GCRMN
Global Coral Reef Monitoring Network

ICRI
International Coral Reef Initiative

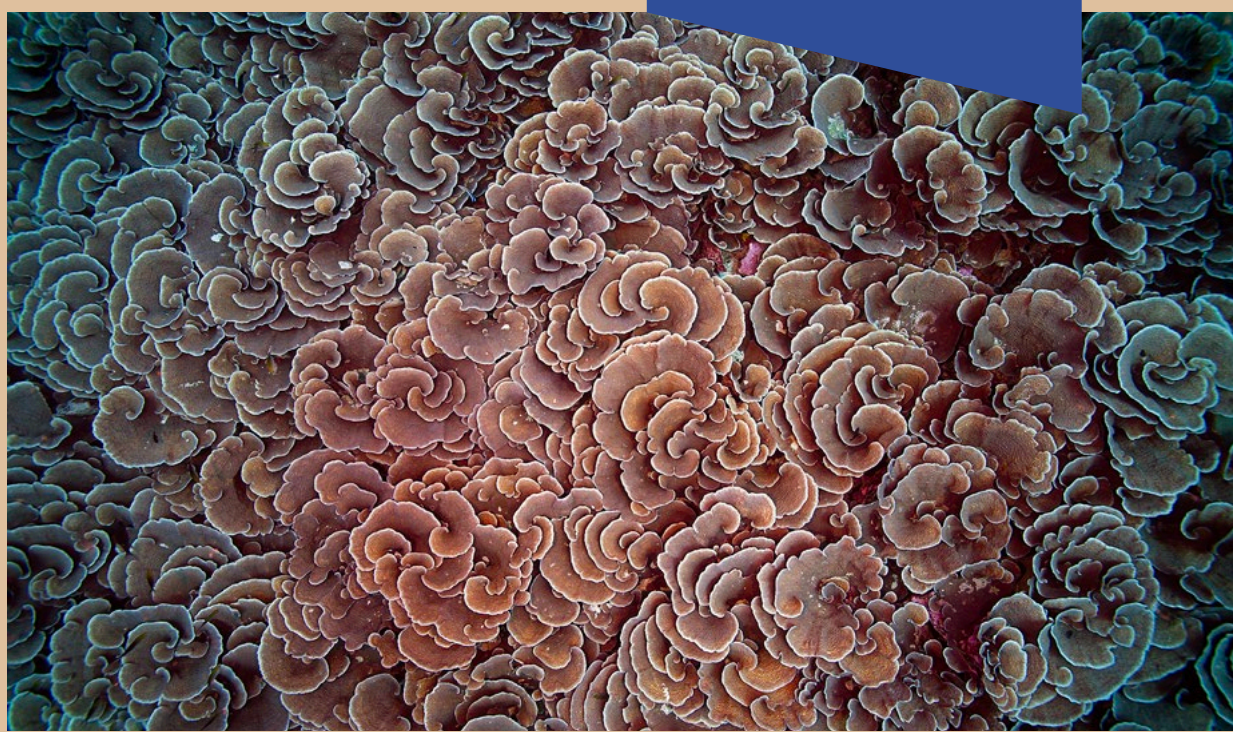
Australian Government

Australian Institute of Marine Science

Status of Coral Reefs of the World: 2020

Chapter 14. Statistical Methods

Edited by: David Souter, Serge Planes, Jérémy Wicquart, Murray Logan, David Obura and Francis Staub



Australian Government
Department of Foreign Affairs and Trade

Gouvernement Princier
Principauté de Monaco

Government Offices of Sweden
Ministry of the Environment and Energy

Prince Albert II of Monaco Foundation

UN environment programme 1972-2022

The conclusions and recommendations of this report are solely the opinions of the authors, contributors and editors and do not constitute a statement of policy, decision, or position on behalf of the participating organizations, including those represented on the cover.

Chapter 14.

Statistical Methods

Author: Murray Logan

1. Modelling considerations

The sampling design of most monitoring programs is typically tailored somewhere along a continuum between a configuration of perpetually fixed sampling sites and a configuration of randomly selected sites, depending on the purpose, resources and logistics of the program. Whist fixed sites act as their own baselines over time and thus provide relatively efficient means for estimating temporal trends, the resulting estimates are biased towards the selected locations (which may not be collectively representative of the broader area). By contrast, a configuration of uniquely random sites is less likely to be biased and hence more representative, but usually requires a considerably greater number of sites in order to detect change from within the noise.

