A new project to foster Japanese-Australian discovery through Data Science for environmental and ecological management

Charis Burridge

Cherry Bud Workshop, Mar ’08, Keio University, Japan

www.csiro.au

CSIRO

Outline of talk

Two parts:

- **Australian-Japan Foundation project (AJF): Keio-CSIRO collaboration including intros to Keio and CSIRO**

- **Recent reef fish monitoring project: involving CMIS statisticians and CMAR marine ecologists**

Keio-CSIRO partnership: to now

- **CSIRO staff at previous Cherry Bud workshops: Richard Jarrett, Harri Kiiveri, John Donnelly**
- **2 PhD students (Hideyasu Shimadzu & Yuki Sugaya) recently undertook 6-week internships with CMIS: Hideyasu worked with Matt Browne, Yuki with Ian Saunders**
- **Hideyasu & Matt developed an AJF project proposal**
- **Hideyasu recruited by GeoScience Australia for collaborative research with CSIRO on national marine biodiversity prediction**
- **CSIRO staff at current Cherry Bud workshop: Ian Saunders, Richard Jarrett and myself**

Keio-CSIRO partnership: from here

- **Apr 2008-Mar 2009: AJF funds 4-week visits by 3 people to/from Australia to promote joint research in marine biodiversity and other applications of joint interest to CSIRO and Keio**
- **One-year program will include mini-workshops in Australia and final workshop in Keio next March**
- **With a combination of CSIRO internships and (hopefully) further AJF travel grants, we plan to develop a longer-term program to engage both current and future Keio Uni postgraduate students in applying statistical methods to important, real-world problems**

Keio University: a snapshot

- **Oldest university in Japan (150 years)**
- **Motto: ‘the pen is mightier than the sword’**
- **Human resource data**
  ~2,600 staff, ~32,000 students (~3,700 doctoral)
- **Graduate School of Science & Technology is home to mathematicians and statisticians**
- **In 2003 Prof Ritei Shibata and others were awarded a 5-year grant as a designated 21st Century Centre of Excellence**
- **The Data Science component has been led by Ritei, part of this being the Data and Description environment he will describe later**

Introducing CSIRO

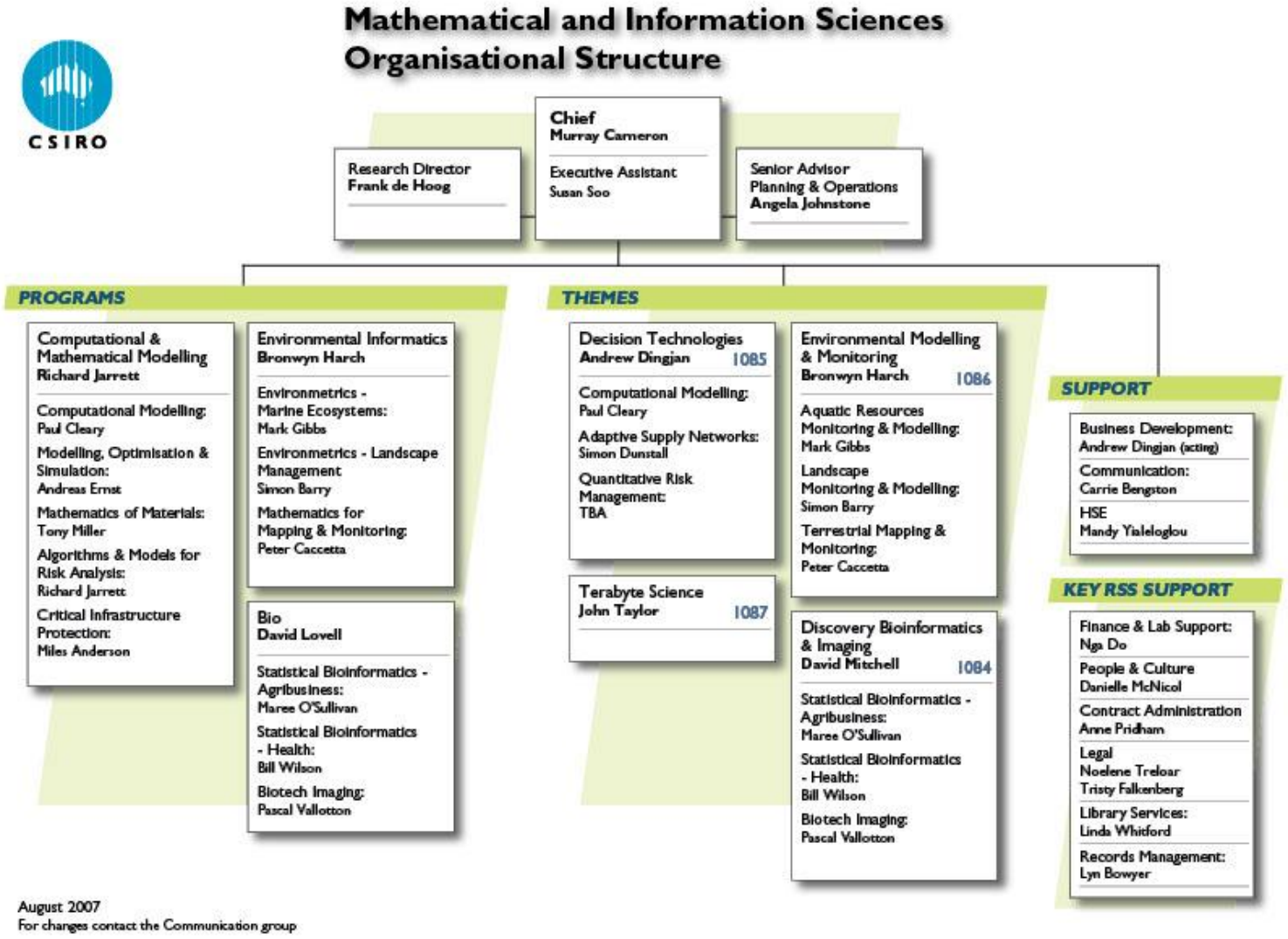
- **Commonwealth Scientific and Industrial Research Organisation**
- **Formed in 1926 (some longevity, but younger than Keio!)**
- **Australia's largest single employer of scientists (>6500; diverse)**
- **Operating from 57 sites in Australia and overseas**
- **A co-founding member of the Global Research Alliance**

