6. HIERARCHICAL AGGREGATIONS

6.1 Theoretical framework

To facilitate the integration of additional input Measures into the report card scores (such as additional Physical or Chemical), or even additional Sub-indicators (such as sediment metals, aquaculture yields etc), we can defined a hierarchical structure in which Measures (such as Chlorpohyll-a, NOx, sediment aluminum and yield etc) are nested within appropriate Sub-indicators. In turn, these Sub-indicators are nested within Indicators.

By progressively abstracting away the details of the Measures and Sub-indicators, a more focused narrative can be formulated around each level of the hierarchy. For example, when discussing the current state (and trend in state) of the Water Quality Indicator, rather than needing to discuss each individual constituent of Water Quality, high-level Grades are available on which to base high-level interpretations. More detailed explorations are thence revealed as required by exploring the Grades at progressively finer scales of the hierarchy. Moreover, the hierarchical structure offers great redundancy and thus flexibility to add, remove and exchange individual measures.

Similar arguments can be made for a spatial hierarchy in which Sites are nested within Zones which in turn are nested within the Whole GBR.

The purpose of aggregation is to combine together multiple items of data. For Nesp 3.2.5, the report card is informed by a triple hierarchical data structure in which Daily observations are nested within Seasonal and Annual aggregates, Measures are nested within Sub-indicators which are nested in Indicators and Sites are nested within Zones (see Figure 74).

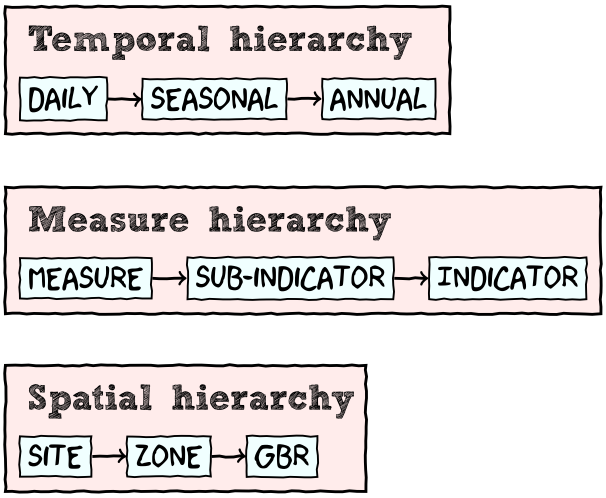


Figure 74: Temporal, measure and spatial aggregation hierarchy

Although the triple hierarchy (temporal, Spatial and Measurement), does offer substantial redundancy and power advantages, it also introduce the complexity of how to combine the hierarchies into a single hierarchical aggregation schedule. Table 15 (a fabricated example), illustrates this complexity for aggregating across Spatial and Measure scales when data availability differs. This simple example demonstrates how different aggregation schedules can result in different Zone Indicator scores:

- calculating Zone 1 Indicator Score as the average of the Site level Water Quality Scores prioritizes that the Zone 1 Indicator Score should reflect the average of the Water Quality Indicator Scores for the Site. This routine will bias the resulting Zone 1 Water Quality Indicator Score towards Sub-indicators represented in more Sites. The current MMP sampling design is unbalanced (some Zones have more Sites than others and not all Measures are observed in all Sites), and there is no guarantee that the design will be maintained over time. If for example, Chemical Measures were not available for certain Zones, then the Whole GBR Water Quality Indicator Score will be biased towards Water Clarity Sub-indicators.
- calculating Zone 1 Water Quality Indicator Score as the average of the Zone 1 level Sub-indicator Scores prioritizes equal contributions of Sub-indicators to the Indicator Score at the expense of being able to relate Zone 1 Scores to the corresponding Site Scores.

The above becomes even more complex when the temporal dimension is include..

Table 15: Fabricated illustration of the discrepancies between total means (i.e. Zone 1 Indicator Score) generated from row means (Site Sub-indicator Scores) and column means (Zone 1 Sub-indicator Scores).

	Sub-indicators		
Site	**Water Clarity**	**Nutrients**	**Indicator**
1	5	2	3.50
2	6		6.00
3	6	4	5.00
Zone 1	5.67	3.00	X

If X (mean) is calculated from the three row means = 4.83
If X (mean) is calculated from the two column means = 4.33

An additional complication is how the different hierarchies integrate together. Specifically, what level of data should be aggregated first and at what point do the aggregations of one hierarchy feed into other hierarchies. For example, should observations first be aggregated from Daily to Seasonal or Annual, then aggregated from Site level to Zone level and then finally aggregated from Measure to Indicator? Some possible configurations are presented in Figure 75.

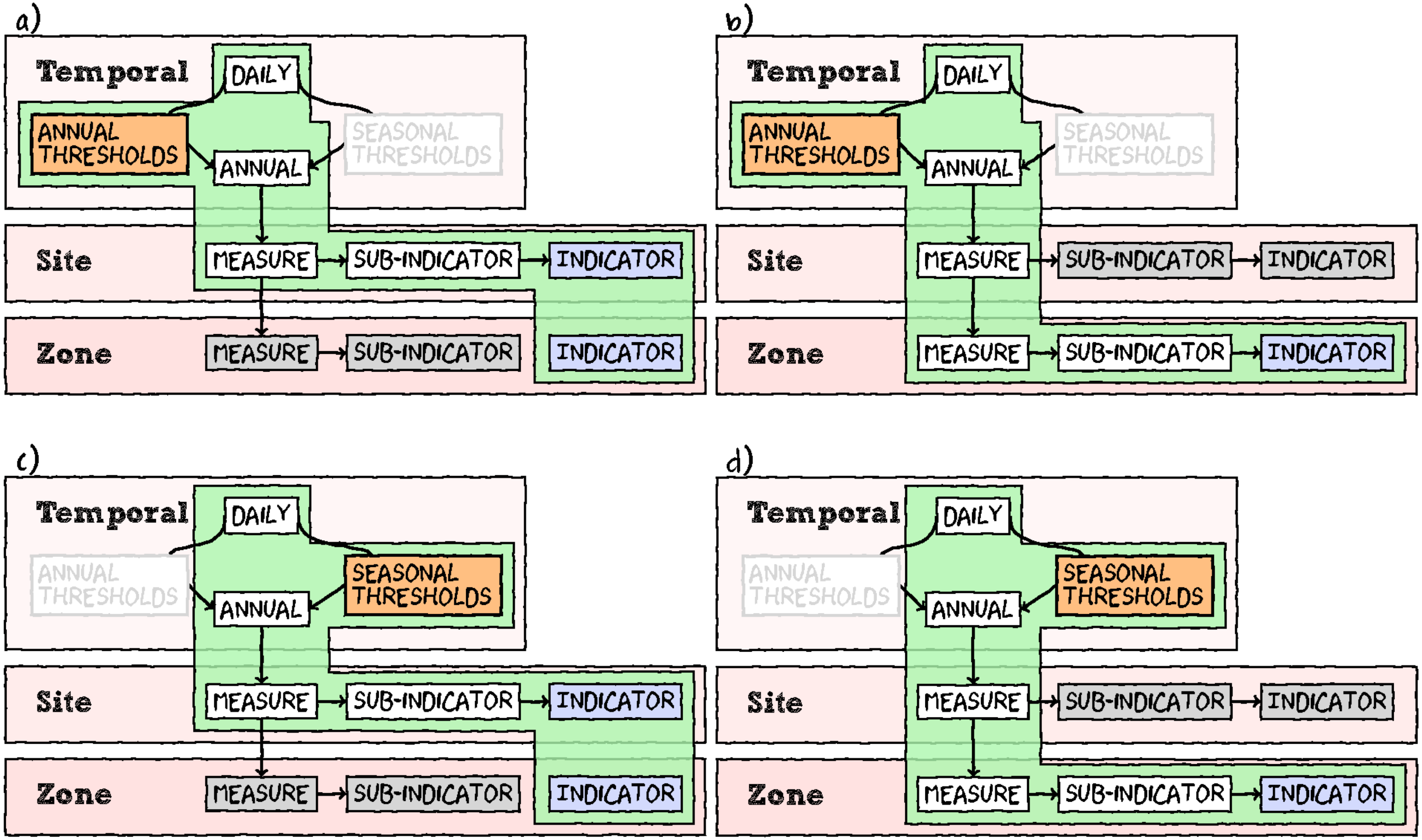


Figure 75: Schematic illustrating four possible aggregation routines through the combination of Temporal (Daily, Seasonal and Annual), Spatial (Site, Zone) and Measure (Measure, Sub-indicator, Indicator) nodes of the triple hierarchical aggregation routine associated with the GBR Report Card. Aggregation directions between nodes are signified by arrows and the main aggregation pathway through the routines is illustrated by the green polygon.

To maximize information retention throughout a series of aggregations, it is preferable to aggregate distributions rather than single properties of those distributions (such as means). The simplest way to perform a hierarchy of aggregations is to interactively calculate the means (or median) of items (means of means etc). At each successive aggregation level only very basic distributional summaries (such as the mean and perhaps standard deviation) are retained, the bulk of upstream information is lost. Alternatively, more complex methods that involve combining data or probability distributions can be effective at aggregating data in a way that propagates rich distributional properties throughout a series of aggregations.

Importantly, if the purpose of aggregation is purely to establish a new point estimate of the combined items, a large variety of methods essentially yield the same outcomes. On the other hand, if the purpose of aggregation is also to propagate a measure of uncertainty or confidence in the point estimate through multiple hierarchical levels of aggregation (as is the case here), then the different methodologies offer differing degrees of flexibility and suitability.

Hierarchical aggregations are essentially a series of steps that sequentially combine distributions (which progressively become more data rich). The resulting distribution formed at each step should thereby reflect the general conditions typified by its parent distributions and by extension, each of the distributions higher up the hierarchy.

Numerous characteristics can be estimated from a distribution including the location (such as mean and median) and scale (such as variance and range). For the current project, the mean and variance were considered the most appropriate[12] distributional descriptions and from these estimates Grades and measures of confidence can be respectively derived. Hence the numerical summaries (mean and variance) at any stage of the hierarchical aggregation are a byproduct rather than the sole property of propagation.

6.1.1 Bootstrap aggregation

Although some of the items to be aggregated together might initially comprise only a few values (or even a single value), it is useful to conceptualize them as continuous distributions. For example, when aggregating multiple *Measures* (such as all Water Quality Chemicals) together to generate a (*Site* level) *Sub-indicator* average, each *Measure* in each *Site* can be considered a distribution comprising the single Score for that *Measure*. Aggregation then involves combining together the multiple distributions into a single amalgam (by adding the distributions together, see Figure 76). Similarly, when aggregating at the *Indicator* level across *Site* to generate *Zone* summaries for each *Indicator*, *Site* distributions are respectively added together to yield a single distribution per *Zone*.

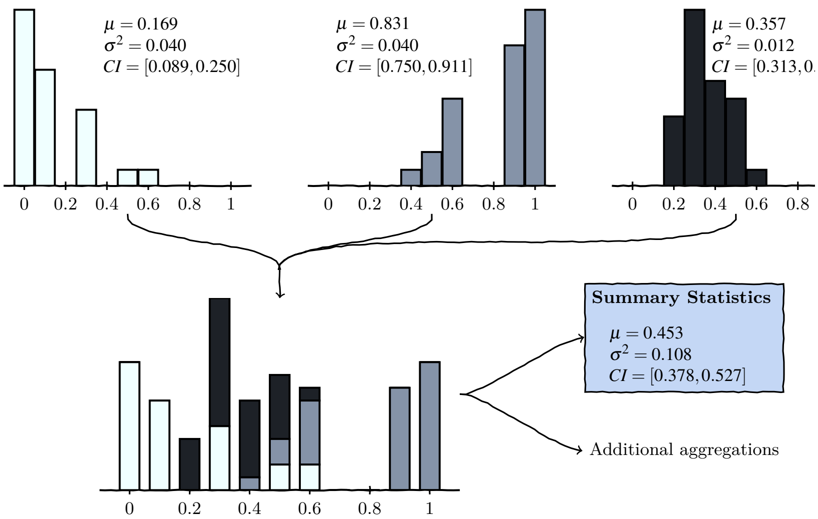


Figure 76: Illustration of Bootstrapped aggregation of three distributions. Simple summary statistics (mean, variance and 95% confidence interval presented for each distribution).

If the distributions being aggregated are all proportional distributions (e.g. density distributions), adding them altogether is trivially simple. However, if, rather than actual distributions, the items to be aggregated are ac-

[12]The aggregations typically involve some Measures with a small number of unique observations (and thus indices) and thus means and variances provide greater sensitivity than medians and ranges. Moreover, the indexing stage effectively removes outliers and standardizes the scale range thereby reducing the need for robust estimators.

tually just small collections of values (as is the case for many of the discrete Measures here) or even large, yet unequally populous collections of values (as could be the case for Continuous Flow Monitoring with missing or suspect observations), then simply aggregating the distributions together will result in amalgams that are weighted according to the size of the collections (larger collections will have more influence). For example, if we were aggregating together three *Zones* (to yield Whole GBR estimates), one of which comprised twice as many Sites, simple aggregation of distributions would result in a distribution that was more highly influenced by the Zone with the more *Sites*. Similarly, when aggregating from the level of Sub-indicator to the level of Indicator, the resulting Indicator would be biased towards the Sub-indicator with the most Measures. Whilst this may well be a useful property (e.g. stratified aggregation), it may also be undesirable.

Bootstrapping is a simulation process that involves repeated sampling (in this case with replacement) of a sample set with the aim of generating a bootstrap sample from a distribution. This bootstrap sample can be used to estimate the underlying probability distribution function that generated the data as well as any other summary statistics. Importantly, bootstrapping provides a way to generate distributions that are proportional and thus un-weighted by the original sample sizes thereby facilitating un-weighted aggregation[13]. Bootstrapped distributions can be aggregated (added together) to yield accumulated child distributions that retain the combined properties of both parents (see Figure 76). As a stochastic process, repeated calculations will yield slightly different outcomes. Nevertheless, the more bootstrap samples are collected, the greater the bootstrap distributions will reflect the underlying Score distribution and provided the number of drawn samples is sufficiently large (e.g. 10,000 re-samples), repeated outcomes will converge.

To reiterate, the advantage of bootstrapping data before concatenating (or averaging) versus simply concatenating data from multiple sources together, is to ensure that source data are all of exactly the same sample size (so as to not weight more heavily towards the more populous source(s)[14]). Bootstrapping also provides a mechanism for propagating all distribution information throughout an aggregation hierarchy and ensures that estimates of variance derived from child distributions are on a consistent scale[15]. The latter point is absolutely critical if variance is going to be used to inform a Confidence Rating system and confidence intervals.

Minimum operator procedures are supported by filtering on the lowest performed indicator prior to bootstrapping. Importantly, the bootstrapping routine simply provides a mechanism to collate all sources together to yield a super distribution. Thereafter, the joint distribution can be summarized in what ever manner is deemed appropriate (arithmetic, geometric, harmonic means, medians, variance, range, quantiles etc). Moreover, different levels of the aggregation can be summarized with different statistics if appropriate.

6.1.2 Beta approximation

Whilst the bootstrap aggregation approach described above does offer a robust way to combine data across scales and sources, for large data sets, it does impose large computational and storage burdens. For such cases (large data such as remote sensing), index distributions can be approximated by beta distributions. The beta distribution is defined on the interval [0,1] and is parameterized by two positive shape parameters (α, β) according to the following:

$$f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

A beta function can manifest as many different shapes and as all of these are described by just two shape parameters. Therefore, rather than store all the bootstrapped values for each distribution, we can alternatively approximate each distribution by a beta and store only the defining shape parameters of each distribution. When combining, rather than randomly sample 10,000 stored values of each distribution, we simple resample 10,000 random draws from each beta distribution[16]. The combined distribution can then be approximated by a beta distribution and so on.