The focus and challenge for the current report was to utilise a collection of datasets from a large number of disparate monitoring programs from around the world in order to provide estimates of status and trends at much broader spatio-temporal scales.

Whenever multiple data sets are integrated together (particularly if each is used to represent different areas), the issues of representativeness and bias are exacerbated. First, quantitative estimates are always driven by sample sizes. Within any well designed monitoring program, efforts are made to ensure the design remains relatively balanced. However, this is not the case across programs. Therefore, when aggregating multiple datasets at a broader scale, it is important to be able to control for varying samples sizes so as to minimise the risks of biasing towards the more heavily replicated datasets. Moreover, sample size and density does not necessarily reflect the density and distribution of the underlying landscape. For example, in the case of coral reefs, sampling intensity is likely to be a function of the relative prosperity of the surrounding populations and proximity to major population centers rather than the density and distribution of the reefs themselves.

Enormous (and complex) spatio-temporal models that employ full positional encoding to evaluate the spatial patterns between all possible pairs of sampling units (sites) have the potential to allow the transferal of information from the fine, observation scale measurements to the broader geographic scales of this report. By assuming that a response variable (such as percent live hard coral cover) varies continuously over an entire two-dimensional surface, such models are potentially able to leverage trends in areas of relatively high sample density to estimate the trends in neighbouring areas of sparse sampling density - albeit with greater uncertainty. However, such models proved to be too computationally burdensome and were incredibly difficult to tune to ensure they yield sensible outcomes. They also assumed that changes over space were relatively gradual and thus, can easily smooth over what would otherwise be considered abrupt local changes. Furthermore, incorporating information about the spatial distribution of reefs as well as physical barriers to auto-correlative process was far from trivial.

As an alternative, we explored hierarchical models in which sampling units were progressively aggregated with their neighbours into larger and larger units. For example, neighbouring quadrats were grouped together into sites, sites into global grid locations (see below) and so on up to the level of the entire globe. This represents a pseudo-spatial model in that although the influence of neighbouring data does deteriorate along the hierarchy, it does so in increments relating to group membership rather than as a continuous function of spatial distance. Hence, in the case of an area comprising of ten sub-areas, each of the sub-areas will share some information with the other sub-areas even though any one of them might be geographically closer to a member of another area than most of the sub-areas in its designated area (the classic nearest vs average neighbour conundrum). In any case, all attempts to fit full, global hierarchical models with the very disparate datasets proved very difficult to stabilise.

Instead, smaller hierarchical models (Fig. 1), fit separately to each Marine Ecosystem of the World (hereafter MEOW) Ecoregion[1], were integrated together within a spatially weighted aggregation hierarchy in which individual model posteriors (annual estimates) were propagated up through the hierarchy. Although this approach does still have some elements of the pseudo-spatial hierarchy that permits data poor areas to leverage patterns off data richer areas, the leveraging is quarantined to within MEOW Ecoregions where processes are more likely to be homogeneous and thus the resulting trends are more likely to be consistent with the observed data. More details about the spatial weights are discussed in section 3 and the statistical models are discussed in section 4.