Research facilities and locations

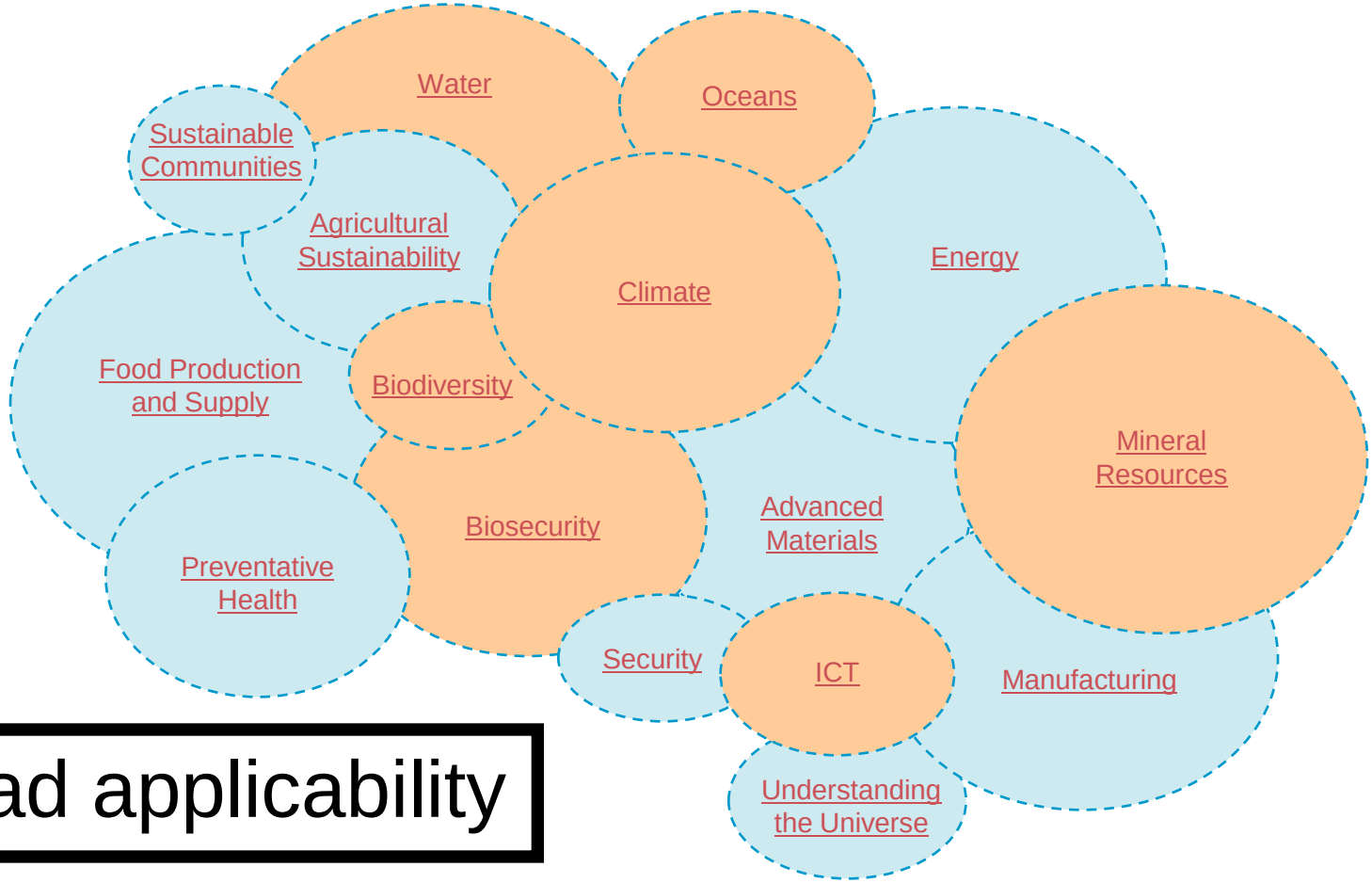
- **CSIRO manages three National Research Facilities:**
  - Australian Animal Health Laboratory at Geelong
  - Australia Telescope at Parkes, Coonabarabran and Narrabri
  - the oceanographic research vessel *Southern Surveyor*.
- **Also have >30 other research facilities such as:**
  - Riverside Life Sciences Centre at Brisbane
  - CSIRO Discovery Centre at Canberra
  - Australian Resources Research Centre at Perth
  - Corporate headquarters in the Australian Capital Territory (Canberra).
- **CSIRO has a laboratory in France and staff located in:**
  - Ireland
  - The Netherlands
  - United States of America.
- **.**

Structure of CSIRO: Divisions & Flagships

CMIS Division: ~120 staff (many statisticians) [Input Programs - R; Output Themes - L]



CMIS Environmental Informatics Program delivers into multiple Outcome Domains



Broad applicability

CMIS Environmental Informatics Program: Core Capabilities

- ***Scaleable technologies with development through multiple investments***

  i. *mathematical frameworks for highly structured stochastic systems*

  ii. *space-time models for environmental monitoring*

  iii. *sample and survey design*

  iv. *eco-scape risk assessment frameworks*

  v. *mathematics for multi-scale space-time data integration*

  vi. *high dimensional data and statistical pattern recognition (discrimination, classification, un-mixing)*

What are we aiming to achieve?