6.1.3 Weights

Standard bootstrapping yields equally weighted distributions, however, specific weighting schemes can also be easily applied by bootstrapping in proportion to the weights. For example, to weight one parent twice as high as

[13] technically, all equally weighted rather than un-weighted
[14] Such weightings should be handled in other ways if at all
[15] Variance is inversely proportional to sample size
[16] Unfortunately there is no closed-form general formula for the sum of multiple independent beta distributions.

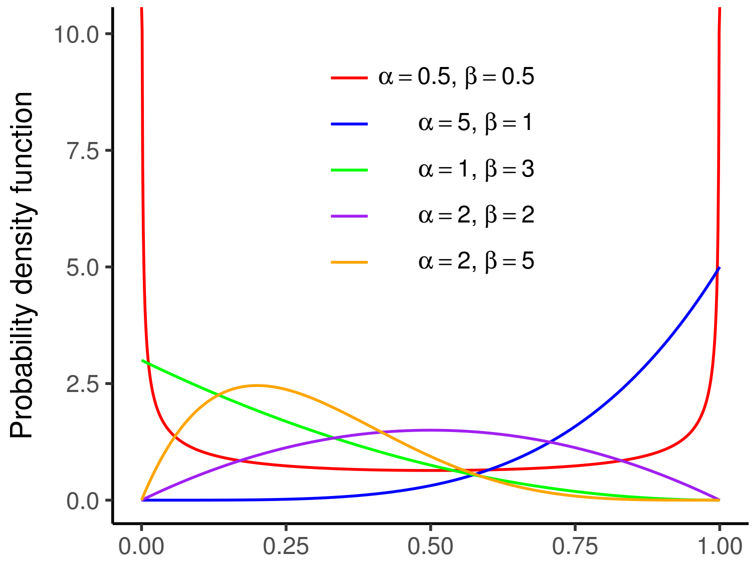


Figure 77: Beta probability densities

another, simply collect twice as many re-samples from the first distribution. To ensure that all resulting distributions have the same size (by default 10,000 items), the number of bootstrap samples collected (n) from each of the (p) parent distributions (i), given the weights (w_i) is calculated as:

$$n_i = (S/p) \times w_i$$

where S is the target size (10,000) and . indicates the ceiling. Qualitative data (such as ratings) can also be incorporated by enumerating the categories before bootstrapping.

In addition to allowing expert driven weights that govern the contribution of different items during aggregations, it is possible to weight according to relative spatial areas during spatial aggregations. Currently, all Sites are equally weighted when aggregating to Zone level and all Zones equal when aggregating to Whole of GBR level. That means that small Zones have an equal contribution as large Zones despite representing a smaller fraction of the water body. Area based weights could be applied such that Sites and Zones contribute in proportion to relative areas.

Weights are defined by a user editable configuration file that is similar in structure to the Water Quality thresholds file.

6.1.4 Expert interventions

The ability for experts and Report Card managers to intervene (exclude or overwrite) Scores/Grades at any Spatial/Measure scale is essential to maintain the quality of a Report Card in the event of unrepresentative or suspect data. The current system is able to support expert interventions in the form of exclusions and overwrites. For example, after reviewing the QAQC, an expert can elect to exclude one or more Measures (or Subindicators etc) from one or more spatial scales. Such interventions are specified via a user editable configuration files[17] (csv) that is similar in structure to the Water Quality thresholds file.

The essential component of this configuration file is that it allows a user to specify what Data are to be excluded or replaced. These can be at any of the levels of the Measure hierarchy (Measures, Sub-indications and Indicators) and any level of the Spatial hierarchy (Sites, Zones and Whole GBR). Settings pertaining to levels further along the aggregation hierarchies have precedence. For example, if Chemicals are excluded (or overridden) in a particular Zone, then all Chemical Measures within all Sites will be excluded irrespective of what the settings are for any specific Measure/Site.

[17]Since aggregation occurs across two hierarchies (the Measure hierarchy and the Spatial hierarchy - see Figures 74 and 75), two configuration files are necessary.

6.1.5 Scores and Grades

The double hierarchy Bootstrap aggregation described above, yields **Score** distributions for each Measure-level/Spatial-level combination. The location and scale of each distribution can thus be described by its mean and variance. Mean **Scores** are then converted into a simple five-point alphanumeric **Grade** scale (and associated colors) using a control chart (see Figure 78).

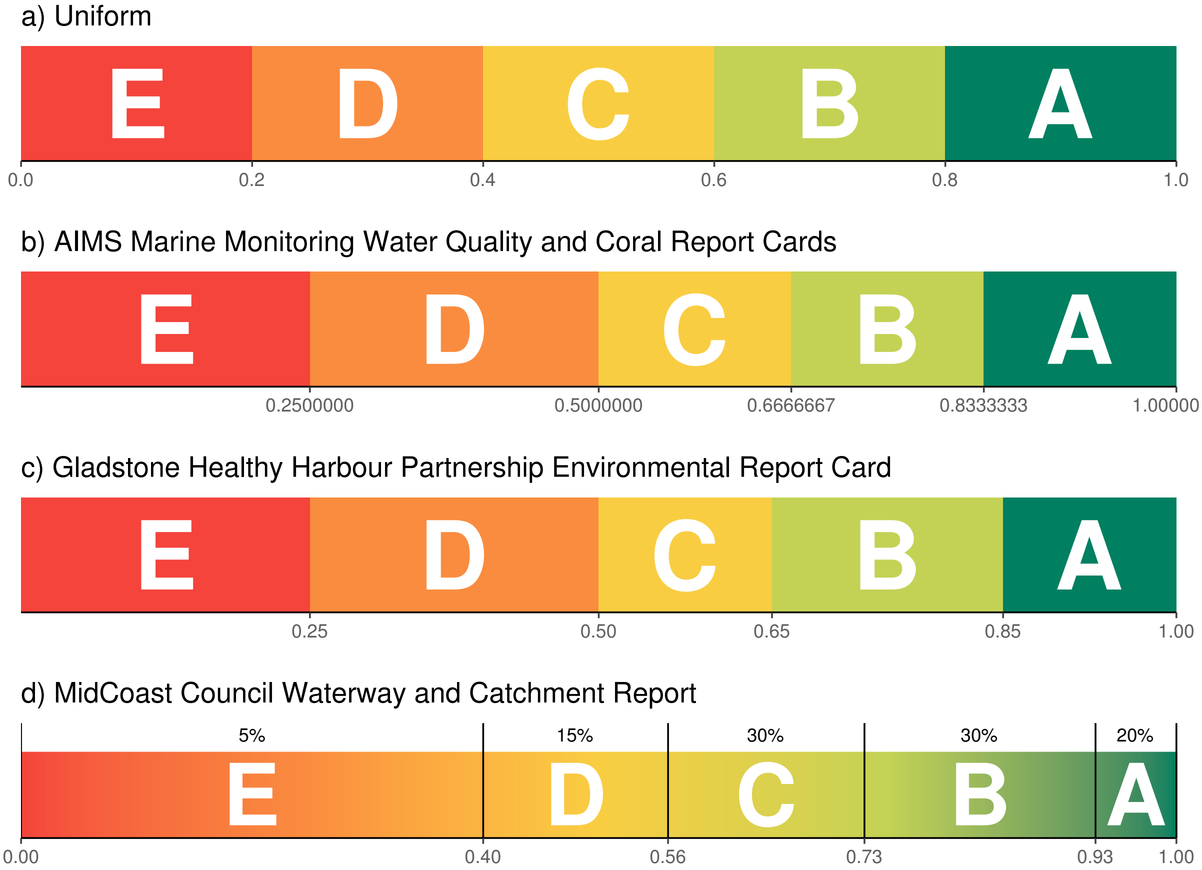


Figure 78: Score to grade conversion control charts. In each case, the scale along the base defines the grade boundaries.

The control charts adopted by the AIMS inshore water quality Marine Monitoring Program (MMP Lønborg et al., 2016) and the Gladstone Healthy Harbour Partnership (Gladstone Healthy Harbour Partnership, 2016) both define two levels (Poor and Very Poor) under the Threshold values and three above (Satisfactory, Good and Very Good). The threshold is purposely placed at the boundary of two grades so as to ease the distinction between 'pass' and 'fail'. The major difference between these two charts is that whereas the AIMS MMP report card control chart partitions the three better than threshold categories, the Gladstone Healthy Harbour Partnership report card control chart employs simpler boundary cutoffs around the 'B' grade (although this does result in arbitrarily unequal category sizes.

By contrast, the MidCoast Council (formally Great Lakes Council) Waterway and Catchment Report (MidCoast Council, 2016) uses grade boundaries based on historical score distribution quantiles associated with definitions of what proportion of total observations (sites) are considered 'Excellent' (A), 'Good' (B), 'Fair' (C), 'Poor' (D) and 'Very Poor' (Fig. 78d). For example, the 'Very Poor' grade was defined as the worst 5% of sites across the entire State of New South Wales and the lowest 5% of sites has a maximum score of 0.4. This approach recognizes the non-linear spread of scores resulting from their particular metrics and attempts to ensure that grades are intuitively interpretable (A grade of A means the site is in Excellent condition). Nevertheless, it does necessitate a of historical data and as well as a very specific and agreed upon set of a priori condition definitions.

In each of the above approaches, grade boundaries are usually determined to some extent by expert panel to ensure that the range of indices represented by each grade classification is congruent with community interpretation of a letter grade report cards. It is far less clear how estimates of uncertainty can be incorporated into such a grading scheme in a manner that will be intuitive to non-technical audiences. That said, statistical uncertainty is just one of many sources of un- certainty that should be captured into a confidence or certainty rating. Hence any expectations of presenting uncertainty in a quantitative manner may well be unrealistic anyway.

In the absence of expert opinion, we have elected to adopt a very simple score-grade control chart in which the score range is simply partitioned into five equal grades (Fig. 78a).

6.1.6 Certainty rating

Incorporating an estimate of scale (variance) into a certainty or confidence rating necessitates re-scaling the estimates into a standard scale. In particular, whereas a scale parameter of high magnitude indicates lower degrees of certainty, for a certainty rating to be useful for end users, larger numbers should probably represent higher degrees of certainty. Thus, the scaling process should also reverse the scale. Furthermore, variance is dependent on the magnitude of the values.

In order to re-scale a scale estimate into a certainty rating, it is necessary to establish the range of values possible for the scale estimate. Whilst the minimum is simple enough (it will typically be 0), determining the maximum is a little more challenging depending on the aggregation algorithm (bootstrapping, Bayesian Network etc). One of the advantages in utilizing proportional distributions (such as is the case for a Bayesian Network or a re-sampled bootstrap distribution) is that the scale parameter for the single worst case scenario can be devised (once the worst case scenario has been determined) independent of sample sizes or weightings. In most situations this is going to be when the distribution comprises equal mass at (and only at) each of the two extremes (for example, values of just 0 and 1).

The measure of confidence rating discussed above is purely an objective metric derived from the variance in the aggregation hierarchy. It is completely naive to issues such as missing data, outliers and Limit of Detection issues - the influences of which on a confidence rating are necessarily subjective. A full Confidence Rating would combine these objective variance component with additional subjective considerations such as climatic and disturbance information, and the perceived influence of missing, Limit of Detection and outlying data. Hence, the statistical scaled statistical variance would form just one component in the Confidence Rating system.

The bootstrap aggregation method provides a mechanism for estimating variance from which to build such an expert considered Confidence Rating system.

Table 16 presents the Water Quality Indicator Scores and associated Grades for each Zone based on three of the grade control chart types described in Figure 78 for the eReefs data indexed using the fsMAMP formulation. Whilst there is some agreement between the different grade types, in general, the Uniform type yields higher grades than either MMP or GHHP.

Table 16: Score and associated Grades based on three different grade control charts (Uniform, MMP and GHHP) for eReefs data indexed via fsMAMP and aggregated to Zone/Indicator level.

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
Cape York	Open Coastal	2014	0.692	B	B	B
Cape York	Open Coastal	2014	0.791	B	B	B
Cape York	Open Coastal	2014	0.891	A	A	A
Cape York	Open Coastal	2014	0.856	A	A	A
Cape York	Open Coastal	2015	0.741	B	B	B
Cape York	Open Coastal	2015	0.906	A	A	A
Cape York	Open Coastal	2015	0.880	A	A	A
Cape York	Open Coastal	2016	0.757	B	B	B
Cape York	Open Coastal	2016	0.837	A	A	B
Cape York	Open Coastal	2016	0.917	A	A	A
Cape York	Open Coastal	2016	0.891	A	A	A
Cape York	Midshelf	2014	0.716	B	B	B
Cape York	Midshelf	2014	0.805	B	A	B
Cape York	Midshelf	2014	0.894	A	A	A
Cape York	Midshelf	2014	0.862	A	A	A

..continued from previous page

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
Cape York	Midshelf	2015	0.764	B	B	B
Cape York	Midshelf	2015	0.838	A	A	B
Cape York	Midshelf	2015	0.913	A	A	A
Cape York	Midshelf	2015	0.891	A	A	A
Cape York	Midshelf	2016	0.781	B	B	B
Cape York	Midshelf	2016	0.855	A	A	A
Cape York	Midshelf	2016	0.929	A	A	A
Cape York	Midshelf	2016	0.903	A	A	A
Cape York	Offshore	2014	0.825	B	A	B
Cape York	Offshore	2014	0.874	A	A	A
Cape York	Offshore	2014	0.923	A	A	A
Cape York	Offshore	2014	0.911	A	A	A
Cape York	Offshore	2015	0.852	A	A	A
Cape York	Offshore	2015	0.896	A	A	A
Cape York	Offshore	2015	0.941	A	A	A
Cape York	Offshore	2015	0.928	A	A	A
Cape York	Offshore	2016	0.895	A	A	A
Cape York	Offshore	2016	0.927	A	A	A
Cape York	Offshore	2016	0.960	A	A	A
Cape York	Offshore	2016	0.951	A	A	A
Wet Tropics	Open Coastal	2014	0.602	C	B	C
Wet Tropics	Open Coastal	2014	0.711	B	B	B
Wet Tropics	Open Coastal	2014	0.819	B	A	B
Wet Tropics	Open Coastal	2014	0.742	B	B	B
Wet Tropics	Open Coastal	2015	0.668	B	B	B
Wet Tropics	Open Coastal	2015	0.760	B	B	B
Wet Tropics	Open Coastal	2015	0.853	A	A	A
Wet Tropics	Open Coastal	2015	0.803	B	A	B
Wet Tropics	Open Coastal	2016	0.692	B	B	B
Wet Tropics	Open Coastal	2016	0.790	B	B	B
Wet Tropics	Open Coastal	2016	0.888	A	A	A
Wet Tropics	Open Coastal	2016	0.836	A	A	B
Wet Tropics	Midshelf	2014	0.711	B	B	B
Wet Tropics	Midshelf	2014	0.789	B	B	B
Wet Tropics	Midshelf	2014	0.866	A	A	A
Wet Tropics	Midshelf	2014	0.826	B	A	B
Wet Tropics	Midshelf	2015	0.760	B	B	B
Wet Tropics	Midshelf	2015	0.826	B	A	B
Wet Tropics	Midshelf	2015	0.893	A	A	A
Wet Tropics	Midshelf	2015	0.866	A	A	A
Wet Tropics	Midshelf	2016	0.796	B	B	B
Wet Tropics	Midshelf	2016	0.864	A	A	A
Wet Tropics	Midshelf	2016	0.933	A	A	A
Wet Tropics	Midshelf	2016	0.905	A	A	A
Wet Tropics	Offshore	2014	0.819	B	A	B
Wet Tropics	Offshore	2014	0.868	A	A	A