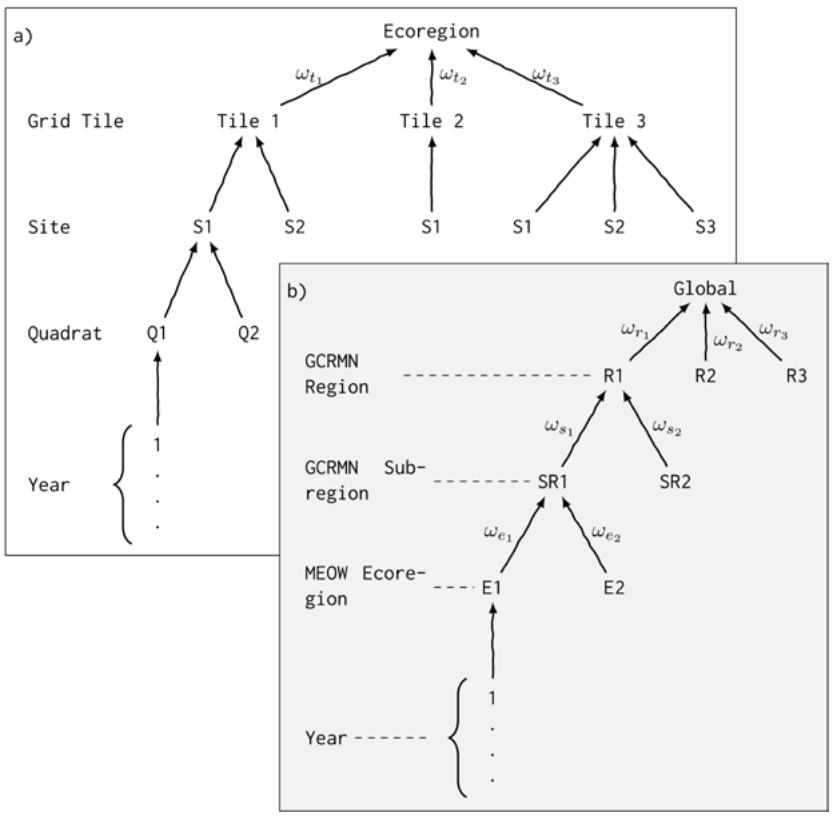
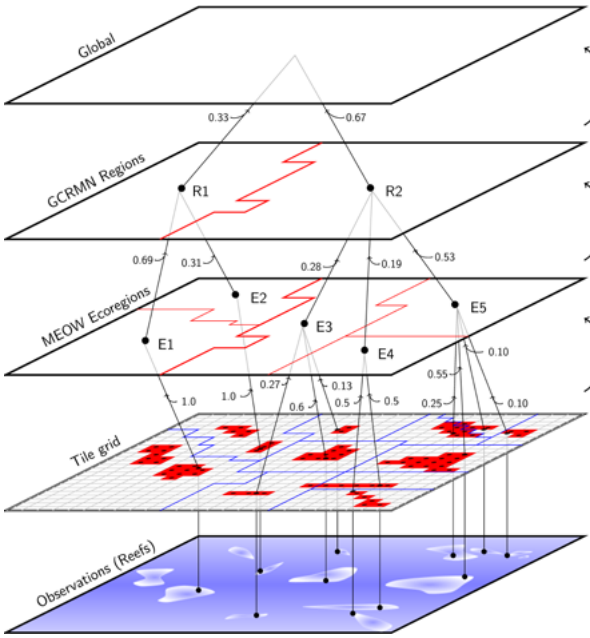


Figure 14.1. Schematic representation of the a) individual Marine Ecosystem of the World Ecoregion Bayesian modelling hierarchies and b) spatial aggregation hierarchy. Note the quadrat-level is de-emphasized to highlight that the quadrat to site level aggregation has occurred outside of the statistical model. The symbolise the use of spatial weights.

[1] Spalding, Mark D., Helen E. Fox, Gerald R. Allen, Nick Davidson, Zach A. Ferdaña, Max Finlayson, Benjamin S. Halpern, et al. 2007. "Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas." BioScience 57 (7): 573–83. https://doi.org/10.1641/B570707.



Region Weights

Region	Area	Weight
R1	26	0.33
R2	53	0.67

Ecoregion Weights

Region	Ecoregion	Area	Weight
R1	E1	18	0.69
R1	E2	8	0.31
R2	E3	15	0.28
R2	E4	10	0.19
R2	E5	29	0.53

Tile Weights

Ecoregion	Tile ID	Area	Weight
E1	T11-20	18	1.00
E2	T16-18	8	1.00
E3	T05-11	4	0.27
E3	T13-11	9	0.60
E3	T21-13	2	0.13
E4	T05-05	5	0.50
E4	T06-04	5	0.50
E5	T15-03	16	0.55
E5	T21-03	3	0.10
E5	T20-07	3	0.24
E5	T22-05	7	0.10

Figure 14.2. Schematic diagram illustrating the hierarchical structure relating the hypothetical observations (bottom layer) to the level of a 10 km x10 km grid tile, Marine Ecosystem of the World (MEOW) Ecoregions, GCRMN Regions and Global scale. The grid tile layer depicts the 10 km x10 km tile containing reef (red) and the voronoi polygons (blue lines) used to partition area zones of sample unit influence. The MEOW Ecoregions layer illustrates five fictitious Ecoregions which are aggregated into two GCRMN Regions in the layer above. Vertical lines illustrate the aggregation of data along the hierarchy and the numbers along these paths represent the aggregation weights (also tabulated).