Through harnessing appropriate capabilities we deploy mathematical and statistical technologies that:

1. increase value of remote sensing information via quantified parameter accuracy & variability estimates
2. incorporate the data streams provided by emerging monitoring and assessment technologies to improve the prediction and quantification of uncertainty at different spatial and temporal scales
3. integrate different models & data sources and reliably represent uncertainty in system wide predictions
4. enable the reliable estimation of condition, trends and extremes in both space and time
5. optimally allocate scarce monitoring resources across the environment to satisfy multiple objectives
6. integrate empirical, process based and expert based information into coherent eco-scape risk assessment frameworks

Science Plan for Environmental Informatics in CMIS

- space-time statistical modelling
- spatio-temporal sample and survey design
- Hierarchical Structured Stochastic Systems
- mathematics for multi-scale space-time data integration
- statistical pattern recognition
- eco-scape risk assessment

+

- Domain knowledge
- Relationships

People

& Processes

Capabilities

Eco-scapes
- Continental Scale Vegetation
- Marine Systems
- Catchment to Coastal Zone
- Urban

Using the mathematical sciences to understand and predict processes dominated by **scale** and complexity

Application areas

Research drivers

- **National Research priorities**
- **International Research priorities**
- **Identified needs of collaborators/clients**
- **CMIS Purpose**

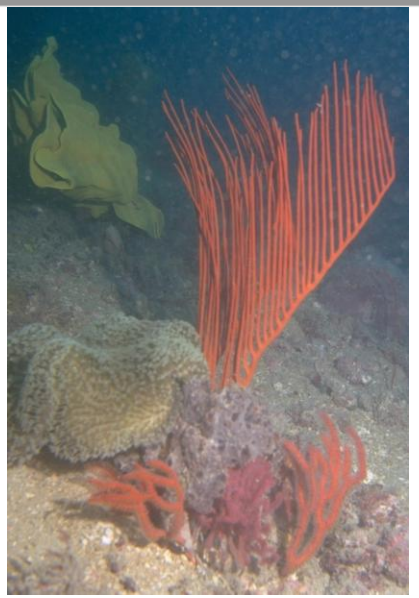
Recent successes of Aquatics Stream

1. Estimating abundance and distribution of minke whales for the International Whaling Commission and Australian Antarctic Division

2. Research to support science-based management of the Northern Prawn Fishery

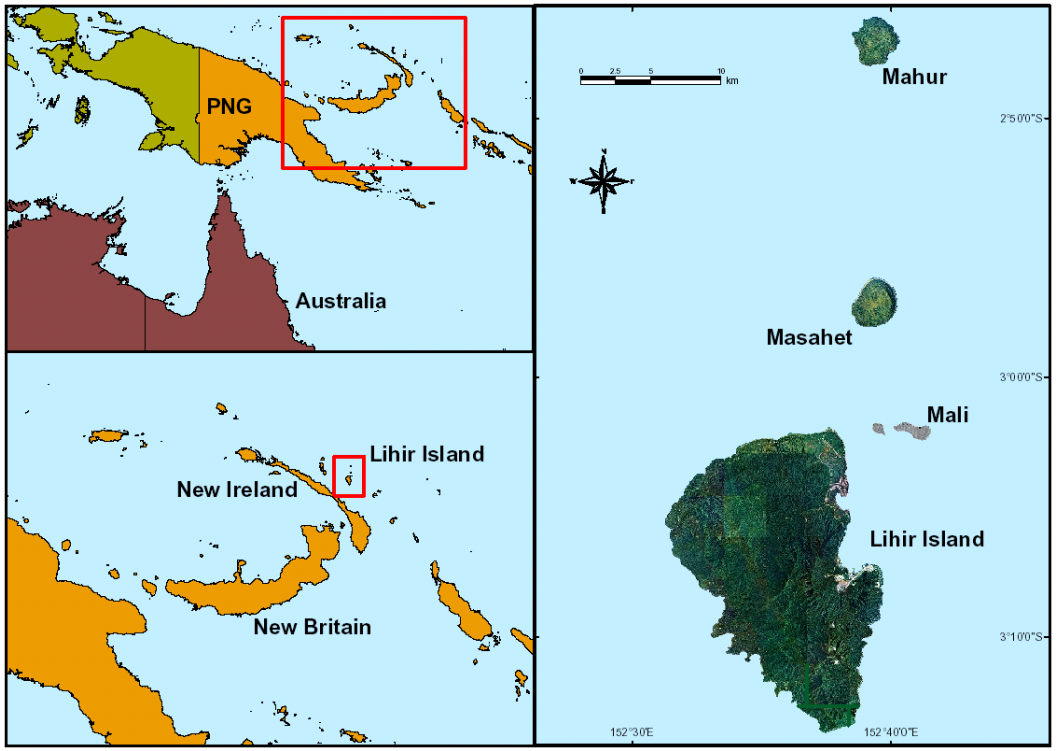
3. Novel methods for mapping and predicting seabed biodiversity in GBR and SE Australia

4. Effective reporting on environmental impact study for Lihir Gold Mine ($$$$$ to PNG but extensive potential ecological footprint)