..continued from previous page

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
Wet Tropics	Offshore	2014	0.918	A	A	A
Wet Tropics	Offshore	2014	0.906	A	A	A
Wet Tropics	Offshore	2015	0.844	A	A	B
Wet Tropics	Offshore	2015	0.886	A	A	A
Wet Tropics	Offshore	2015	0.929	A	A	A
Wet Tropics	Offshore	2015	0.923	A	A	A
Wet Tropics	Offshore	2016	0.873	A	A	A
Wet Tropics	Offshore	2016	0.910	A	A	A
Wet Tropics	Offshore	2016	0.947	A	A	A
Wet Tropics	Offshore	2016	0.940	A	A	A
Dry Tropics	Open Coastal	2014	0.580	C	C	C
Dry Tropics	Open Coastal	2014	0.686	B	B	B
Dry Tropics	Open Coastal	2014	0.793	B	B	B
Dry Tropics	Open Coastal	2014	0.772	B	B	B
Dry Tropics	Open Coastal	2015	0.624	C	B	C
Dry Tropics	Open Coastal	2015	0.726	B	B	B
Dry Tropics	Open Coastal	2015	0.829	B	A	B
Dry Tropics	Open Coastal	2015	0.813	B	A	B
Dry Tropics	Open Coastal	2016	0.639	C	B	C
Dry Tropics	Open Coastal	2016	0.746	B	B	B
Dry Tropics	Open Coastal	2016	0.852	A	A	A
Dry Tropics	Open Coastal	2016	0.827	B	A	B
Dry Tropics	Midshelf	2014	0.758	B	B	B
Dry Tropics	Midshelf	2014	0.806	B	A	B
Dry Tropics	Midshelf	2014	0.853	A	A	A
Dry Tropics	Midshelf	2014	0.817	B	A	B
Dry Tropics	Midshelf	2015	0.799	B	B	B
Dry Tropics	Midshelf	2015	0.841	A	A	B
Dry Tropics	Midshelf	2015	0.884	A	A	A
Dry Tropics	Midshelf	2015	0.870	A	A	A
Dry Tropics	Midshelf	2016	0.821	B	A	B
Dry Tropics	Midshelf	2016	0.863	A	A	A
Dry Tropics	Midshelf	2016	0.906	A	A	A
Dry Tropics	Midshelf	2016	0.896	A	A	A
Dry Tropics	Offshore	2014	0.778	B	B	B
Dry Tropics	Offshore	2014	0.842	A	A	B
Dry Tropics	Offshore	2014	0.907	A	A	A
Dry Tropics	Offshore	2014	0.885	A	A	A
Dry Tropics	Offshore	2015	0.809	B	A	B
Dry Tropics	Offshore	2015	0.867	A	A	A
Dry Tropics	Offshore	2015	0.924	A	A	A
Dry Tropics	Offshore	2015	0.905	A	A	A
Dry Tropics	Offshore	2016	0.829	B	A	B
Dry Tropics	Offshore	2016	0.885	A	A	A
Dry Tropics	Offshore	2016	0.940	A	A	A
Dry Tropics	Offshore	2016	0.921	A	A	A

..continued from previous page

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
Mackay Whitsunday	Open Coastal	2014	0.464	D	C	D
Mackay Whitsunday	Open Coastal	2014	0.559	C	C	C
Mackay Whitsunday	Open Coastal	2014	0.654	C	B	B
Mackay Whitsunday	Open Coastal	2014	0.686	B	B	B
Mackay Whitsunday	Open Coastal	2015	0.491	D	C	D
Mackay Whitsunday	Open Coastal	2015	0.586	C	C	C
Mackay Whitsunday	Open Coastal	2015	0.682	B	B	B
Mackay Whitsunday	Open Coastal	2015	0.709	B	B	B
Mackay Whitsunday	Open Coastal	2016	0.505	C	C	C
Mackay Whitsunday	Open Coastal	2016	0.600	C	B	C
Mackay Whitsunday	Open Coastal	2016	0.696	B	B	B
Mackay Whitsunday	Open Coastal	2016	0.720	B	B	B
Mackay Whitsunday	Midshelf	2014	0.568	C	C	C
Mackay Whitsunday	Midshelf	2014	0.633	C	B	C
Mackay Whitsunday	Midshelf	2014	0.699	B	B	B
Mackay Whitsunday	Midshelf	2014	0.687	B	B	B
Mackay Whitsunday	Midshelf	2015	0.607	C	B	C
Mackay Whitsunday	Midshelf	2015	0.670	B	B	B
Mackay Whitsunday	Midshelf	2015	0.733	B	B	B
Mackay Whitsunday	Midshelf	2015	0.722	B	B	B
Mackay Whitsunday	Midshelf	2016	0.602	C	B	C
Mackay Whitsunday	Midshelf	2016	0.666	C	B	B
Mackay Whitsunday	Midshelf	2016	0.729	B	B	B
Mackay Whitsunday	Midshelf	2016	0.734	B	B	B
Mackay Whitsunday	Offshore	2014	0.470	D	C	D
Mackay Whitsunday	Offshore	2014	0.590	C	C	C
Mackay Whitsunday	Offshore	2014	0.710	B	B	B
Mackay Whitsunday	Offshore	2014	0.689	B	B	B
Mackay Whitsunday	Offshore	2015	0.474	D	C	D
Mackay Whitsunday	Offshore	2015	0.596	C	C	C
Mackay Whitsunday	Offshore	2015	0.718	B	B	B
Mackay Whitsunday	Offshore	2015	0.702	B	B	B
Mackay Whitsunday	Offshore	2016	0.475	D	C	D
Mackay Whitsunday	Offshore	2016	0.596	C	C	C
Mackay Whitsunday	Offshore	2016	0.716	B	B	B
Mackay Whitsunday	Offshore	2016	0.693	B	B	B
Fitzroy	Open Coastal	2014	0.377	D	D	D
Fitzroy	Open Coastal	2014	0.484	D	C	D
Fitzroy	Open Coastal	2014	0.592	C	C	C
Fitzroy	Open Coastal	2014	0.559	C	C	C
Fitzroy	Open Coastal	2015	0.382	D	D	D
Fitzroy	Open Coastal	2015	0.491	D	C	D
Fitzroy	Open Coastal	2015	0.600	C	B	C
Fitzroy	Open Coastal	2015	0.608	C	B	C
Fitzroy	Open Coastal	2016	0.442	D	C	D
Fitzroy	Open Coastal	2016	0.542	C	C	C

..continued from previous page

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
Fitzroy	Open Coastal	2016	0.643	C	B	C
Fitzroy	Open Coastal	2016	0.671	B	B	B
Fitzroy	Midshelf	2014	0.589	C	C	C
Fitzroy	Midshelf	2014	0.646	C	B	C
Fitzroy	Midshelf	2014	0.703	B	B	B
Fitzroy	Midshelf	2014	0.570	C	C	C
Fitzroy	Midshelf	2015	0.595	C	C	C
Fitzroy	Midshelf	2015	0.654	C	B	B
Fitzroy	Midshelf	2015	0.713	B	B	B
Fitzroy	Midshelf	2015	0.627	C	B	C
Fitzroy	Midshelf	2016	0.631	C	B	C
Fitzroy	Midshelf	2016	0.681	B	B	B
Fitzroy	Midshelf	2016	0.731	B	B	B
Fitzroy	Midshelf	2016	0.683	B	B	B
Fitzroy	Offshore	2014	0.528	C	C	C
Fitzroy	Offshore	2014	0.638	C	B	C
Fitzroy	Offshore	2014	0.749	B	B	B
Fitzroy	Offshore	2014	0.699	B	B	B
Fitzroy	Offshore	2015	0.541	C	C	C
Fitzroy	Offshore	2015	0.651	C	B	B
Fitzroy	Offshore	2015	0.761	B	B	B
Fitzroy	Offshore	2015	0.727	B	B	B
Fitzroy	Offshore	2016	0.571	C	C	C
Fitzroy	Offshore	2016	0.675	B	B	B
Fitzroy	Offshore	2016	0.779	B	B	B
Fitzroy	Offshore	2016	0.739	B	B	B
Burnett Mary	Open Coastal	2014	0.624	C	B	C
Burnett Mary	Open Coastal	2014	0.702	B	B	B
Burnett Mary	Open Coastal	2014	0.780	B	B	B
Burnett Mary	Open Coastal	2014	0.779	B	B	B
Burnett Mary	Open Coastal	2015	0.639	C	B	C
Burnett Mary	Open Coastal	2015	0.721	B	B	B
Burnett Mary	Open Coastal	2015	0.803	B	A	B
Burnett Mary	Open Coastal	2015	0.797	B	B	B
Burnett Mary	Open Coastal	2016	0.653	C	B	B
Burnett Mary	Open Coastal	2016	0.735	B	B	B
Burnett Mary	Open Coastal	2016	0.817	B	A	B
Burnett Mary	Open Coastal	2016	0.805	B	A	B
Burnett Mary	Midshelf	2014	0.646	C	B	C
Burnett Mary	Midshelf	2014	0.690	B	B	B
Burnett Mary	Midshelf	2014	0.734	B	B	B
Burnett Mary	Midshelf	2014	0.749	B	B	B
Burnett Mary	Midshelf	2015	0.675	B	B	B
Burnett Mary	Midshelf	2015	0.717	B	B	B
Burnett Mary	Midshelf	2015	0.759	B	B	B
Burnett Mary	Midshelf	2015	0.785	B	B	B

..continued from previous page

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
Burnett Mary	Midshelf	2016	0.745	B	B	B
Burnett Mary	Midshelf	2016	0.780	B	B	B
Burnett Mary	Midshelf	2016	0.815	B	A	B
Burnett Mary	Midshelf	2016	0.835	A	A	B
Burnett Mary	Offshore	2014	0.733	B	B	B
Burnett Mary	Offshore	2014	0.802	B	A	B
Burnett Mary	Offshore	2014	0.871	A	A	A
Burnett Mary	Offshore	2014	0.842	A	A	B
Burnett Mary	Offshore	2015	0.771	B	B	B
Burnett Mary	Offshore	2015	0.831	B	A	B
Burnett Mary	Offshore	2015	0.891	A	A	A
Burnett Mary	Offshore	2015	0.872	A	A	A
Burnett Mary	Offshore	2016	0.848	A	A	B
Burnett Mary	Offshore	2016	0.893	A	A	A
Burnett Mary	Offshore	2016	0.938	A	A	A
Burnett Mary	Offshore	2016	0.912	A	A	A

6.1.7 Confidence intervals

Confidence intervals (CI) represent the intervals in which we have a certain degree of confidence (e.g. 95%) that repeated estimates will fall. Hence the 95% CI of the mean is the range defined by the quantiles representing 95% of repeated estimates of the mean.

To calculate 95% confidence intervals for bootstrap aggregated distributions (e.g. Site 1/Chemical distribution), we repeatedly[18] draw a single sample from each of the constituent distributions (e.g. a single value from the Site 1 Ammonia, Chlorophyll-a and NOx distributions) and from each set of draws, calculate the weighted [19] mean of the values. The 95% CI is thus calculated as the quantiles (p=0.025 and p=0.975) of the means.

Confidence intervals are used to represent uncertainty in estimations. For example, 95% confidence intervals associated with a estimated mean roughly express a range of values over which we have the nominated degree of confidence that the true value is likely to lie[20].

Uncertainty arrises from multiple sources. Firstly, it arrises from the accuracies of the measured data and secondly, from the inprecisions introduced by the statistical methodologies for processing and summarizing the dat. Hence encapsulating and communicating full uncertainty requires information about both of these sources of uncertainty.

Estimates (such as sample means) are typically calculated from very small (yet ideally representative) samples drawn from a much larger population. In such cases, the statistically derived confidence intervals are used to provide an indication of the range of estimates in which we are confident the true value is likely to lie. That is, they depict the statistical uncertainty that arrizes from the need to estimate parameters from small amounts of the total possible spatial/temporal domain.

If measurement uncertainty is also known, then it is possible to incorporate and propogate this through the aggregation schedule so as to yeild total uncertainty. Measurement uncertainty is very typically very difficult to obtain. Nevertheless, it is usually assumed to be relatively small compared to the statistical uncertainty.

However, in the case of the Satellite and eReefs data, we have a virtual saturation of sample data. That is, with respect to the spatial and temporal extent of the data, we essentially have the entire population. Consequently, the statistical uncertainty is virtually zero. We are not estimating a mean, we are calculating the mean. Hence

[18]The more repeated draws the closer the distribution of means will converge. For the current project, the number of repeated draws is 10,000.

[19]Weights according to the weights defined for that level of the aggregation hierarchy

[20]From a frequentist perspective, 95% confidence intervals techinically indicate that 95% of intervals of the calculated extent will contain the true mean

measurement uncertaintly is of elevated importance. Unfortunately, we do not have any information about the measurement uncertainty at a spatial and temporal scale appropriate. As a result, we have elected not to represent uncertainty (as it would only be based on statistical uncertainty which would give the misleading impression of extremely low levels of uncertainty).

6.2 Summary of adopted methodologies

The aggregation schedule can be summarized as:

A. Calculation of Zone level Score and Grades

1. Collect raw data (= **Measures**) at each fixed monitoring site and compare individual observations to associated threshold/benchmark/reference or set of expectation ranges
2. Create indexed data as an expression of degree of difference (*scaled modified amplitude method*) to yield a **Score** for each **Measure** per sampling location (e.g. Site) (applies to *Measures* in all *Indicators*, Water Quality). In the absence of thresholds (e.g. Measures within Plankton), observed data are rescaled to a range defined by historical quantiles (20th and 80th percentiles) for each Measure.
3. Apply any expert opinion interventions
4. Combine **Measure** Scores into **Site**-level **Sub-indicator** Scores by averaging taking into account any weightings, i.e. aggregate into observation-level Sub-indicator Scores. This step involves **Bootstrapping** each input to distributions of 10,000 re-samples (or fewer if weighted), combining distributions and finally Bootstrapping again into a single 10,000 size distribution.
5. Combine **Sub-indicator** Scores into **Site**-level **Indicator** Scores by averaging, i.e. aggregate into Site-level Indicator Scores.
6. Convert Scores into coloured **Grades** (A-E) for visual presentation in report card

B. Calculation of Zone level Grades

1. Aggregate **Site**-level *Indicator* Scores from step A.5 into **Zone**-level *Indicator* Scores by averaging (incorporating spatial weights)
2. Aggregate **Zone**-level *Indicator* Scores into **Zone**-level *Component* Scores by averaging (incorporating weights)

C. Calculation of Whole GBR Grades

1. Aggregate **Zone**-level *Indicator* Scores from step B.1 into **Whole GBR**-level *Indicator* Scores by averaging (incorporating spatial weights)
2. Aggregate **Whole GBR**-level *Indicator* Scores into **Whole GBR**-level *Component* Scores by averaging (incorporating weights)