2. Spatial hierarchy

The pseudo-spatial hierarchy outlined above necessitates incremental jumps in spatial scale from the level at which observations were collected up to the Global (or even regional) scale. However, if the jumps are too large, the information (temporal patterns) shared across neighbouring spatial units might be driven by very different underlying conditions and thus, are not appropriate.

The original datasets collated in this study were provided at scales of either quadrat/transect or spatial aggregations thereof. These can be group ed naturally into sites (or individual reefs) as the first incremental scale jump; however, subsequent increments are less obvious.

There are numerous ways of grouping coral reef locations into broader geographic areas, or alternatively, dividing the globe up spatially. Some candidates considered were: Exclusive Economic Zones[2], Veron Ecoregions[3] or Marine Ecosystems of the World[1]. Consensus amongst a large panel of coral reef regional representatives was that the MEOW global classification system was the most appropriate as it has a strong bio-geographic focus capturing important, community, evolutionary, dispersal and isolation processes[1]. The MEOW Ecoregions were further grouped up into GCRMN subregions and regions (Tab. 1) to provide additional modelling and reporting granularity.

[2] Flanders Marine Institute. 2019. "Maritime Boundaries Geodatabase: Maritime Boundaries and Exclusive Economic Zones (200nm), Version 11." Available online at http://www.marineregions.org/. https://doi.org/10.14284/386.

[3] Veron, J. E. N. 2000. Corals of the World. Vol. 1–3. Australian Institute of Marine Science, Townsville.

Table 14.1. Spatial hierarchy relating Marine Ecosystems of the World Ecoregions† to GCRMN Regions and Subregions.

GCRMN Region	GRCRMN Subregion	MEOW Ecoregion
Australia	Australia.1	142: Torres Strait Northern Great Barrier Reef
		143: Central and Southern Great Barrier Reef
		202: Tweed-Moreton
	Australia.2	140: Arnhem Coast to Gulf of Carpenteria
		141: Bonaparte Coast
		144: Exmouth to Broome
		145: Ningaloo
		210: Shark Bay
		211: Houtman
	Australia.3	120: Cocos-Keeling/Christmas Island
	Australia.4	151: Lord Howe and Norfolk Islands
Brazil	Brazil.1	074: Fernando de Naronha and Atoll das Rocas
	Brazil.2	075: Northeastern Brazil
	Brazil.3	076: Eastern Brazil
		077: Trindade and Martin Vaz Islands
	Brazil.4	071: Guianan
		072: Amazonia
Caribbean	Caribbean.1	062: Bermuda
		063: Bahamian
	Caribbean.2	064: Eastern Caribbean
		066: Southern Caribbean
	Caribbean.3	065: Greater Antilles
	Caribbean.4	067: Southwestern Caribbean
		068: Western Caribbean
	Caribbean.5	043: Northern Gulf of Mexico
		069: Southern Gulf of Mexico
		070: Floridian

GCRMN Region	GRCRMN Subregion	MEOW Ecoregion
East Asia	East Asia.1	126: Palawan/North Borneo
		127: Eastern Philippines
		128: Sulawesi Sea/Makassar Strait
	East Asia.2	129: Halmahera
		130: Papua
		131: Banda Sea
		133: Northeast Sulawesi
		138: Gulf of Papua
		139: Arafura Sea
	East Asia.3	115: Gulf of Thailand
		116: Southern Vietnam
		117: Sunda Shelf/Java Sea
		118: Malacca Strait
	East Asia.4	119: Southern Java
		132: Lesser Sunda
	East Asia.5	109: Andaman and Nicobar Islands
		110: Andaman Sea Coral Coast
		111: Western Sumatra
	East Asia.6	112: Gulf of Tonkin
		113: Southern China
		114: South China Sea Oceanic Islands
	East Asia.7	121: South Kuroshio
ETP	ETP.1	060: Cortezian
		061: Magdalena Transition
		164: Revillagigedos
		165: Clipperton
	ETP.2	166: Mexican Tropical Pacific
		167: Chiapas-Nicaragua
		168: Nicoya
	ETP.3	170: Panama Bight
		171: Guayaquil
	ETP.4	169: Cocos Islands
		172: Northern Galapagos Islands
		173: Eastern Galapagos Islands
		174: Western Galapagos Islands
PERSGA	PERSGA.1	087: Northern and Central Red Sea
	PERSGA.2	088: Southern Red Sea
	PERSGA.3	089: Gulf of Aden