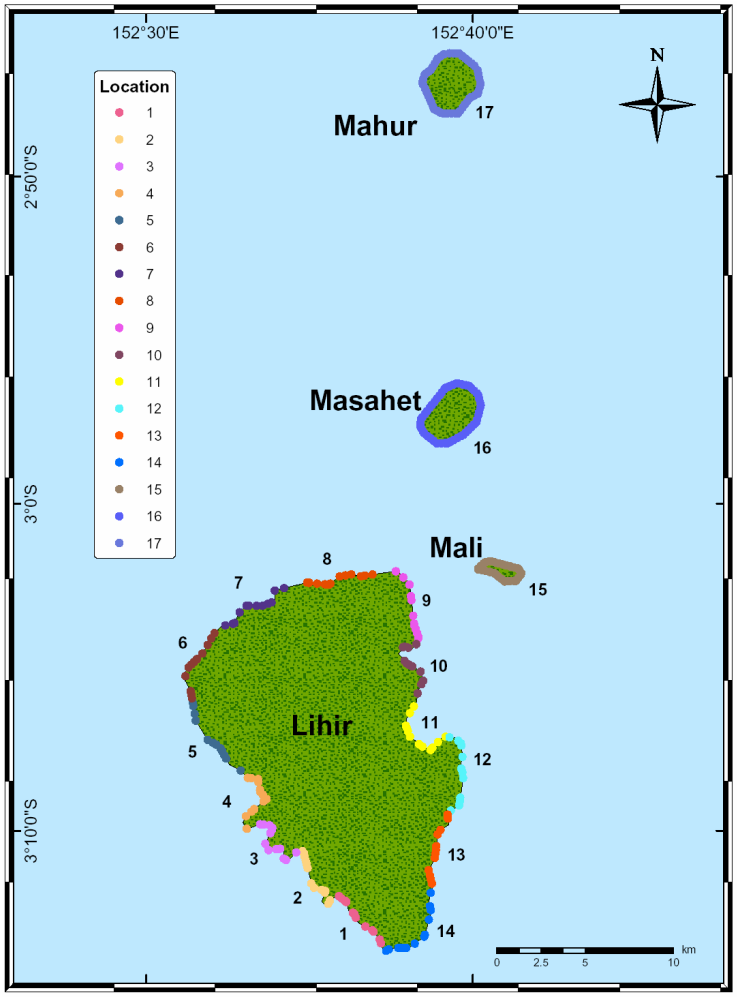
Background to Lihir environmental study

- **Study area**
  - Lihir Island group, about 900 km NE of Port Moresby, capital of Papua New Guinea – now mainly monitoring Niolam (main island) and Mali.



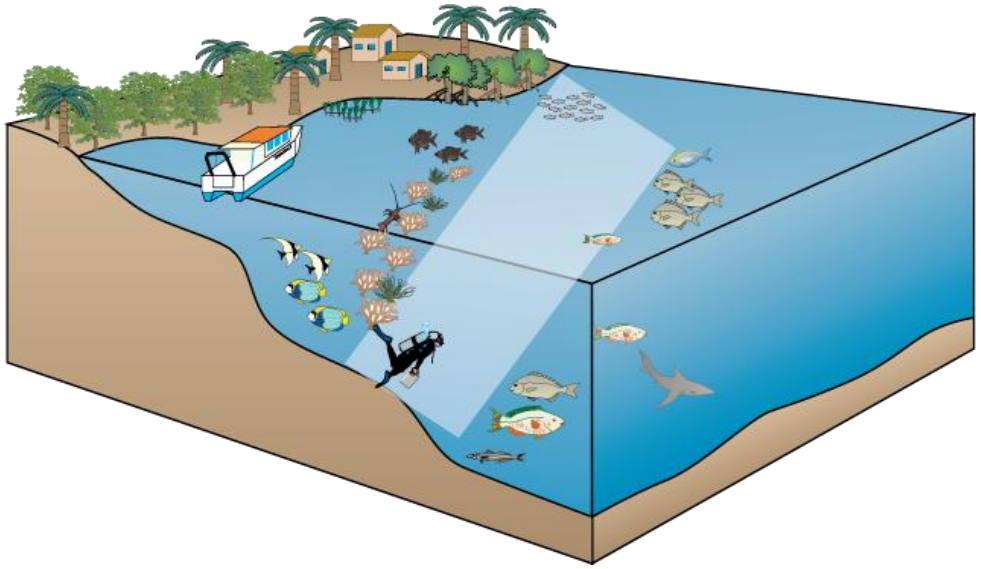
Survey sites for inshore fish monitoring

- **Phase I (1999-2002)**
  - Three surveys at >100 sites on 4 islands, grouped into 17 ‘Locations’
- **Phase II (2005-2007)**
  - **Three surveys at subset of Phase I sites, visited on each survey – longitudinal design to optimise ability to detect change**
  - **Locations 2-5, 16 and 17 no longer surveyed**
  - **Approx 75 sites per visit, and 52 sites have been sampled in all six surveys**