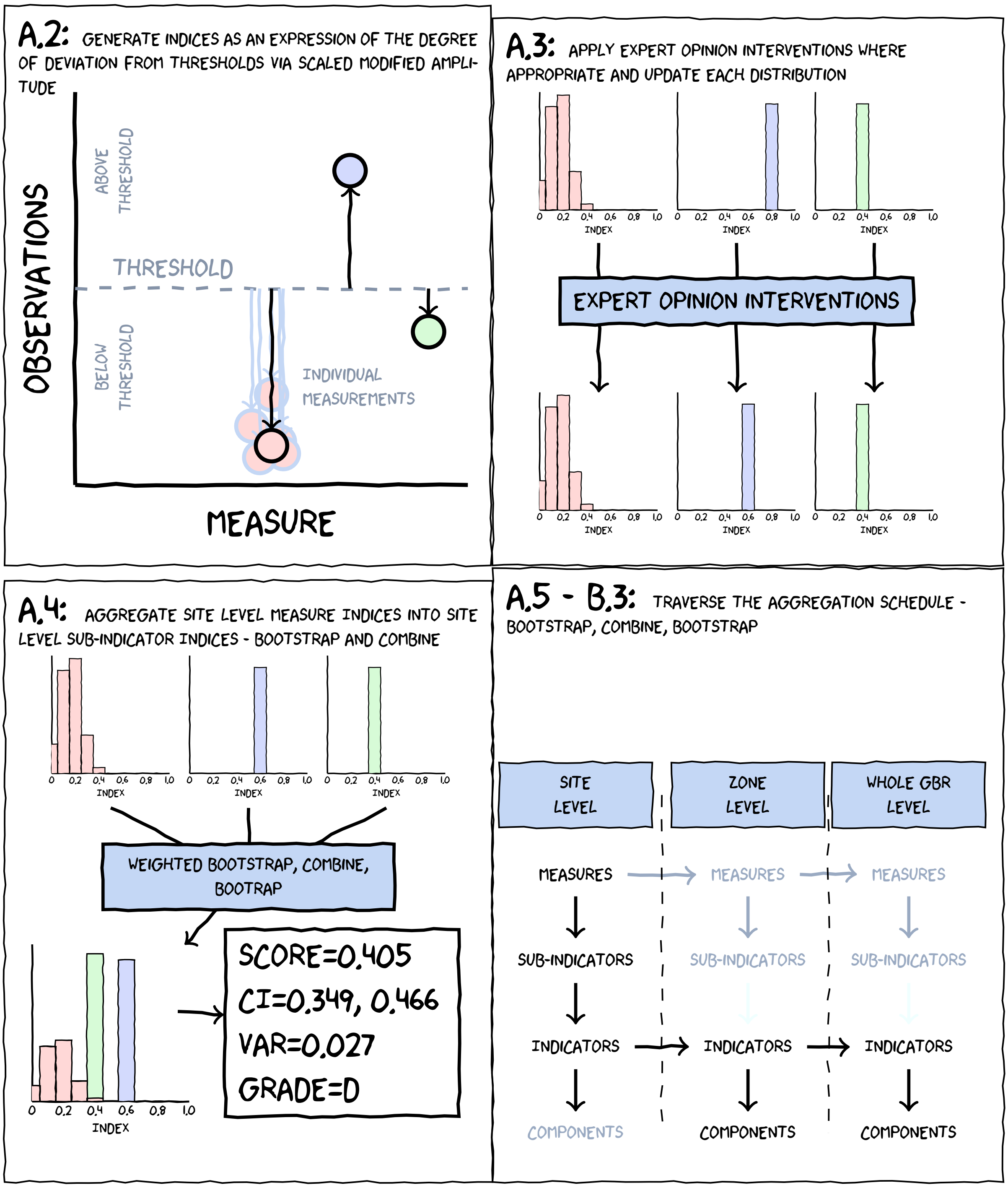
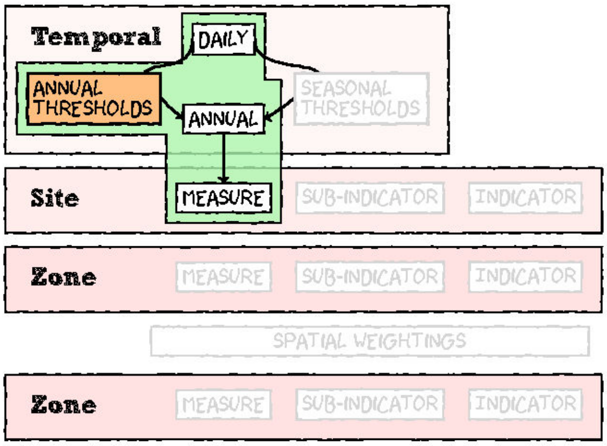


Figure 79: Schematic illustrating the major steps of the GBR Report Card. In this fabricated example, there are three Measures (Red, Green and Blue). Each of the Blue and Green Measures are represented by a single discrete observation, whereas the Red Measure is represented by a large collection of observations. Expert option intervened to lower the blue Measure distribution from observed values at 0.8 to 0.6.

6.3 Aggregation summaries

The ISP have indicated that the Water Quality metric should be based purely on eReefs fsMAMP indexed Chlorophyll-a and Secchi Depth and that the conversion of scores to grades should follow a uniform control chart. Consequently, this section will only present graphical summaries for these metric determinants. Other aggregation combinations can be found in Appendix ??.

6.3.1 Site/Measure level



6.3.1.1 Site level maps

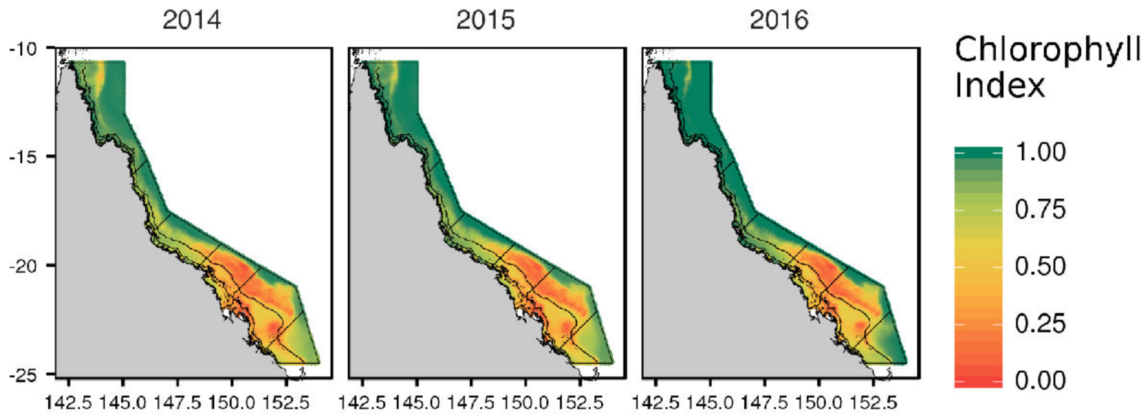


Figure 80: Spatio-temporal patterns in eReefs fsMAMP Chlorophyll-a index grades (Uniform grade type control chart applied).

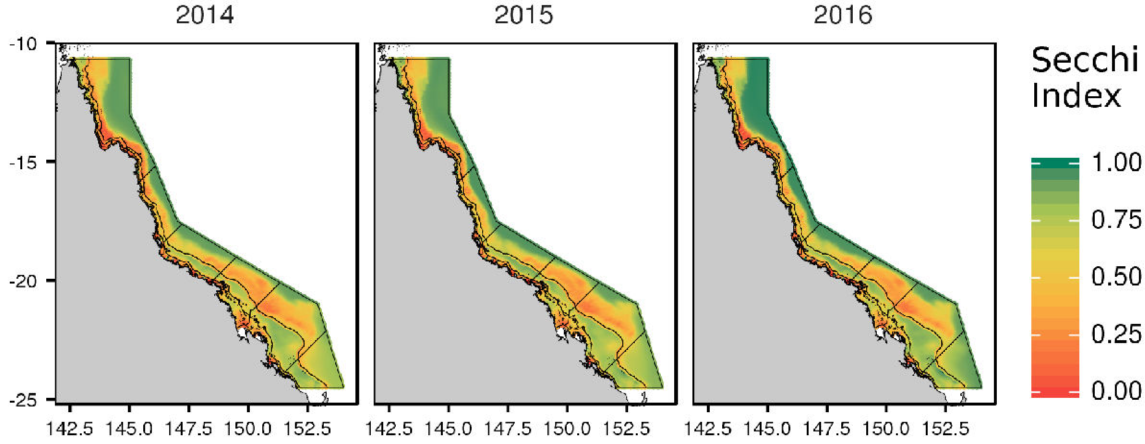
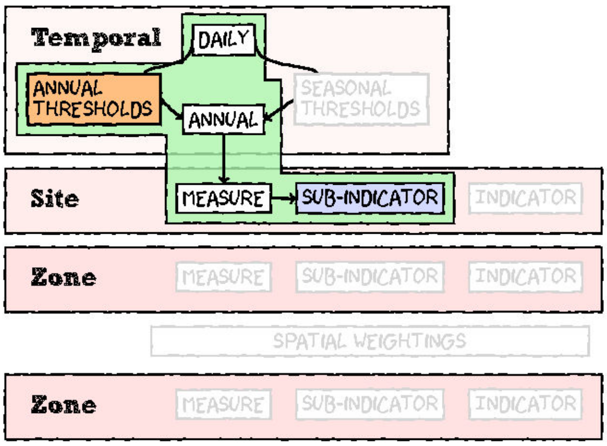


Figure 81: Spatio-temporal patterns in eReefs fsMAMP Secchi Depth index grades (Uniform grade type control chart applied).

6.3.2 Site/Subindicator level



6.3.2.1 Site level maps

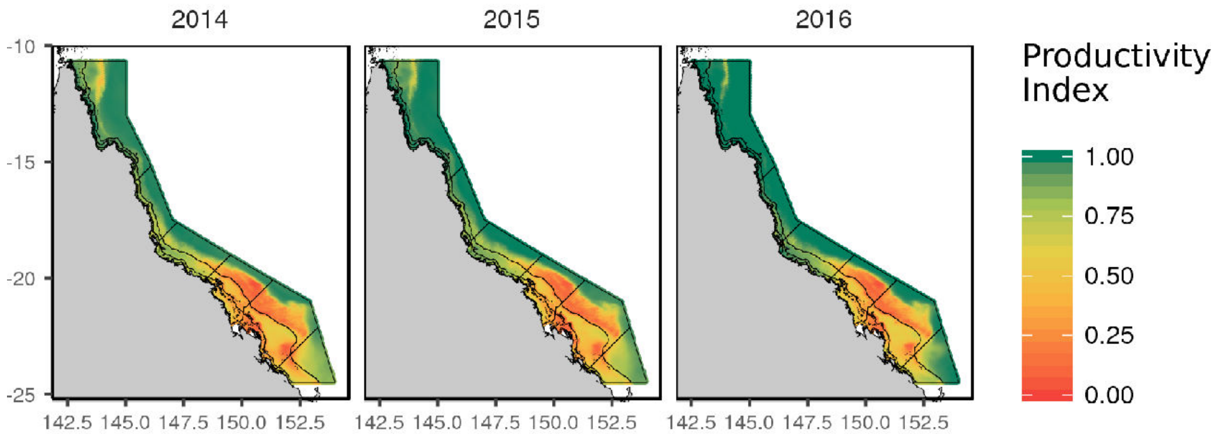


Figure 82: Spatio-temporal patterns in eReefs fsMAMP Productivity index grades (Uniform grade type control chart applied).

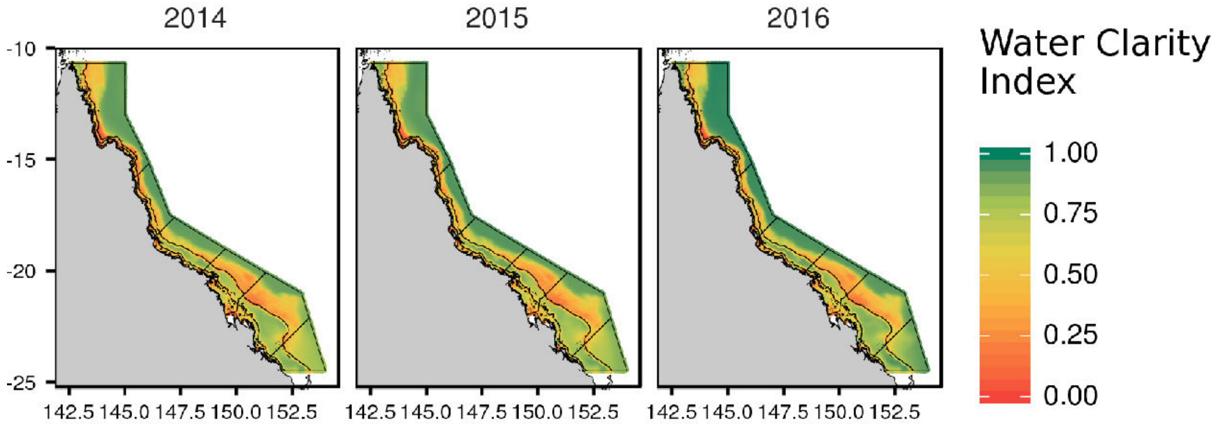
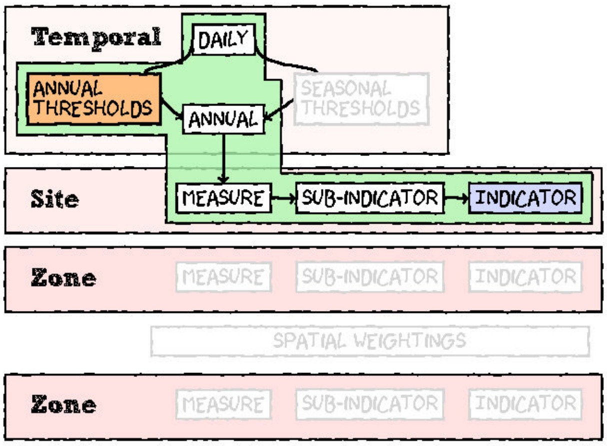


Figure 83: Spatio-temporal patterns in eReefs fsMAMP Water Clarity index grades (Uniform grade type control chart applied).

6.3.3 Site/Indicator level



6.3.3.1 Site level maps

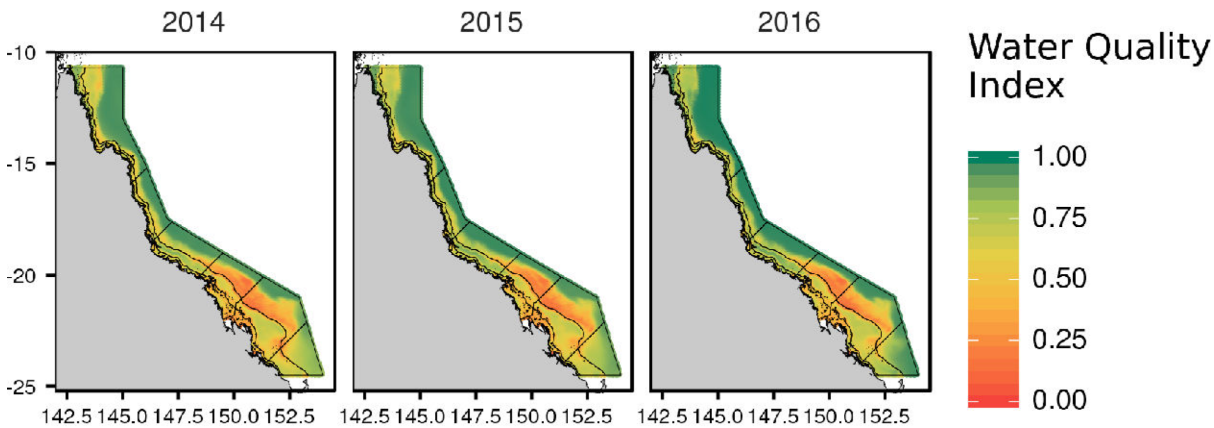
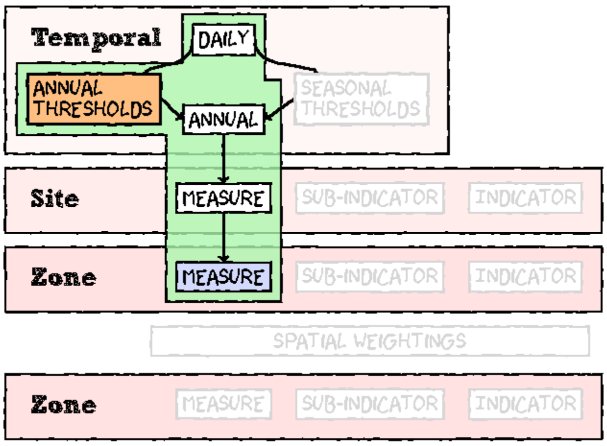


Figure 84: Spatio-temporal patterns in eReefs fsMAMP Water Quality index grades (Uniform grade type control chart applied).