GCRMN Region	GRCRMN Subregion	MEOW Ecoregion
ROPME	ROPME.1	090: Arabian (Persian) Gulf
	ROPME.2	091: Gulf of Oman
	ROPME.3	092: Western Arabian Sea
South Asia	South Asia.1	106: Chagos
	South Asia.2	105: Maldives
	South Asia.3	103: Western India
		104: South India and Sri Lanka
	South Asia.4	107: Eastern India
		108: Northern Bay of Bengal
WIO	WIO.1	093: Central Somali Coast
		094: Northern Monsoon Current Coast
		095: East African Coral Coast
	WIO.2	096: Seychelles
	WIO.3	097: Cargados Carajos/Tromelin Island
		098: Mascarene Islands
	WIO.4	099: Southeast Madagascar
		100: Western and Northern Madagascar
	WIO.5	101: Bight of Sofala/Swamp Coast
		102: Delagoa

† **Spalding, Mark D., Helen E. Fox, Gerald R. Allen, Nick Davidson, Zach A. Ferdaña, Max Finlayson, Benjamin S. Halpern, et al. 2007. "Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas." BioScience 57 (7): 573–83.** https://doi.org/10.1641/B570707.

The jump in spatial scale from Site (reef) to MEOW Ecoregion can be very large and encompass a wide range of influential processes and drivers. Therefore, we sought an additional intermediate scale. Such a scale could be based on collections of reefs or broad communities, although such information was not universally available. An intermediate scale could also be achieved by dividing the globe up into an array (grid) of cells or tiles of a constant size. Moreover, the use of grid tiles provided a way of abstracting away design differences between fixed and random annual site selections thus, providing a mechanism by which multiple sampling designs could be incorporated in the one model.

3. Spatial Weights

In order to help maximise the chances that the hierarchical aggregations were reflective of broad spatial patterns and not heavily biased by sampling effort alone, the aggregations were weighted by the proportion of reef area represented by each spatial unit.

Estimating the distribution and area of global coral reefs is a challenging problem. As is the case with sampling effort consistency across the globe, the granularity and accuracy of coral reef mapping varies substantially from region to region. New initiatives such as the Allen Coral Atlas will help to address these challenges as satellite imagery improves and algorithms mature and achieve recognition

and acceptance within the broader scientific community. However, for the purposes of weighting analyses underpinning this *Status of Coral Reefs of the World: 2020* report, Tropical Coral Reefs of the World[4,5] digital shapefiles were used to provide a potentially less biased and more uniform method of estimating coral reef area. The intermediate spatial scale between observed sites and MEOWs was provided by generating a 10 km x10 km grid of tiles across the entire globe and assigning a unique identifier to each tile.

3.1 Tile level weights

All observed site level locations were assigned to a grid tile on the basis of nearest neighbour within 10 km. To estimate the amount of reef area within each MEOW that was represented by each of the observed sites, voronoi polygons were generated from the unique site locations and overlayed onto the grid (Fig. 3). The reef area associated with each voronoi cell was then expressed as a proportion of the total MEOW reef area, thereby representing the relative weight that each grid tile should carry in the analyses.