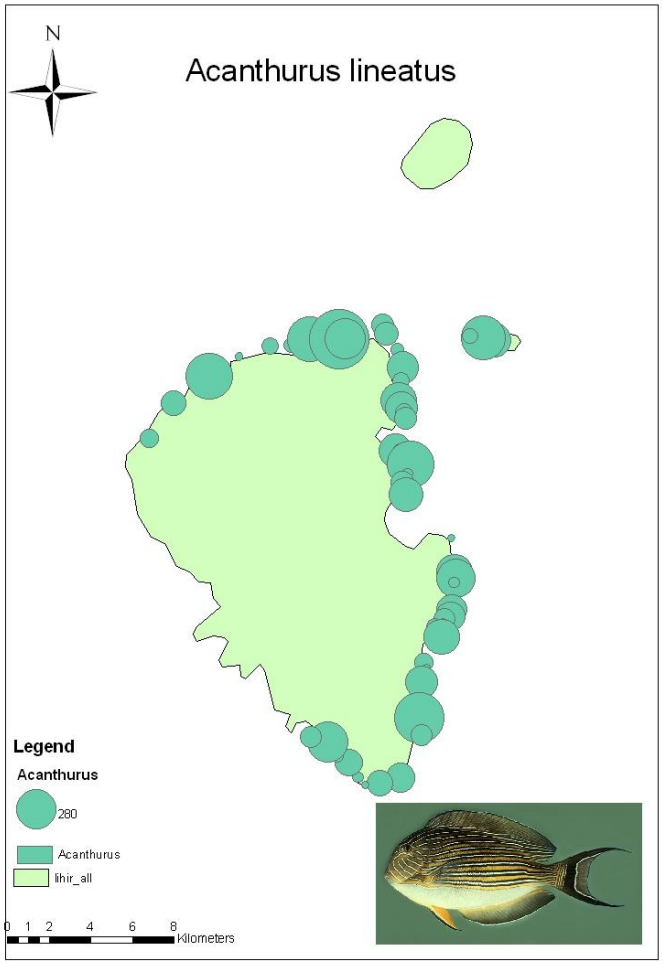


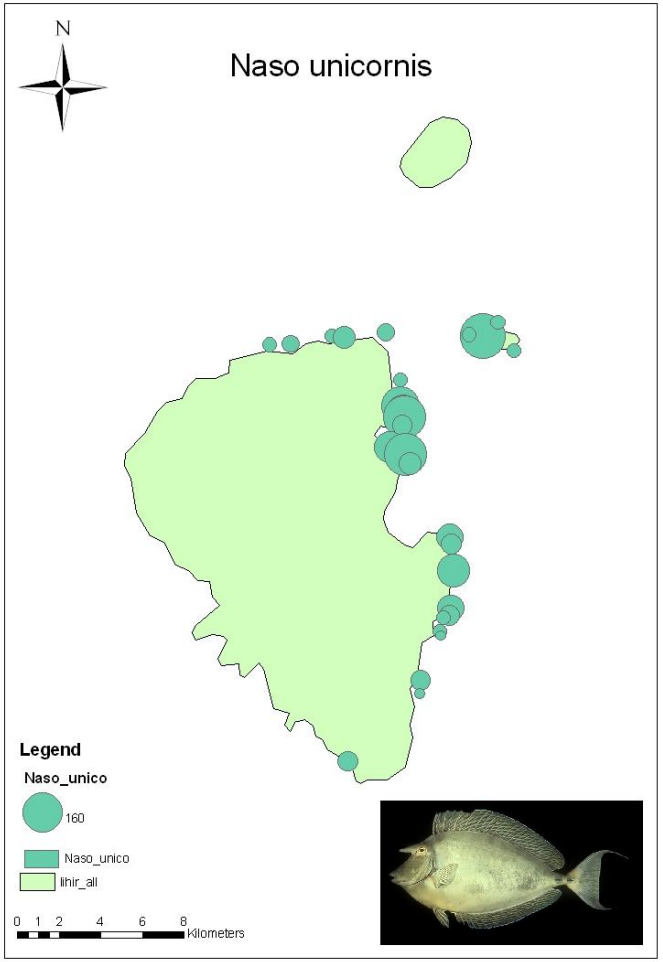
Underwater Visual Census (UVC) sampling

- **Method**
  - Diver swims 100m transect, identifying, assessing size and counting conspicuous fish within a 10m belt.
  - Starts at 15m depth and swims obliquely until reaching reef crest.
  - Also records substrate & habitat information, eg state of coral.

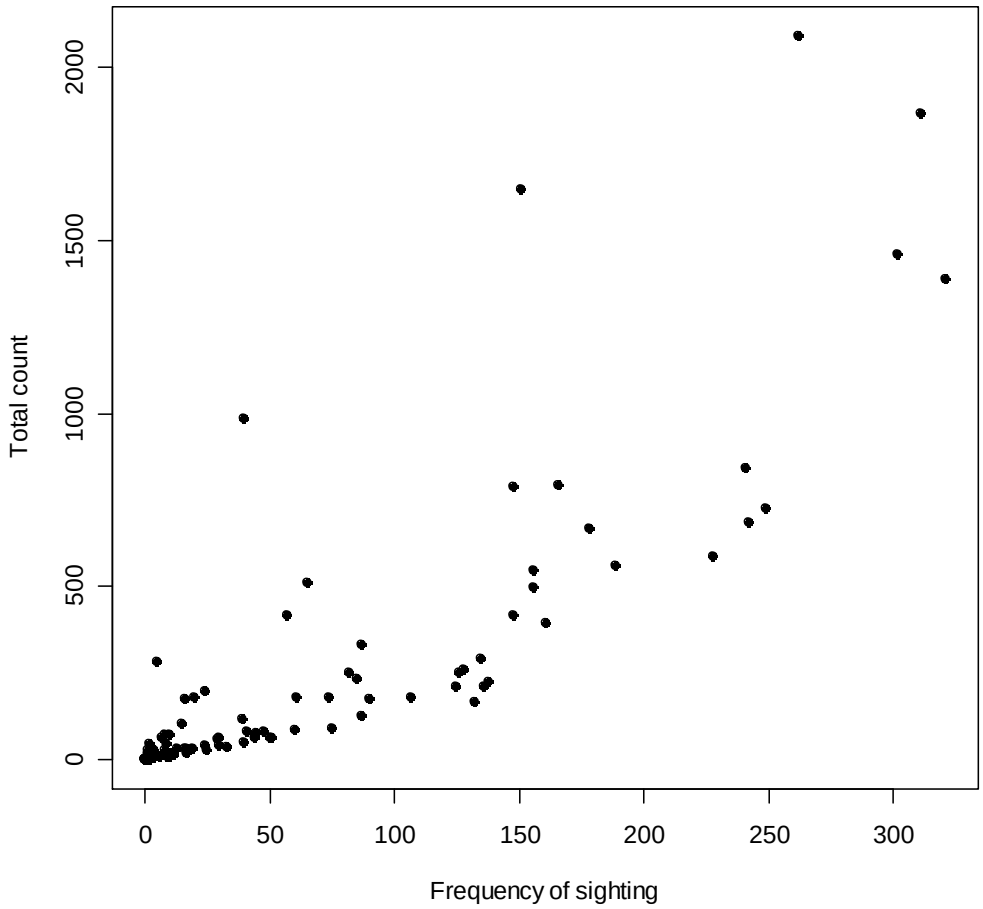


Spatial distribution of two species (raw counts)





Species vary greatly in presence & abundance



Data selection before model-fitting

- **Selected transects visited at least 4 times**
  - 71, 56, 55, 72, 71 and 71 sites per survey
  - Trading off ability to model longitudinal trends at transect level with enough sites to model spatial pattern in density
- **Selected species with at least 220 specimens seen, in at least 55 transects and in every survey**
  - 24 species for individual modelling (out of 117)
  - Future monitoring likely to focus on small number of commonly-seen species
- **Definition of ‘mine’ location (Luise harbour) expanded from 3 sites to 6**
  - to improve confidence interval for mean density
  - without introducing too much bias in the (mostly) very low mean density in Luise Harbour

