelling in Torres Strait and the Gulf of Papua. In *Recent Advances in Marine Science and Technology '94* O. Bellwood, H. Choat, and N. Saxena, ed. Townsville, Queensland, pp. 55-65.

Breiman, L. (2001), *Random Forests*, Machine Learning 45(1), 5–32.

Burford, M. A. and P. C. Rothlisberg, 1999. Factors limiting phytoplankton production in a tropical continental shelf ecosystem. *Estuarine, Coastal and Shelf Science*, 48, 541-549.

Clark, A. J., 1990. Application of a frictional channel flow theory to flow in the Prince of Wales Channel, Torres Strait. *Journal of Physical Oceanography,* 20**,** 890-899.

Dunn, J. R. and K. R. Ridgway, 2002. Mapping ocean properties in regions of complex topography. *Deep-Sea Research I,* 49**,** 591-604.

Harris, P. T. and E. Baker, 1991. The nature of sediments forming the Torres Strait turbidity maximum. *Australian Journal of Earth Science,* 38**,** 65-78.

Harris, A.N.M., Dews, G.J., Poiner, I.R., Kerr, J.D. 1994. The traditional and island based catch of the Torres Strait Protected Zone: Final report on CSIRO research, 1990-1993. CSIRO Division of Fisheries Report, Australia. pp. 47.

Harris, P. T., 1991. Reversal of subtidal dune asymmetries caused by seasonally reversing wind-driven currents in Torres Strait, northeastern Australia. *Continental Shelf Research,* 11**,** 655-662.

Harris, P. T., 1999. Environmental management of Torres Strait: A marine geologist's perspective. In *Gondwana to Greenhouse: Environmental Geoscience - an Australian Perspective*, V. A. Gostin, ed. pp. 149-160.

Harris, P. T., E. Baker, A. Cole and S. Short, 1993. A preliminary study of sedimentation in the tidally dominated Fly River Delta, Gulf of Papua. *Continental Shelf Research,* 13**,** 441-472.

Harris, P. T.. Environmental management of Torres Strait: A marine geologist's perspective. In *Gondwana to Greenhouse: Environmental Geoscience - an Australian Perspective*, V. A. Gostin, ed. pp. 149-160.

Hemer, M. A., P. T. Harris, R. Coleman and J. Hunter, 2003. Sediment mobility due to currents and waves in the Torres Strait - Gulf of Papua region. *Continental Shelf Research,*(in press).

Kaufman, L. and Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.

Kloser, R. J., Bax, N. J., Ryan, T., Williams, A., and Barker, B. A. (2001) Remote sensing of seabed types in the Australian South East Fishery; development and application of normal incident acoustic techniques and associated 'ground truthing'. Marine and Freshwater Research 52(4):475 – 489.

Long, B.G. Pitcher, C.R., Bode, L., Mason, I. (1997) Seabed Current Stress as a Predictor of the Distribution and Abundance of Epibenthos in Torres Strait. *CSIRO Division Of Marine Research Report* MR-GIS 97/6

Pitcher C. R., Gordon, S. R., Kloser, R. J., and Jones, P. N., 1999. Development of an acoustic system for remote sensing of benthic fisheries habitat for mapping, monitoring and impact assessment. *CSIRO Report to Fisheries Research and Development Corporation* T93/237: 74 pp. (CSIRO Marine Research: Cleveland, Australia.) ISBN 0 643 06196 7.

Pitcher, C.R., Venables, W., Ellis, N., McLeod, I., Cappo, M., Pantus, F., Austin, M., Doherty, P., Gribble, N., 2002. GBR Seabed Biodiversity Mapping Project: Phase 1 Report to CRC-Reef. *CSIRO/AIMS/QDPI Report*, pp. 192.

Pitcher, C.R., Skewes, T.D., Dennis, D.M., Prescott, J.H. (1992) Distribution of seagrasses, substratum types and epibenthic macrobiota in Torres Strait, with notes on pearl oyster abundance. *Aust. J. Mar. Freshw. Res.* 43: 409–419

Ridgway, K. R., J. R. Dunn and J. L. Wilkin, 2002. Ocean interpolation by four-dimensional least squares - Application to the waters around Australia. *Journal of Atmospheric and Oceanic Technology,* 19**,** 1357-1375.

Rothlisberg, P. C., P. C. Pollard, P. D. Nichols, D. J. W. Moriarty, A. M. G. Forbes, C. J. Jackson and D. Vaudrey, 1994. Phytoplankton community structure and productivity in relation to the hydrological regime of the Gulf of Carpentaria, Australia, in summer. *Australian Journal of Marine and Freshwater Research,* 45**,** 265-282.

Scott, D. W. (1992) *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley.

Trefethen, L. N. 1980. Numerical Computation of the Schwarz-Christoffel Transformation, *Siam J. Sci. Stat. Comp.* 1, 82–102

Wolanski, E., A. Norro and B. King, 1995. Water circulation in the Gulf of Papua. *Continental Shelf Research,* 15**,** 185-212.

Wolanski, E., P. Ridd and M. Inoue, 1988. Currents through Torres Strait. *Journal of Physical Oceanography,* 18**,** 1535-1545.

Wolanski, E., S. Spagnol, B. King and T. Ayukai, 1999. Patchiness in the Fly River plume in Torres Strait. *Journal of Marine Systems,* 18**,** 369-381.

5. PROJECT STAFF

Name	**% time**
Roland Pitcher, Project leader: co-ordination, analyses, design, reporting	25%
Scott Condie, Biological Oceanographer: water column characterisation, analyses, reporting	10%
Jeff Dunn, Oceanographic data analyst: water column database & mapping	15%
Ian McLeod, Spatial modeller: physical database, modelling & gridding physical data	35%
Mick Haywood, Seabed biologist, characterisation analyses, reporting	10%
Scott Gordon, Acoustics engineer: analysis & mapping of existing acoustics data	20%
Tim Skewes, Seabed biologist: integration of disparate seabed habitat & biological data	10%
Darren Dennis, Seabed biologist: integration of disparate seabed habitat & biological data	5%
Malcolm Austin, Database programmer: management & integration of habitat/biological data	20%
Bill Venables, Spatial statisitician: biophysical characterisation & modelling	5%
Nick Ellis, Mathematical modeller: biophysical modelling	20%
Tom Taranto, GIS modeller: mapping habitat & biological data	5%
Liz Cotterell, data entry, assist in the integration of disparate data	10%