6.3.4 Zone/Measure level



6.3.4.1 Simple time series

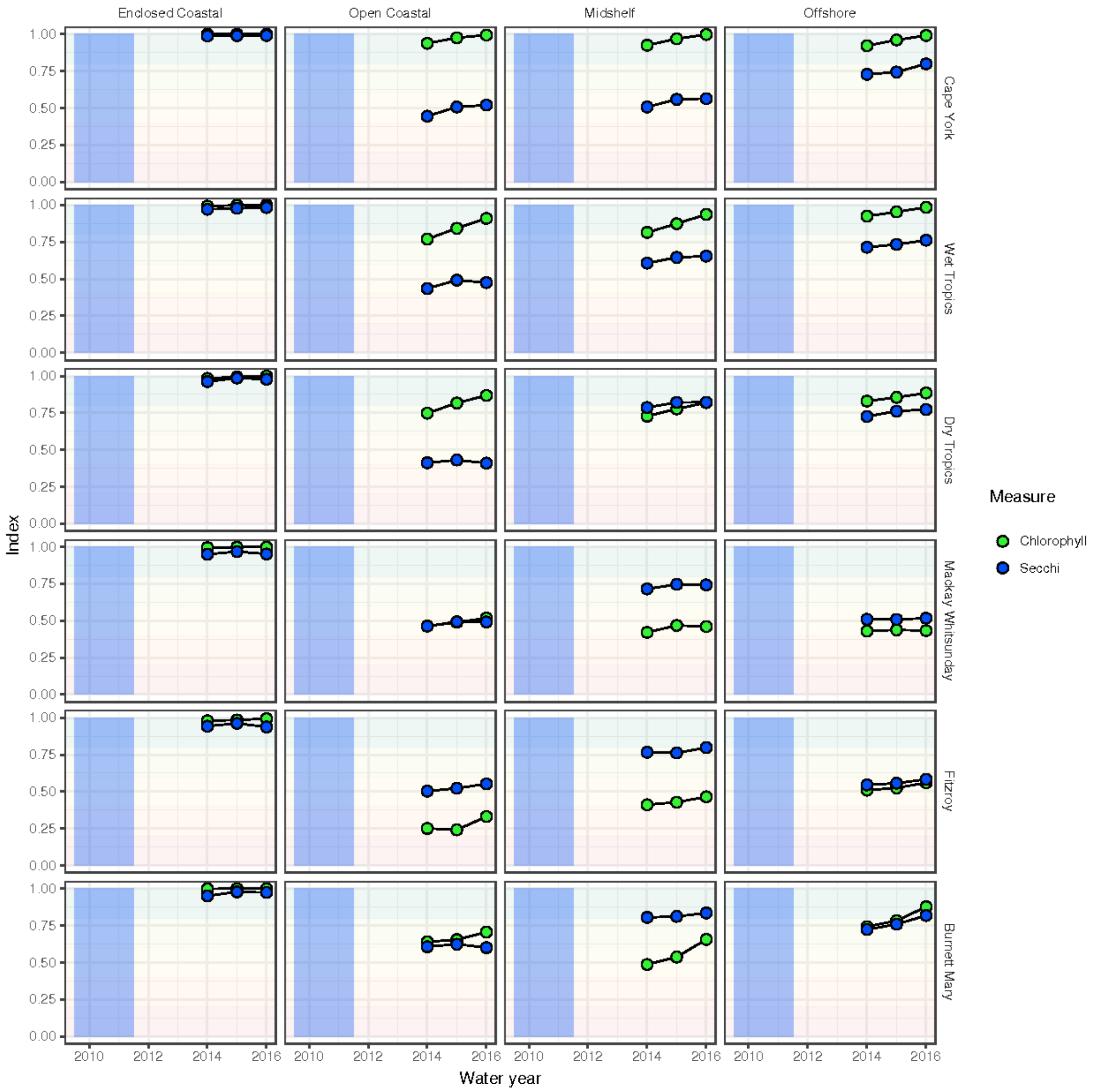


Figure 85: Time series of fsMAMP measures (Chlorophyll-a and Secchi Depth) index scores by zone. The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.3.4.2 *Flat map*

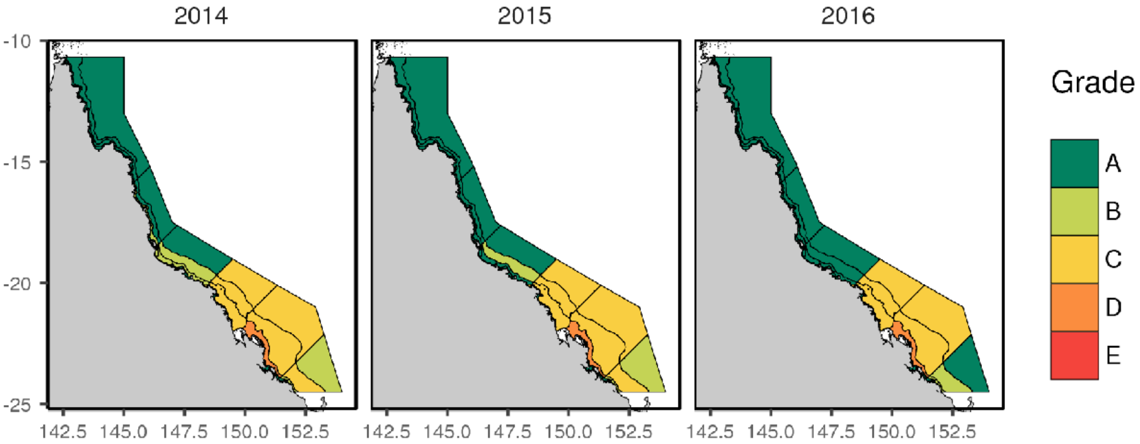


Figure 86: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Chlorophyll-a index grades (Uniform grade type control chart applied).

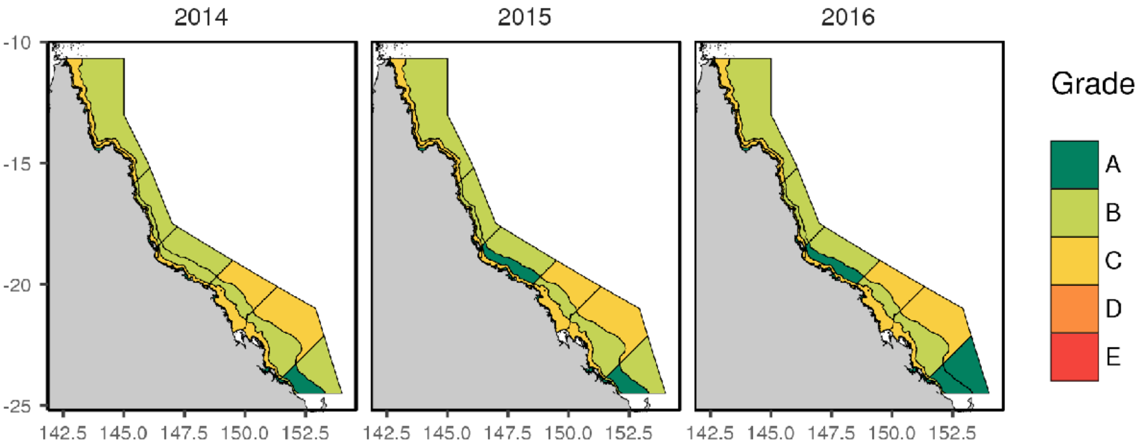


Figure 87: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Secchi Depth index grades (Uniform grade type control chart applied).

6.3.4.3 Mosaic plots level

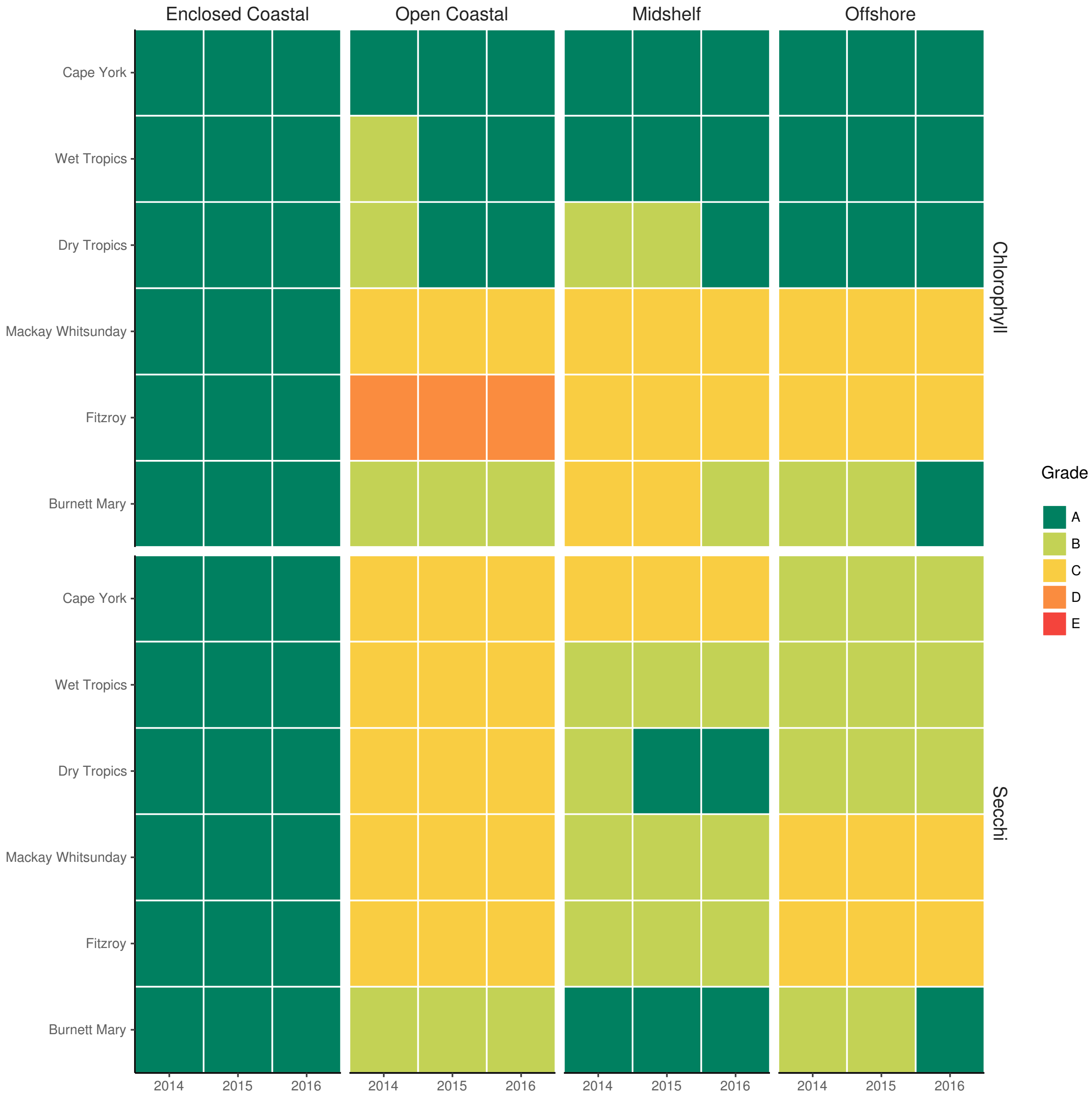
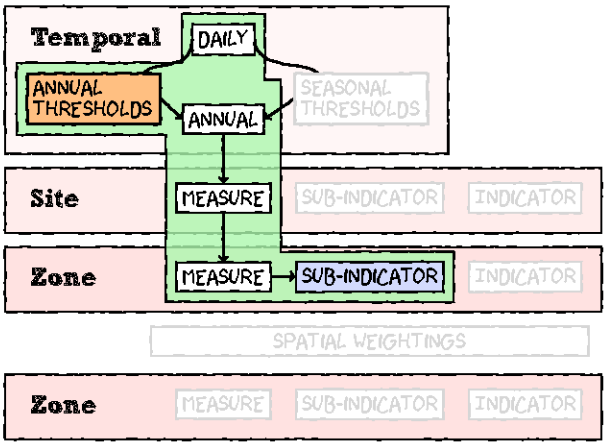


Figure 88: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Chlorophyll-a index grades (Uniform grade type control chart applied).

6.3.5 Zone/Subindicator



6.3.5.1 Simple time series

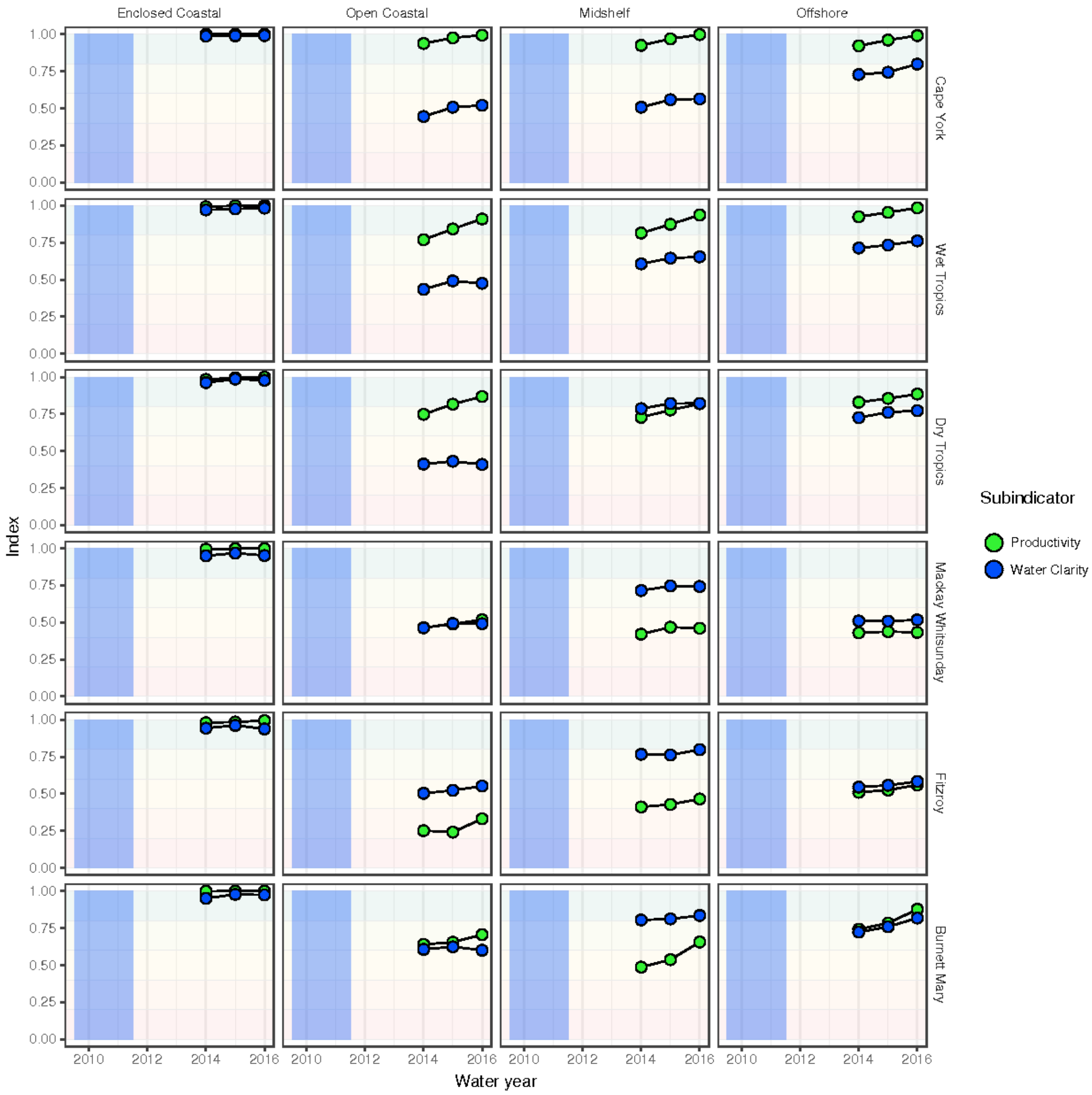


Figure 89: Time series of fsMAMP Productivity and Water Clarity index scores by zone. The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.3.5.2 *Flat map*

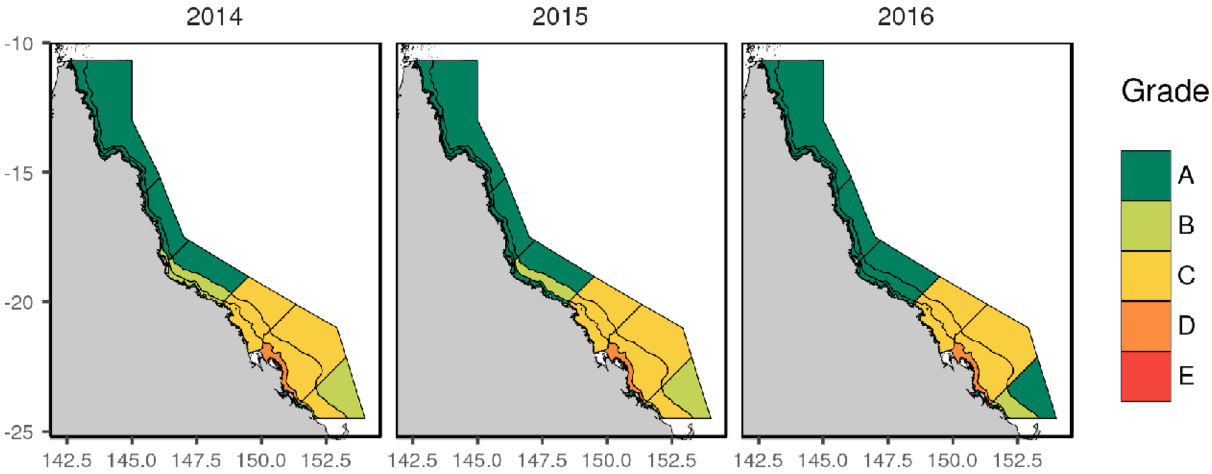


Figure 90: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Productivity index grades (Uniform grade type control chart applied).