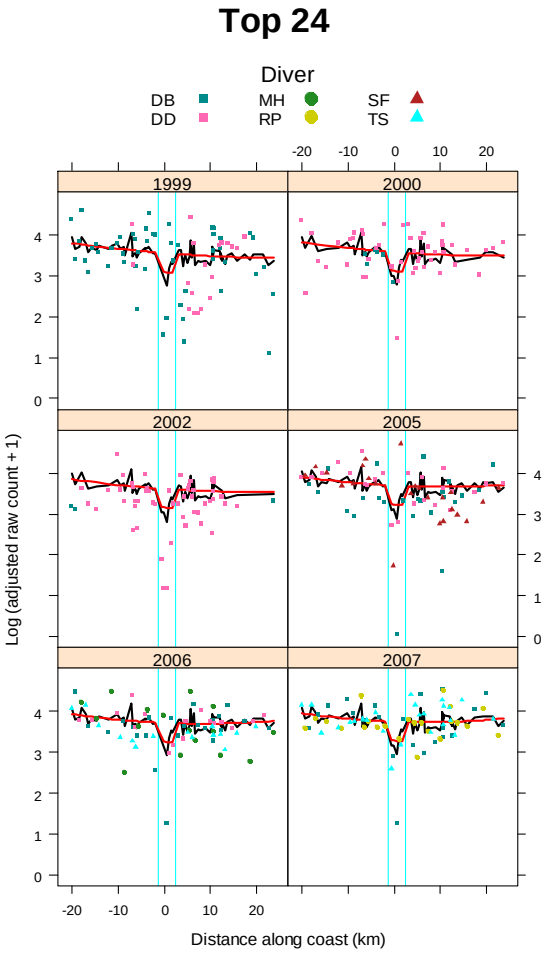
Characteristics modelled

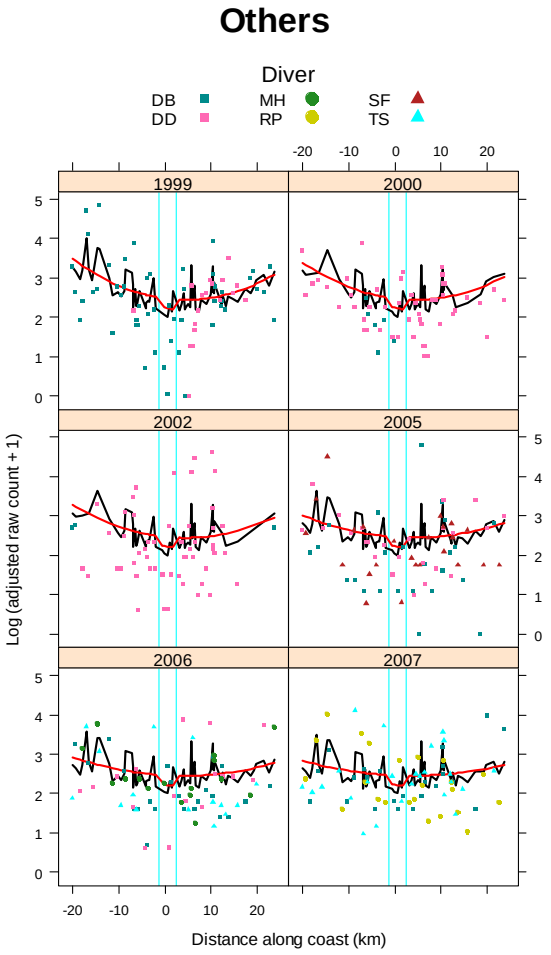
- **Abundance of individual fish species**
- **Total abundance of larger groups:**
  all species; top 24 species; remaining species;
  6 feeding/trophic guilds:
  - benthic invertebrate feeders (8 spp.); herbivorous croppers (5 spp.); herbivorous scrapers (7 spp.); omnivores (2 spp.); Planktivorous feeders (*Naso brevirostris*); Pelagic piscivores (*Aphareus furca*)
- **Also, percent live coral cover**

Basic components of model

- **Statistical blocking factors**
  - 6 different divers, 2-3 per survey
  - Visibility (m) categorised into low (4-8 m) & high (> 8m)
- **Spatial terms**
  - linear/quadratic coastline covariate (‘Dist’ & ‘Dist_sq’)
  - Different intercept in Luise Harbour (‘Mine’)
- **Temporal term**
  - linear time covariate (‘Time’)
- **Spatio-temporal terms**
  - ‘Time x Dist’ and ‘Time x Dist_sq’
- **Quasi-Poisson for count (per transect), with log-link**
- **Random effect(s) for transect (up to 3 parameters: intercept variance, time-slope variance and correlation)**

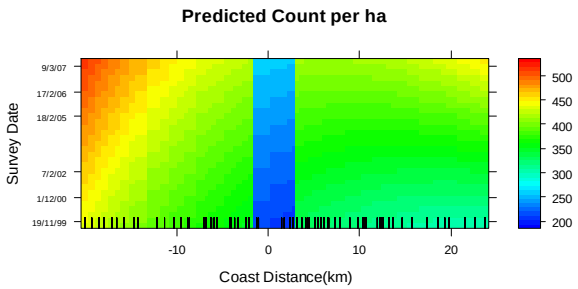
Count data (adjusted for diver & visibility) compared with transect-level/smoothed distance/time model

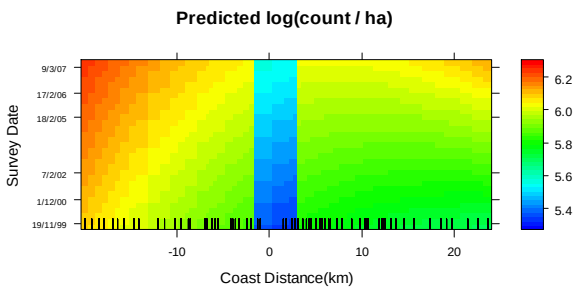


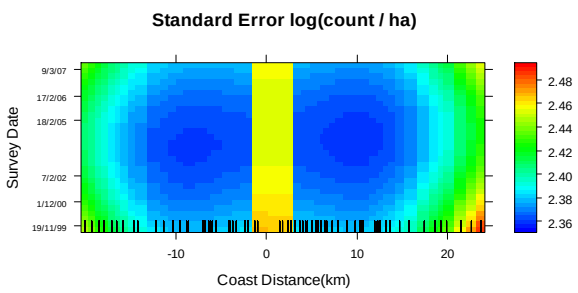


Smoothed space-time model for total of top 24 species and total of other 93 species