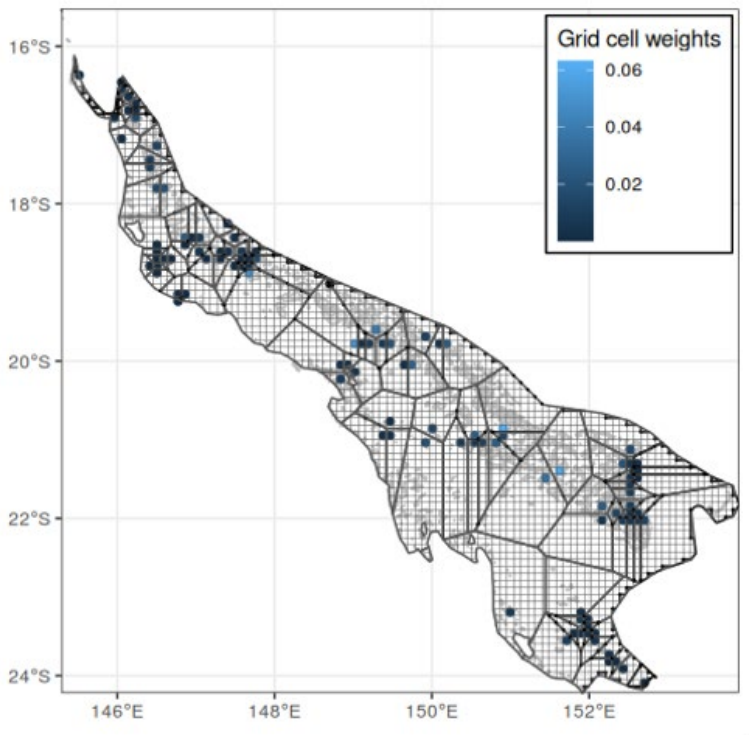


Figure 14.3. Illustration of voronoi polygons overlayed on the 10 km x10 km grid and reefs (grey). Shaded grid tiles represent grid tiles containing observed sites and colour of the grid tile shading represents the relative weights (proportion of reef area in each grid tile).

[4] World Resources Institute. Tropical Coral Reefs of the World (500-m resolution grid), 2011. Global Coral Reefs composite dataset compiled from multiple sources for use in the Reefs at Risk Revisited project incorporating products from the Millennium Coral Reef Mapping Project prepared by IMaRS/USF and IRD.

https://datasets.wri.org/dataset/tropical-coral-reefs-of-the-world-500-m-resolution-grid

[5] Burke, L., K. Reytar, M. Spalding, and A. Perry. 2011. "Reefs at Risk Revisited." Washington, DC, USA: Word Resources Institute.

3.2 Larger scale weights

The weights (relative contributions) of each MEOW Ecoregion in aggregating up to GCRMN subregions was calculated as the proportion of MEOW reef area within each GCRMN subregion (see Fig. 4). Similarly, GCRMN subregion and region weights (used in aggregations to GCRMN Region and Global levels respectively) were calculated from the respective proportions of reef areas in GCRMN regions and globally.

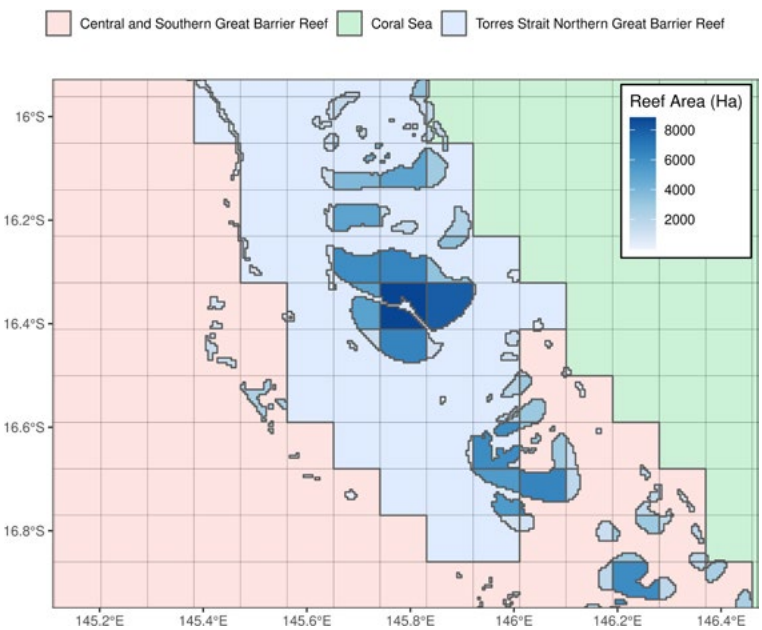


Figure 14.4. Illustration of the relative reef area represented by each 10 km x10 km grid tile within three example MEOW Ecoregions in the GCRMN Australia Region. The colour of reef fill is proportional to the relative area of reef in the MEOW.