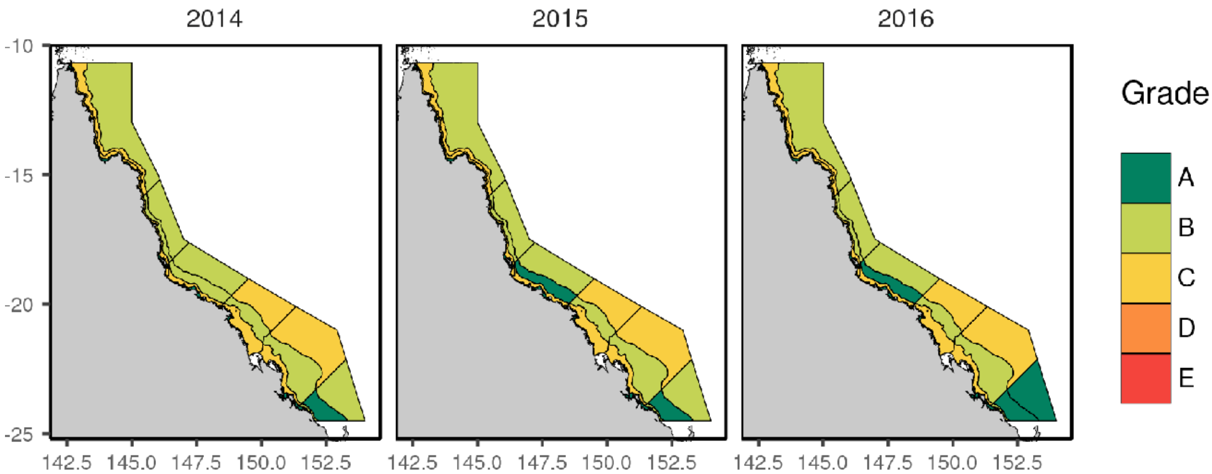


Figure 91: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Water Clarity index grades (Uniform grade type control chart applied).

6.3.5.3 *Mosaic plots*

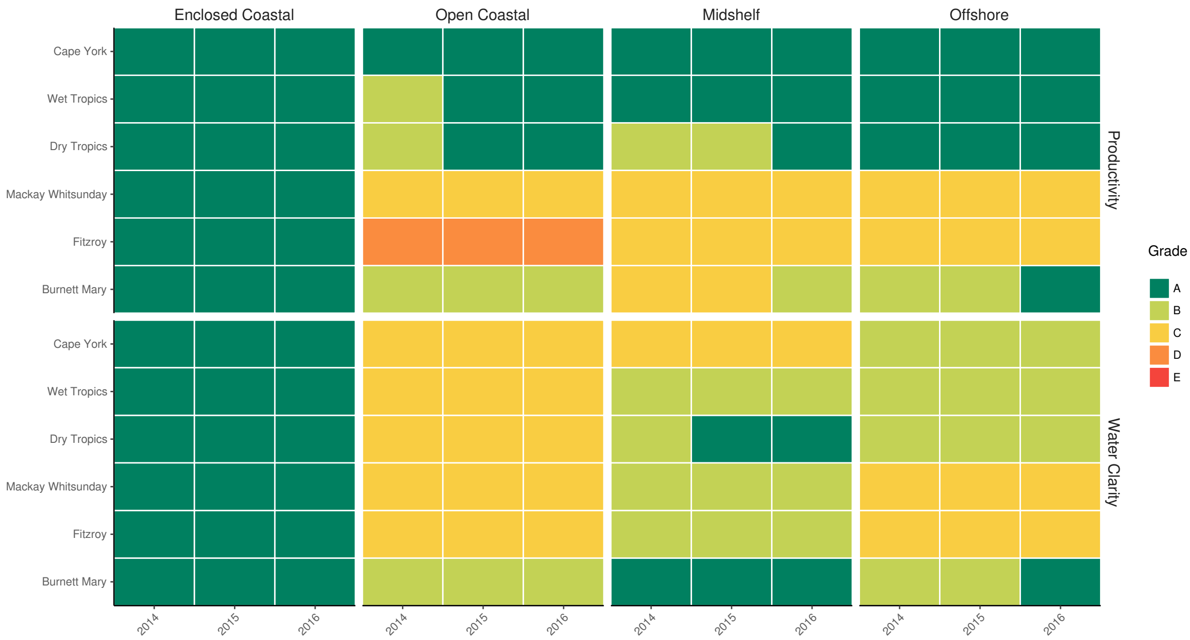
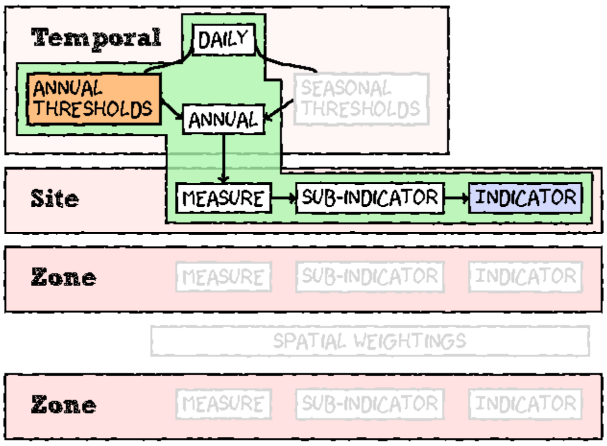


Figure 92: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Subindicator index grades (Uniform grade type control chart applied).

6.3.6 Zone/Indicator level



6.3.6.1 Simple time series

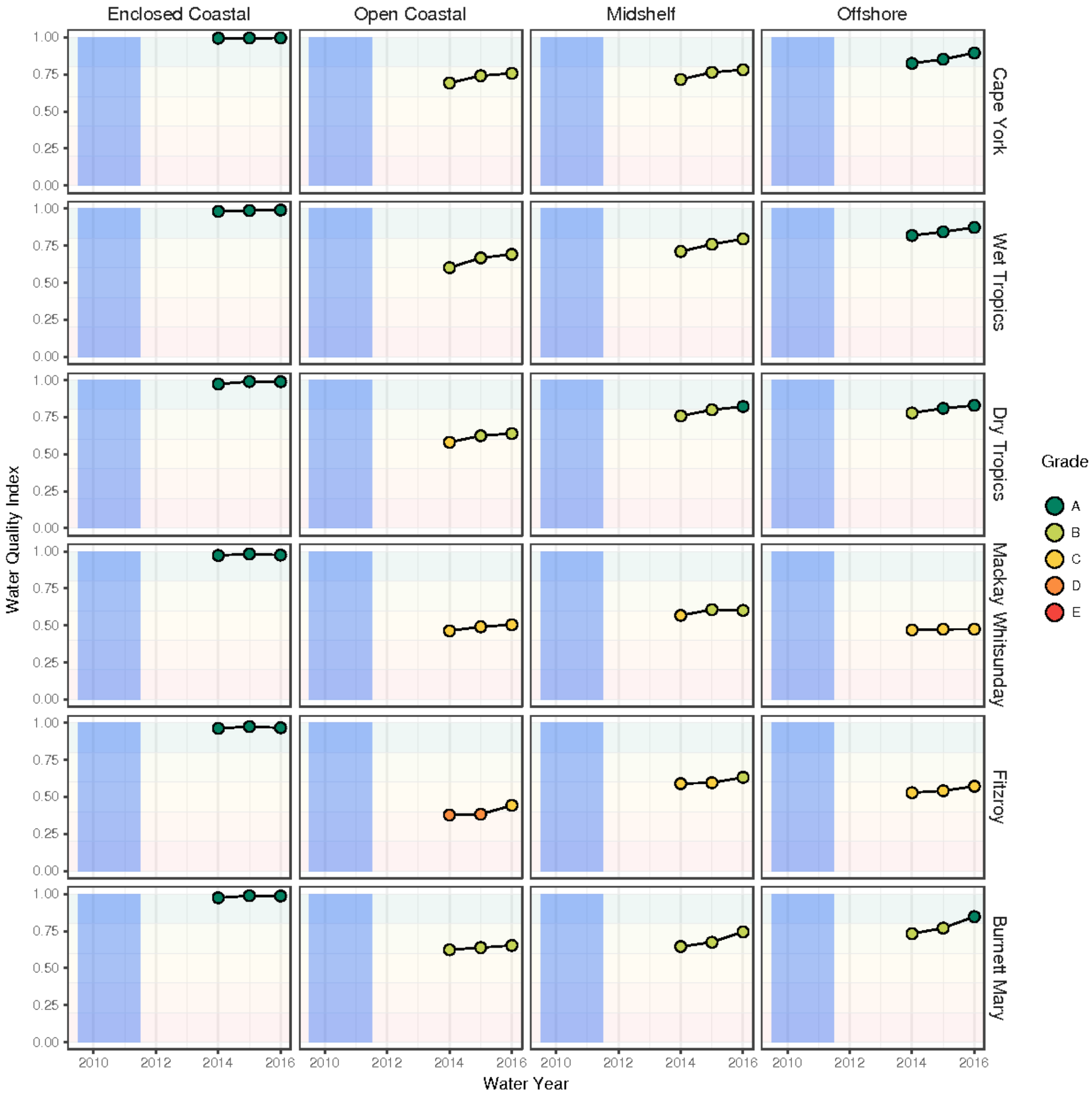


Figure 93: Time series of fsMAMP Water Quality index scores by zone. The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.3.6.2 *Flat map*

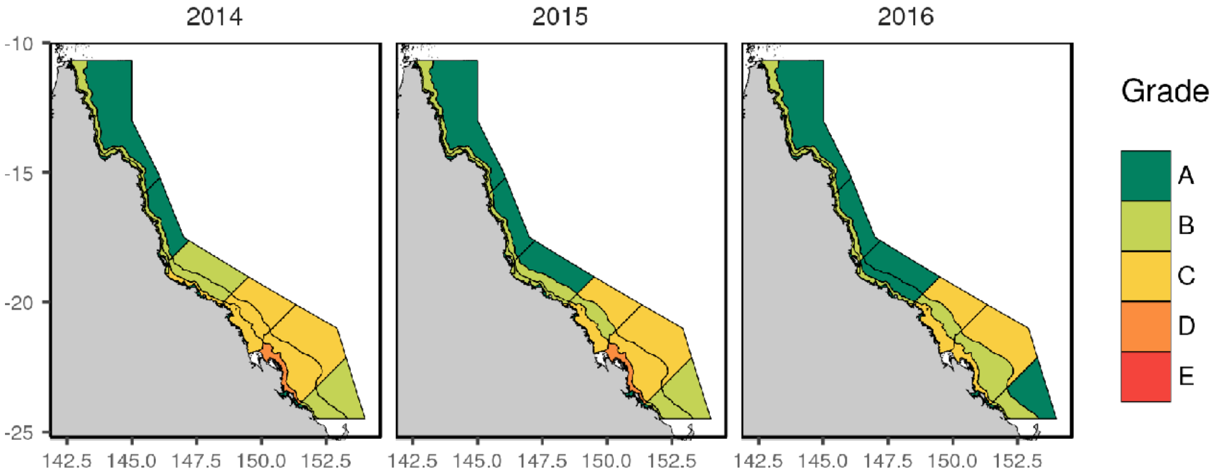


Figure 94: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Productivity index grades (Uniform grade type control chart applied).

6.3.6.3 *Mosaic plots*

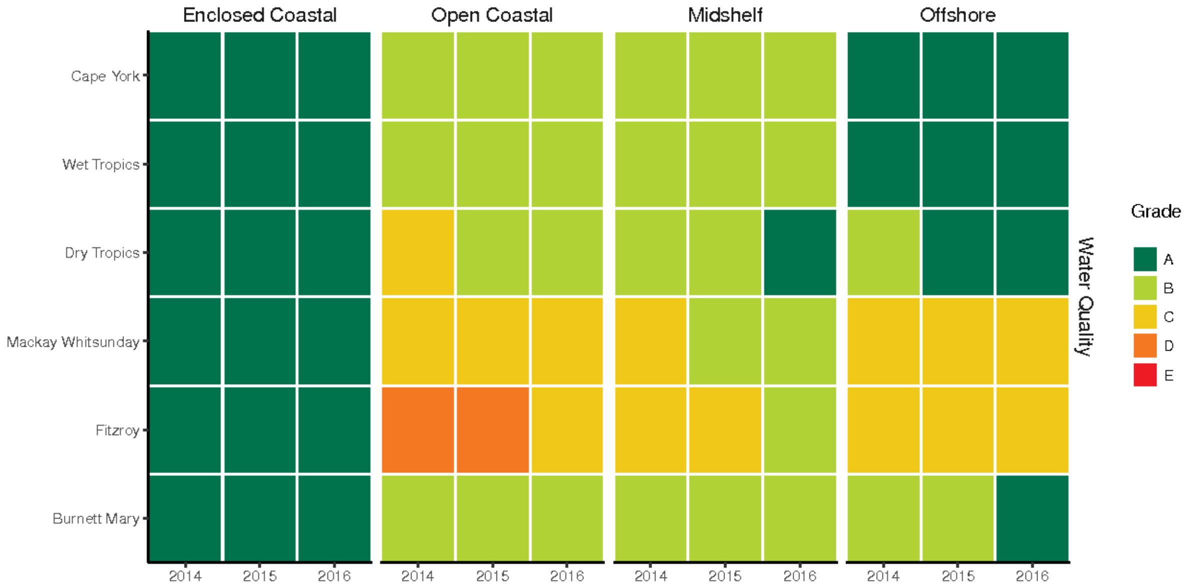
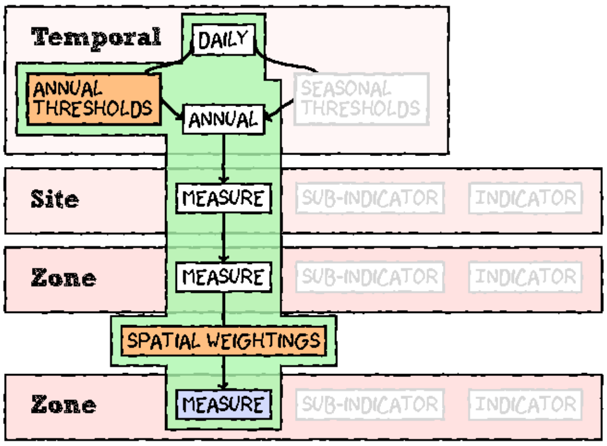


Figure 95: Simplified (Zone mean) eReefs spatio-temporal fsMAMP indicator index grades (Uniform grade type control chart applied).