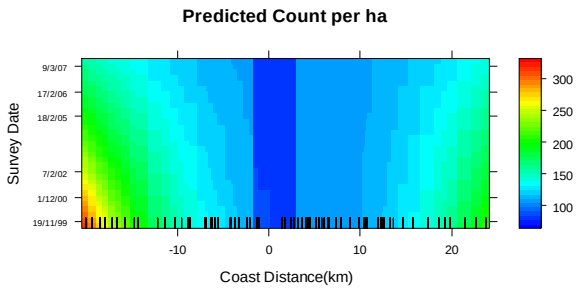
Top 24

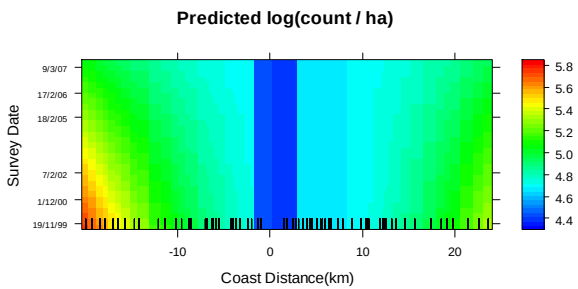


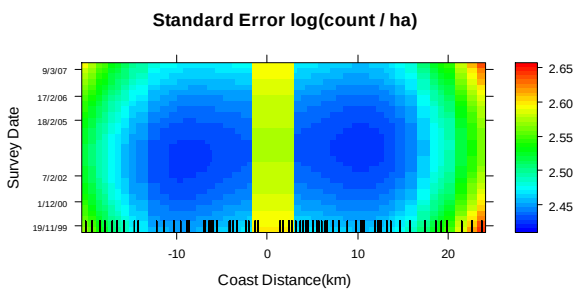




Others

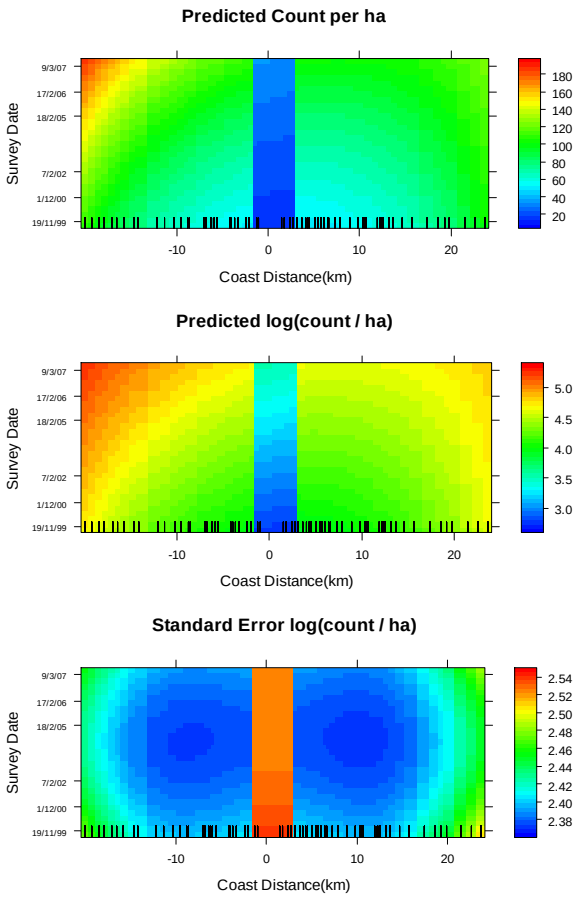




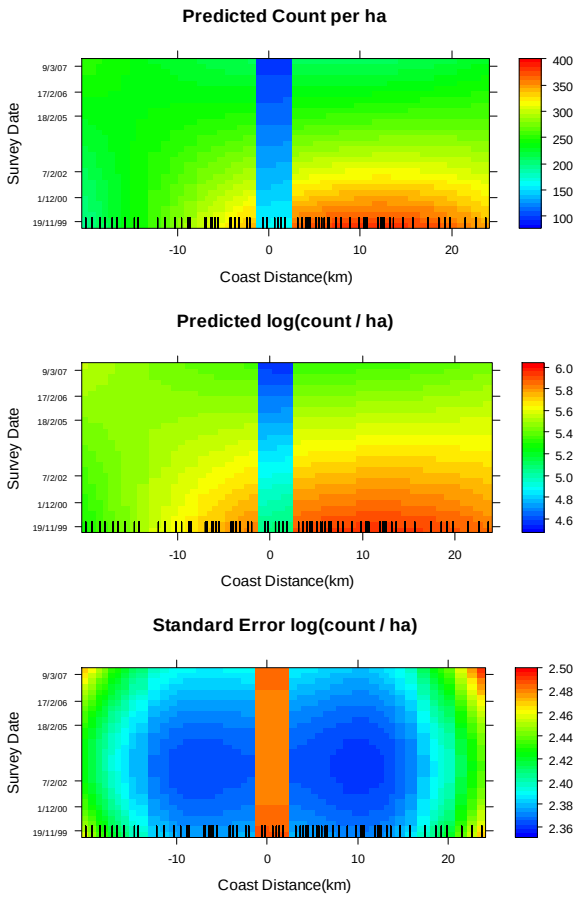


Smoothed trends for herbivorous scraper abundance & percent live coral cover

HS



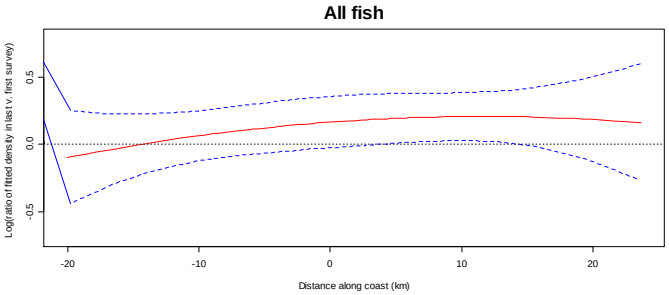
Live coral cover

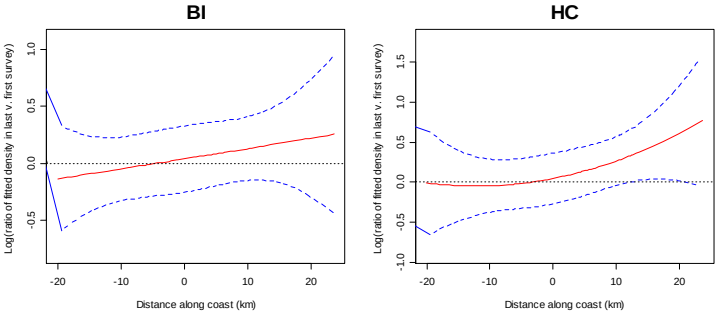


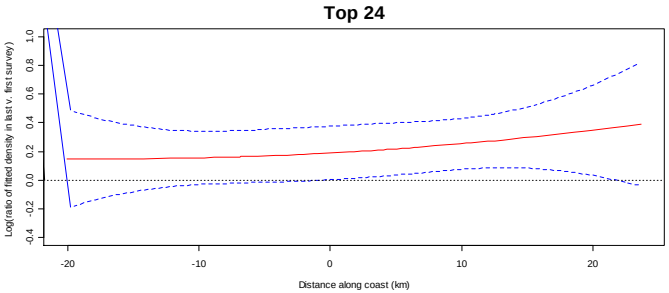
Spatial assessment of mine impact

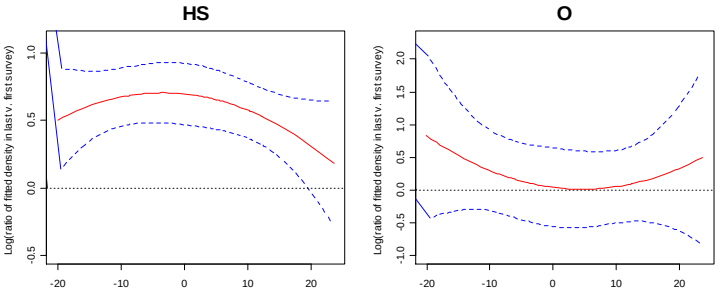
- **Generally, fish were less abundant in Luise Harbour than nearby areas, e.g.**
  - 7 out of 9 fish species with $p<0.1$ for 'Mine' effect – ranging from 45% less for *Naso lituratus* to 94% less abundant for *Naso unicornis*.
  - Every group or guild with $p<0.1$ for 'Mine' effect – ranging from 35% less for All fish, Top 24 and herbivorous croppers to 70% less for herbivorous scrapers
- **Live coral cover was much lower in Luise Harbour**
  - $p < 0.0001$ for 'Mine' effect – 45% of cover relative to that observed elsewhere

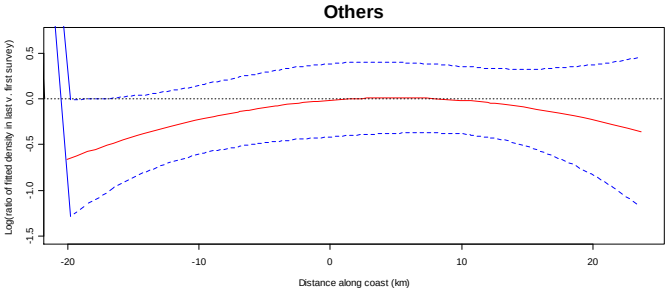
Abundance in 2007 compared with that in 1999 survey, using log-ratio of smoothed predictions

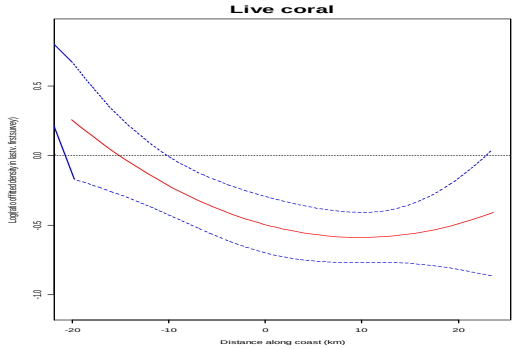












Temporal assessment of mine impact

- **For many groups and individual species, abundance increased over time, though few groups had a statistically significant increase along the entire coastline – often, a mixture of increase / no change / decrease in different places.**
- **Herbivorous scrapers increased almost universally, as did the Top 24 as a group.**