4. Statistical Models

Live hard coral cover and algal cover were calculated by summing observation level data across associated taxonomic groupings.

Separate MEOW Ecosystem Bayesian hierarchical models were constructed within the stan statistical modelling platform[6] via the rstan[7] interface. Each model comprised a model matrix representing year dummy coded as cell means contrasts, a model matrix representing Dataset coded as sum to zero contrasts as well as varying effects representing the hierarchical structure of Sites nested within grid tiles (Fig. 1). Weights were also applied to the grid tiles in order to allow the influence of each grid tile to be proportional to the relative area of reef present within each grid tile.

Separate models were fitted to explore trends in live hard coral cover (HCC) and algae cover (A). In each case, cover was modelled against a beta distribution (logit link). Cover values of either 0 or 1 were first

[6] Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." Journal of Statistical Software, Articles 76 (1): 1–32. https://doi.org/10.18637/jss.v076.i01.

[7] Stan Development Team. 2019. "RStan: The R Interface to Stan." http://mc-stan.org/.

shrunk by 0.01 for compatibility with the beta distribution. Weakly informative priors were applied to the beta shape parameters as well as the varying effects parameters and their standard deviations.

In order to impute missing year combinations and smooth over short-term oscillations in estimates resulting from short-term fluctuations in sampling designs and data availability, priors on Year effects (except that associated with the first observed year of data in a MEOW Ecoregion) were weakly informative normal priors centred around the posterior of either previous Year (in the case of Years after the initial observed year) or after (in the case of Years prior to the initial observed year). For the initial observed Year, standard (zero centred) weakly informative priors were applied.

The Dataset effects were included to act as proxies for all the many and varying ways that different datasets differ including depth, sampling unit type (quadrats, transects, etc) and observer experience. Weakly informative normal priors were applied to the Dataset effects.

The statistical models can be summarised as:

$$
\begin{aligned}
y &\sim Beta(\mu\phi, (1-\mu)\phi) \\
logit(\mu) &= \boldsymbol{\beta_y X_y} + \boldsymbol{\beta_d X_d} + \boldsymbol{\gamma_s} Z_s + \omega \boldsymbol{\gamma_t} Z_t \\
\boldsymbol{\gamma_s} &= \sigma_s \times z_s \\
\boldsymbol{\gamma_t} &= \sigma_t \times z_t \\
\phi &\sim \Gamma(0.01, 0.01) \\
z_s, z_t &\sim N(0,1) \\
\sigma_s, \sigma_t &\sim t(3,0,10) \\
\boldsymbol{\beta_{y_i}} &\begin{cases} i = O_y, \sim N(0, \sigma_y) \\ i > O_y, \sim N(\beta_{s_{i-1}}, \sigma_{y1}) \\ i < O_y, \sim N(\beta_{s_{i+1}}, \sigma_{y1}) \end{cases} \\
\boldsymbol{\beta_d} &\sim N(0, 0.1)
\end{aligned}
$$

where y is the cover of either live hard coral or algae, β_y and β_d represent the effects of Year and Dataset respectively, X_y and X_d represent cell-means Year and sum-to-zero Dataset model matrices respectively, γ_s and γ_t are the sum-to-zero varying effects, and Z_s and Z_t represent the Site and grid tile codes respectively. o_y represents the initial observed Year within the MEOW Ecoregion and i is a year iterator.

Trends in hard coral cover to algae ratio (HCC:A) employed structurally very similar models to those described above, yet the ratio was modelled against a Gaussian distribution.

All models were run with 10,000 no-u-turn MCMC iterations, a warmup of 5000 and a thinning rate of 5 for each of three chains (each with random initial values). Diagnostics indicated that all chains converged on stable, well mixed posteriors (Rhat values < 1.05) and low MCMC sample auto-correlation (< 0.2).

4.1 Hierarchical aggregation

The full posteriors for the Year effects (on the logit scale) of each MEOW Ecoregion were averaged together within each GCRMN subregion (Fig. 2). The resulting posteriors were then summarised by back-transforming to the response scale (inverse logit transform in the case of beta models) and calculating the means and highest probability density intervals (80% and 95%). Similarly, the un-standardised GCRMN subregion posteriors were aggregated (with weights) up to GCRMN Region and then Global level (Fig. 2).