6.4 Aggregations to water body level

6.4.1 Water body/Measure level



6.4.1.1 Simple time series

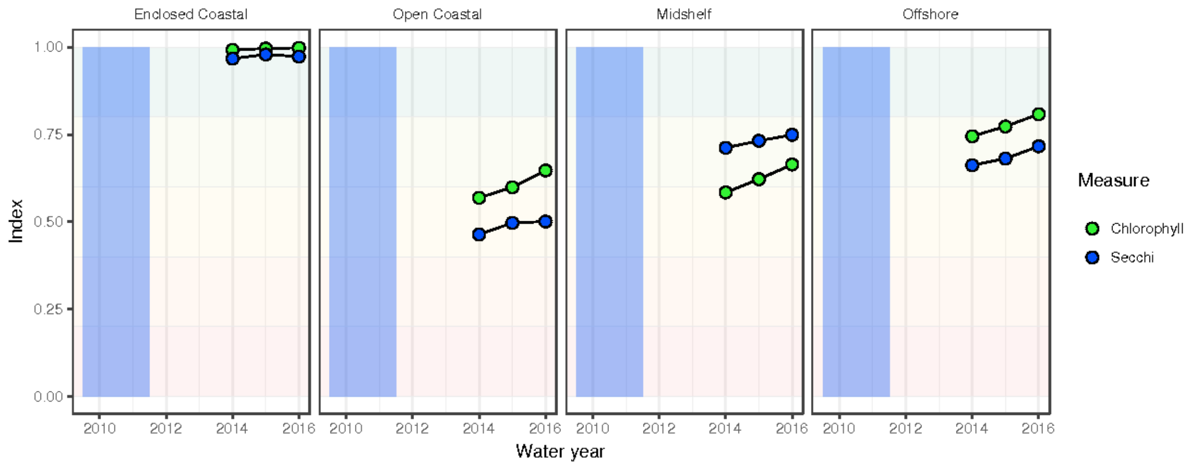


Figure 96: Time series of fsMAMP Measure index scores by water body (aggregated over management region weighted by area). The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.4.1.2 *Mosaic plots*

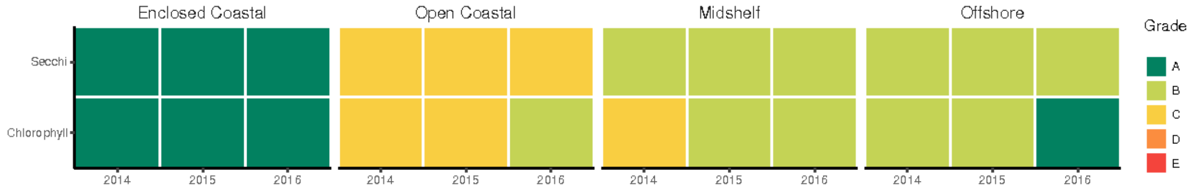
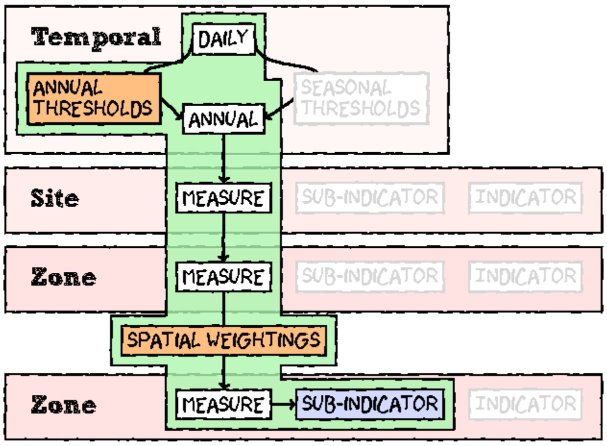


Figure 97: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Measurement index grades (Uniform grade type control chart applied).

6.4.2 Water body/Subindicator level



6.4.2.1 *Simple time series*

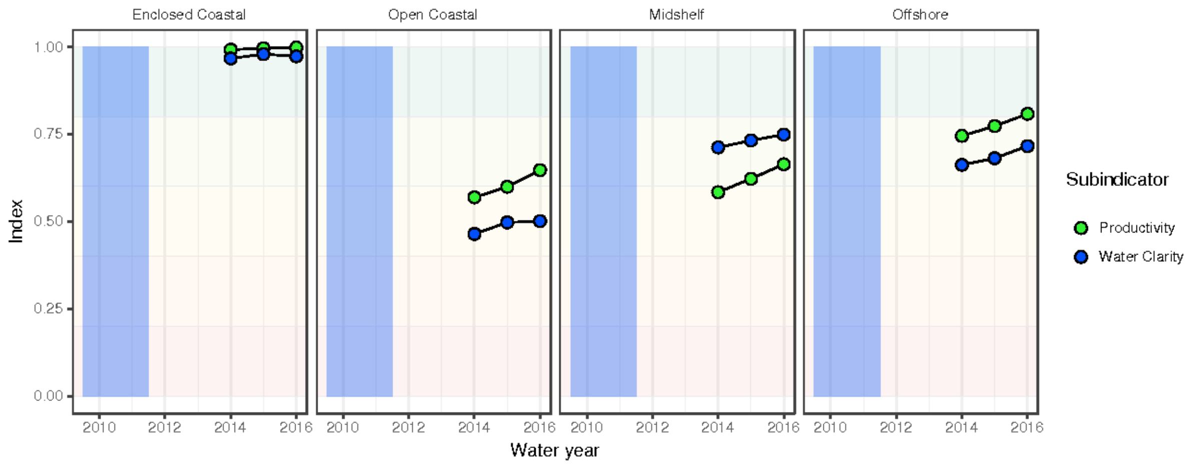


Figure 98: Time series of fsMAMP Subindicator index scores by water body (aggregated over management region weighted by area). The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.4.2.2 Mosaic plots

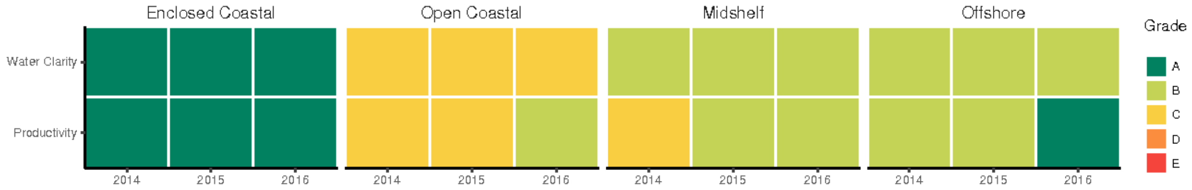
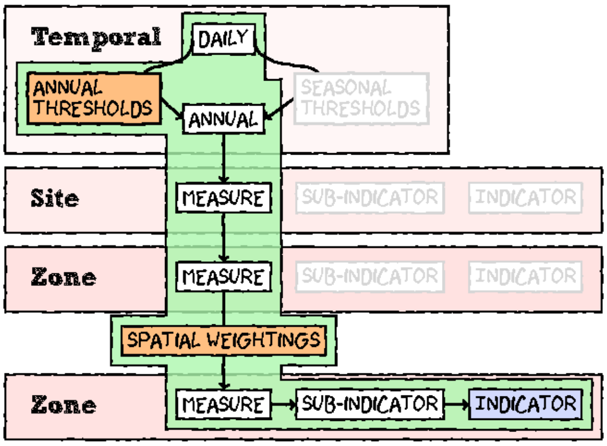


Figure 99: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Subindicator index grades (Uniform grade type control chart applied).

6.4.3 Water body/Indicator level



6.4.3.1 Simple time series

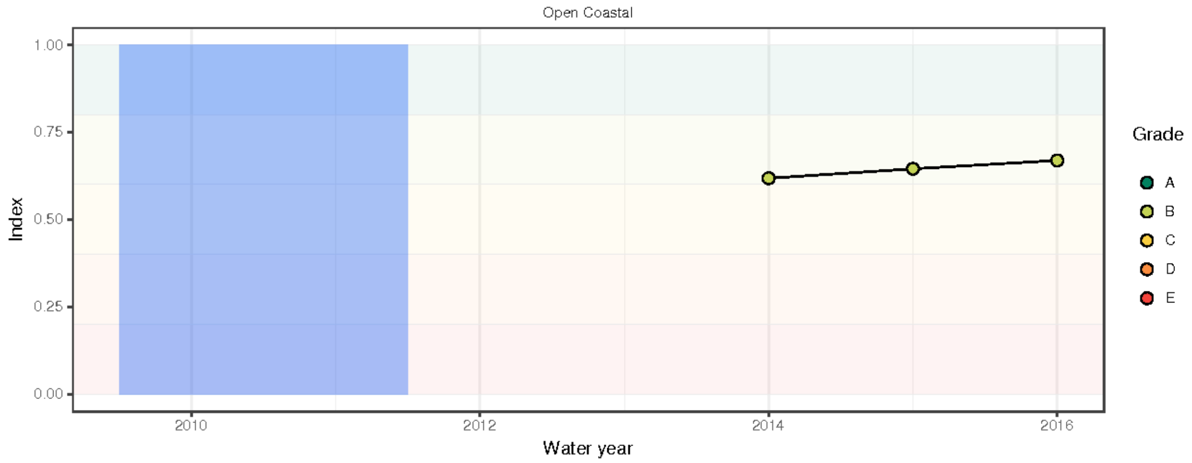


Figure 100: Time series of fsMAMP Indicator index scores by water body (aggregated over management region weighted by area). The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.4.3.2 Mosaic plots

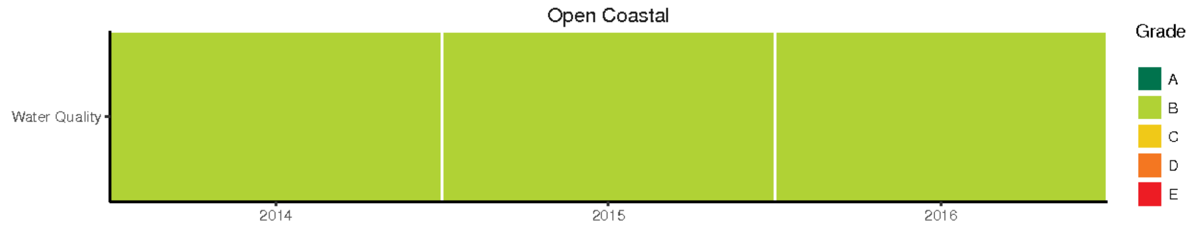
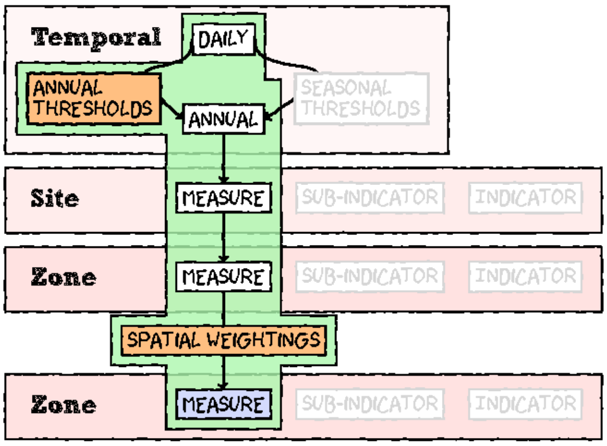


Figure 101: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Indicator index grades (Uniform grade type control chart applied).

6.5 Aggregations to GBR level

6.5.1 GBR/Measure level



6.5.1.1 Simple time series

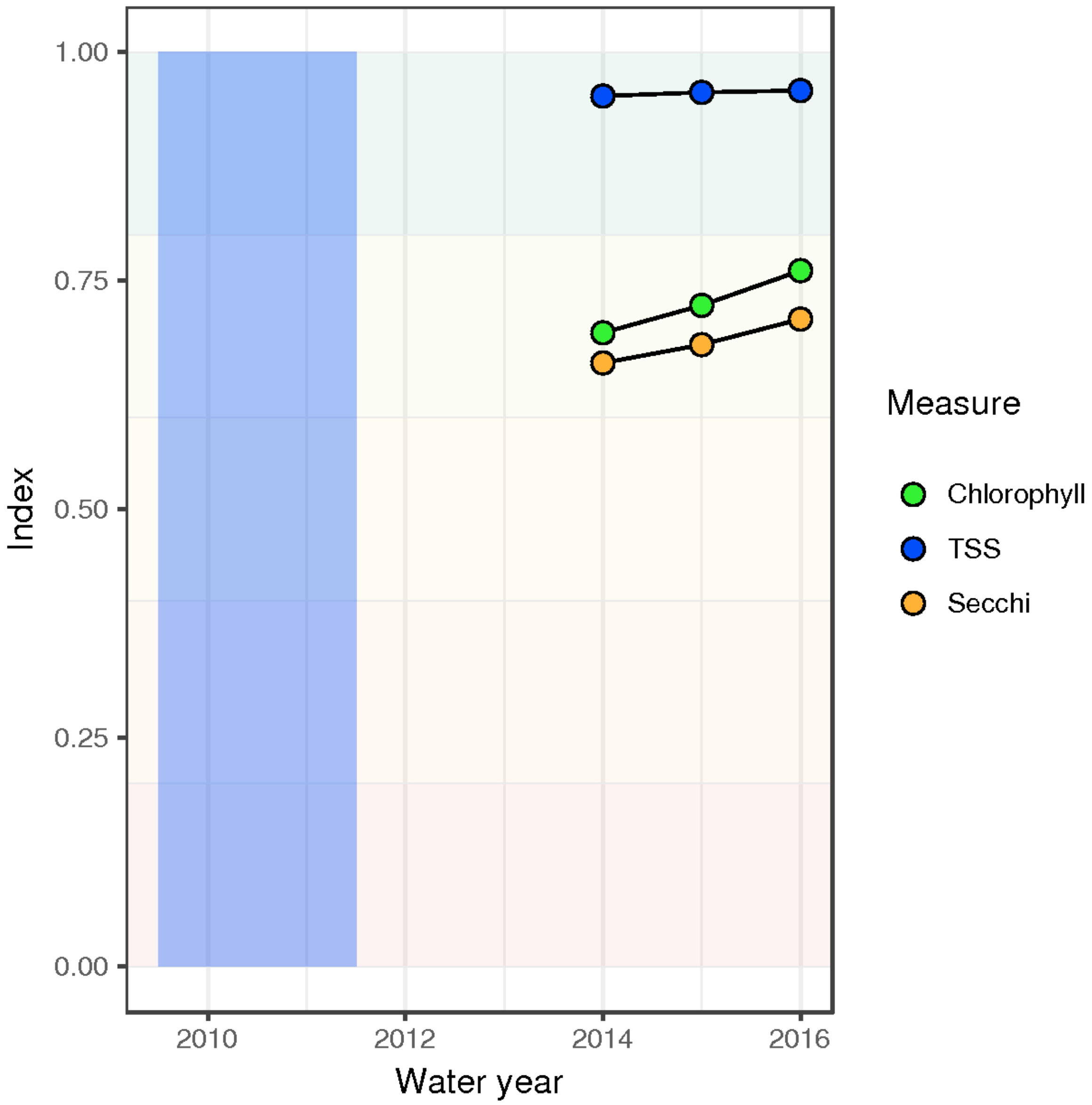


Figure 102: Time series of fsMAMP Measure index scores by GBR (aggregated over management region weighted by area). The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.5.1.2 Mosaic plots