- **Two lutjanids (*L. gibbus* and *L. monostigma*) showed a significant decrease along almost all of the coastline.**
- **Live coral cover decreased markedly along the northern section of the coast.**

Statistical significance of temporal and spatio-temporal effects, and average change between first and last surveys

Species or group	Time	Time x Dist	Time x Dist_sq	Average change*
Top 24 fish species	0.0477	> 0.2	> 0.5	+
Rarer 93 fish species	> 0.5	> 0.2	> 0.2	(−)
Herbivorous scrapers (7 spp.)	< 0.0001	> 0.2	> 0.1	+
Lutjanus gibbus	< 0.0001	> 0.1	> 0.5	−
Lutjanus monostigma	0.0445	0.0715	> 0.2	−
Live coral cover	< 0.0001	0.0013	0.0559	−

* () indicates 95% C.L. for log-ratio overlap 0 for most of coastline.

Minimum detectable change (90% power) between two surveys

Species or group	Decrease	Increase
All fish	−17%	+20%
Top 24 species	−18%	+21%
Herbivorous scrapers	−22%	+25%
Benthic invertebrate feeders	−27%	+33%
Omnivores	−33%	+40%
Individual species	−37% to − 72%	+42% to +140%

Conclusions

- **Although fish community appears impoverished in Luise Harbour, impact appears not to be encroaching on other parts of the main island.**
- **Inter-diver differences are uncomfortably large; so recommended that future surveys incorporate a training/calibration day and a number of sites to be assessed by all divers.**
- **Grouping species into guilds improves power to detect change; useful if all species in the group respond similarly.**

CMIS/EI
Charis Burridge
Research statistician

Phone: +61 7 3826 7186
Email: charis.burridge@csiro.au
Web: www.csiro.au/group

Thank you

Contact Us
Phone: 1300 363 400 **or** +61 3 9545 2176
Email: Enquiries@csiro.au **Web:** www.csiro.au

CSIRO