4.2 Half-decadal pairwise contrasts

The modelled trends in the covers of live hard coral and algae provide a visual representation of the annual fluctuations within the long-term patterns. In addition, there was a need to be able to provide quantified estimates of the degree of medium to long term changes over time. To achieve this, we combined together the modelled posteriors (hard coral and algae separately) into half-decadal time

units (working backwards from the most recent year of available data) and compared the three most recent half-decadal time units (2004-2009, 2010-2014 and 2015-2019) together in a pairwise manner. From each pairwise contrast, we calculated the exceedance probabilities of both positive and negative changes as well as the associated mean absolute and percentage changes.

Half-decadal time units were chosen as they provided a convenient way to evenly partition the time since the last GCRMN *Status of Coral Reefs of the World* report in 2008, encompassed the time span for which reefs where more extensively monitored and provided a good compromise between short and long-term intervals. Whilst it is recognised that the selected half-decadal time boundaries might mask the impact of some local disturbance events, we considered that it was important to maintain consistent time units across the entire scope of the analyses for the purpose of comparability of discussions.

4.3 Proportion of sampling units not recovering

Disturbances are a natural driver within any ecosystem. Nevertheless, over time, a healthy and resilient ecosystem would be expected to recover from disturbances. For sampling units (typically quadrats or transects) that were repeatedly monitored for at least 15 years, we enumerated the number of these units that had experienced a relative decline in raw (un-modelled) hard coral cover of at least 20 percent. We then calculated the percentage of these units that had subsequently recovered to within 90 percent of their pre-decline cover. To provide greater insights about the changes within these units, we also calculated the mean maximum absolute and percentage coral cover declines as well as the long-term (difference in cover between first and last sampling time) mean maximum absolute and relative declines in coral cover. Similar analyses were performed on incidences of algae cover increases and subsequent declines.

For the above calculations it was critical that only fine-scale sampling units (e.g. quadrats/transects) were used rather than higher scale locations such as Sites. This is because benthic data can vary enormously even at fine scales and thus comparing Site level data that comprise different sampling units over time will likely yield very distorted apparent declines and recoveries.

4.4 Sea Surface Temperature Anomalies

The above analyses provided the first large scale, quantitative estimates of the status and trends in the covers of live hard coral and algae. The resulting trends showed clear indications of fluctuations in hard coral cover at a global scale. Since numerous incidences of coral cover decline (both regionally and globally) had reportedly been attributed to coral bleaching resulting from elevated sea surface temperatures, we explored associations between the global trends in live hard coral cover and global trends in sea surface temperature and other climatic indices (e.g. ENSO).

HadSST4 is a global dataset that provides gridded (5x5 degrees) sea surface temperature anomalies across the world as well as global monthly averages[8]. The HadSST4 data were restricted to the temporal range of 1977 to 2020 so as to coincide with the availability of observed benthic data collated for this report. An 18 month rolling mean was used to smooth the trend in HadSST4 anomaly. The relative rate of change in smoothed HadSST4 per unit of time was estimated by calculating derivatives via finite differences.

[8] Kennedy, J. J., N. A. Rayner, C. P. Atkinson, and R. E. Killick. 2019. "An Ensemble Data Set of Sea□surface Temperature Change from 1850: The Met Office Hadley Centre HadSST.4.0.0.0 Data Set." Journal of Geophysical Research: Atmospheres 124: 7719–63. https://doi.org/10.1029/2018JD029867.

After overlaying the smoothed HadSST4 trend and associated derivatives over the trend in global hard coral cover, a number of features became apparent. Periods of coral cover decline appeared to be associated with either smoothed HadSST4 anomalies that exceeded 0.45 or when the rate of smoothed HadSST4 change exceeded 0.15 for two consecutive months. Whilst these are not strictly statistical tests, they do provide the basis of future statistical explorations.

The above associations were communicated visually by plotting smoothed HadSST4 anomaly trend over the trend in global hard coral cover and overlaying vertical light red bars (50% opacity) to indicate when the rate of smoothed HadSST4 change exceeded 0.15 for two consecutive months and vertical dark red bars (20% opacity) to indicate when the smoothed SST anomaly exceeded a value of 0.45.

CORDIO

CRIOBE
USR3278
Centre de Recherches Insulaires et Observatoire de l'Environnement