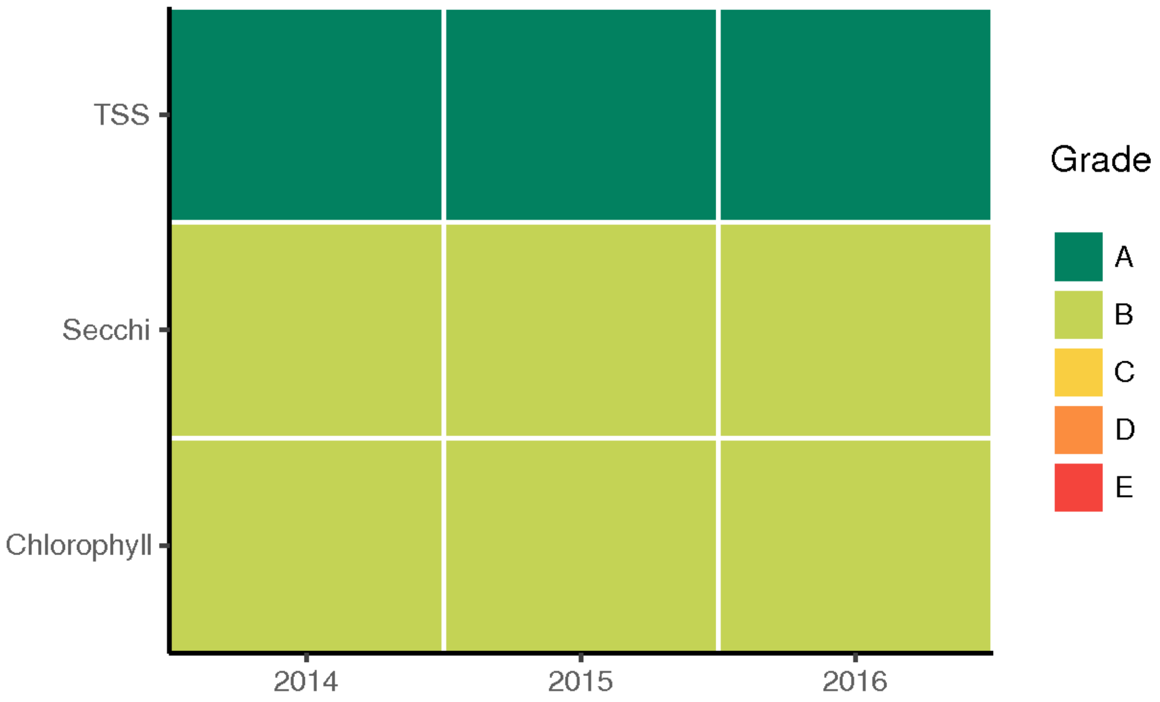
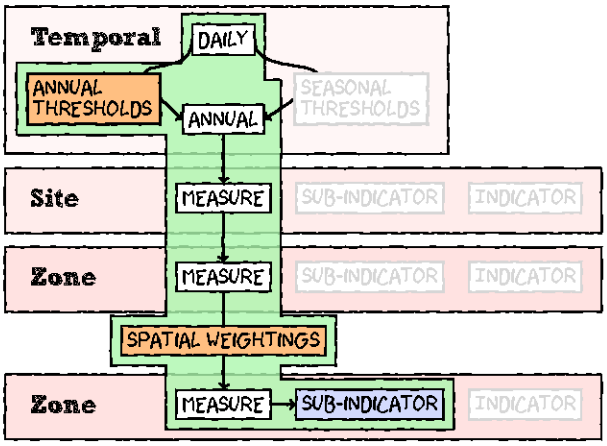


Figure 103: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Measurement index grades (Uniform grade type control chart applied).

6.5.2 GBR/Subindicator level



6.5.2.1 Simple time series

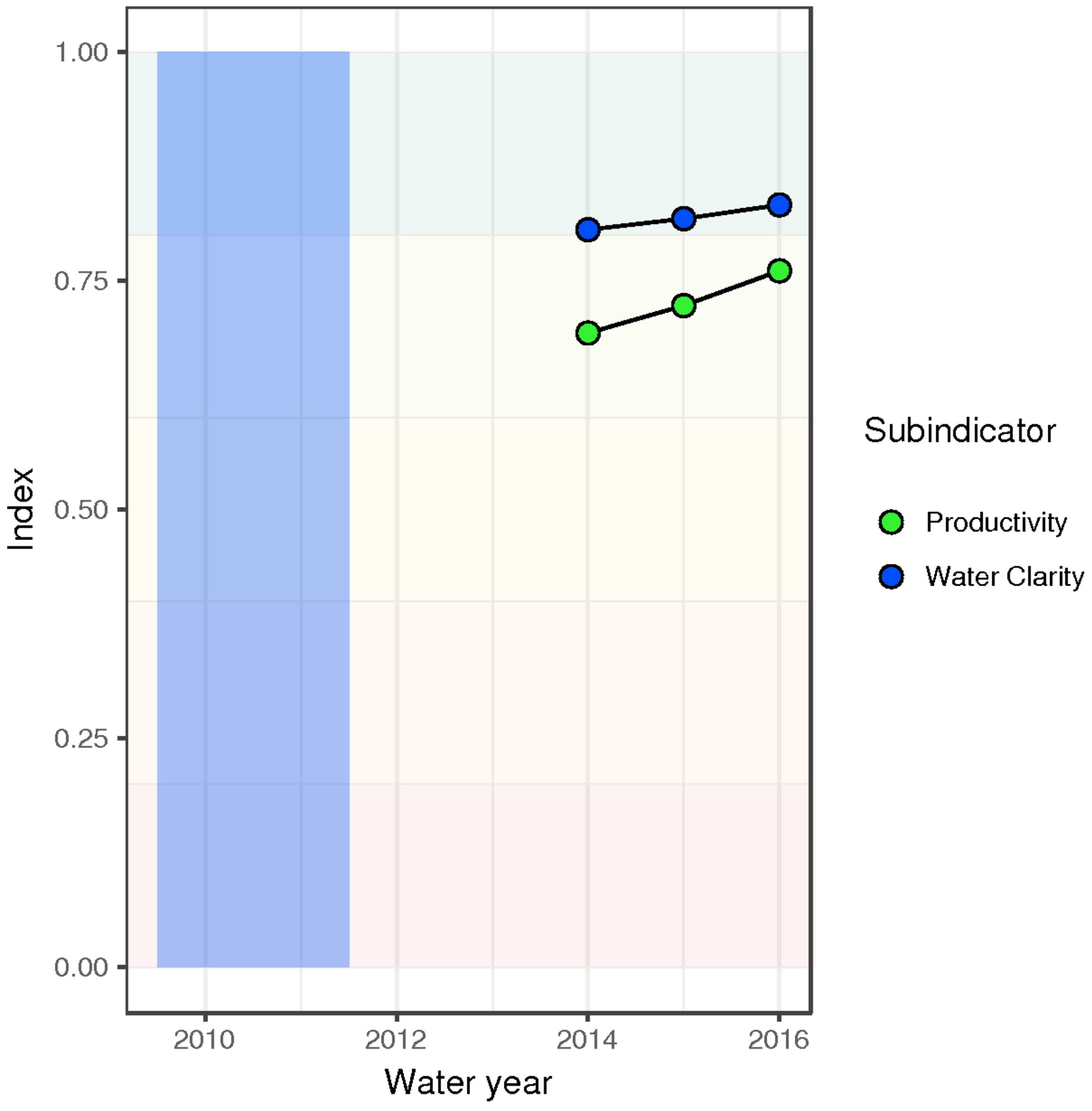


Figure 104: Time series of fsMAMP Subindicator index scores by GBR (aggregated over management region weighted by area). The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.5.2.2 *Mosaic plots*

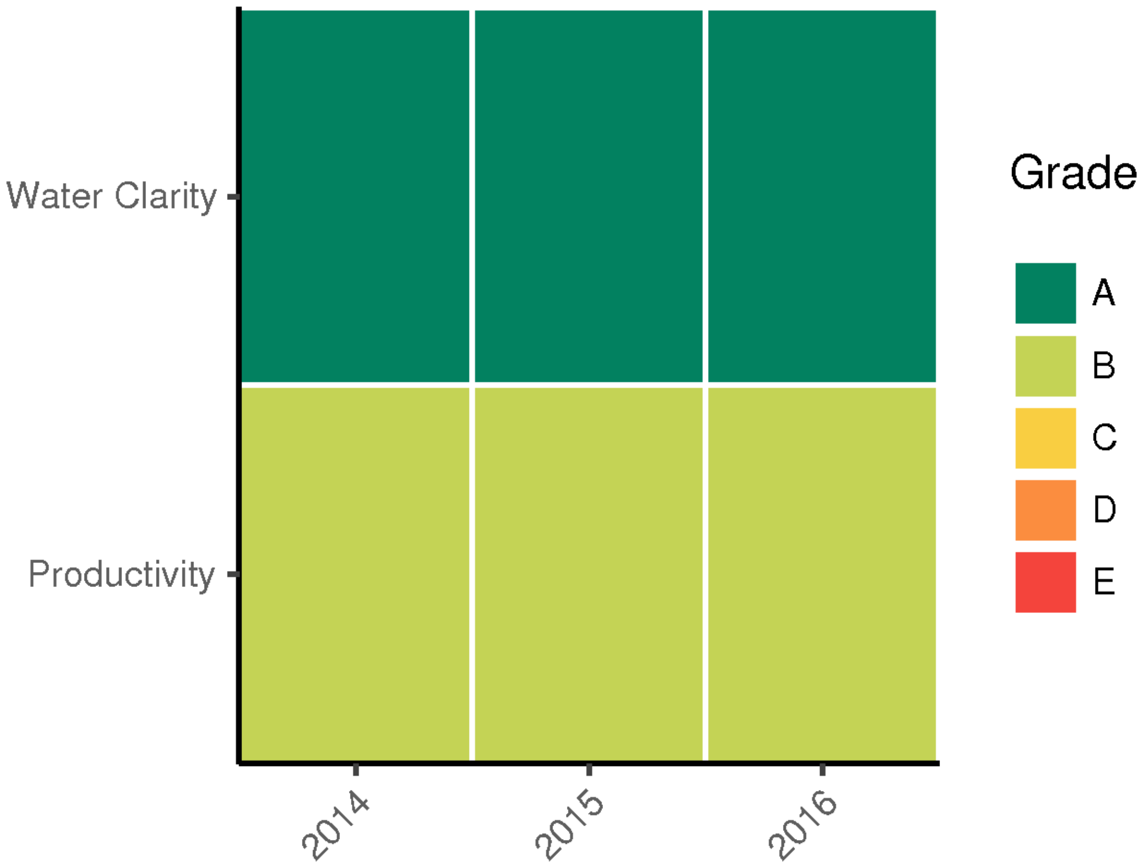
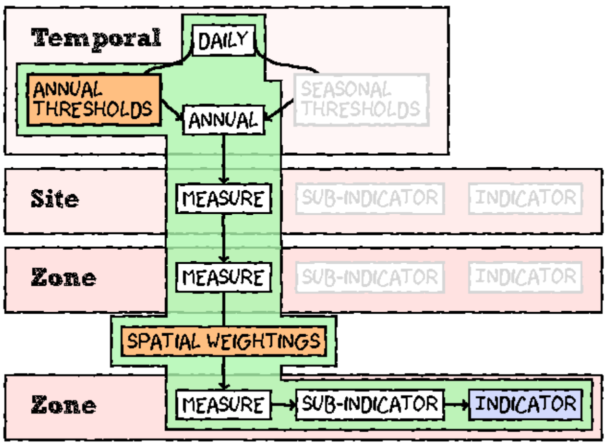


Figure 105: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Subindicator index grades (Uniform grade type control chart applied).

6.5.3 GBR/Indicator level



6.5.3.1 *Simple time series*

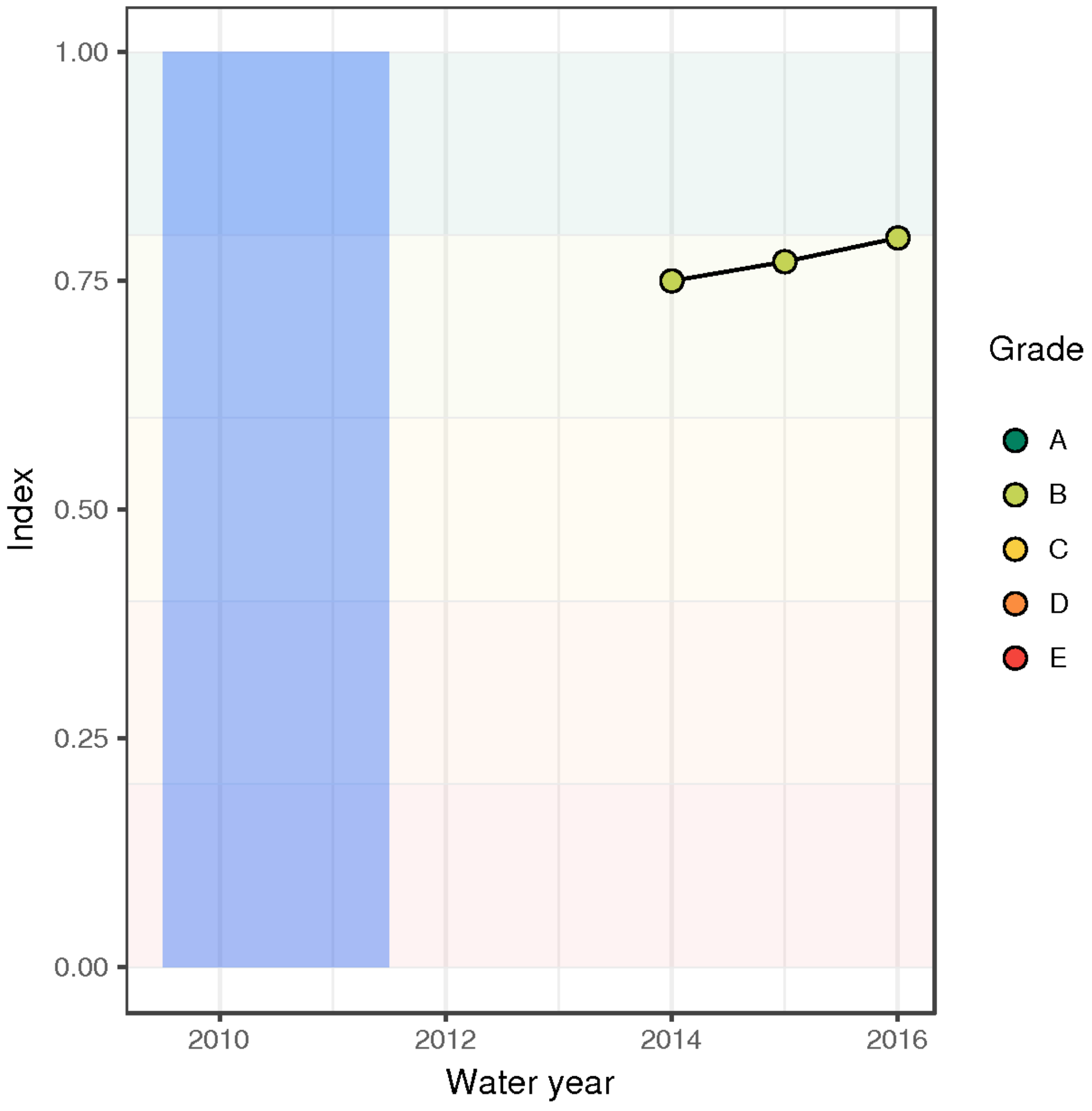


Figure 106: Time series of fsMAMP Indicator index scores by GBR (aggregated over management region weighted by area). The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.5.3.2 *Mosaic plots*

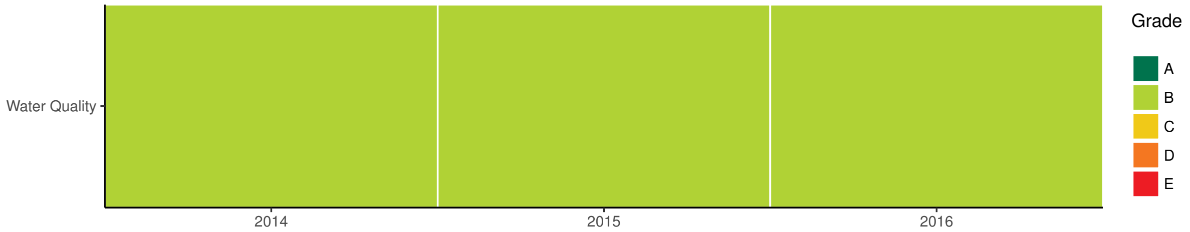


Figure 107: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Indicator index grades (Uniform grade type control chart applied).

Compare Measures aggregated to Zone level

- dependent on selection of sources etc..

6.6 Summary of recommendations