5. INDEX METRICS

5.1 Theoretical framework

Each individual indicator (or sub-indicator) addresses a different aspect of the state of an ecosystem. Hence, even a modest number of (sub)indicators will yield multiple perspectives on ecosystem health. Capturing the essence of the ecosystem health or an indicator thereof, necessitates integrating (aggregating) each of these perspectives together into a single *index*. There are numerous methods that have been applied to index aggregation, the most popular of which are itemized by Fox (2013) and described and evaluated in the context of water quality indices by either Walsh and Wheeler (2012) (from the perspective of cost benefit analyses) or Whittaker et al. (2012).

5.1.1 Multivariate health indicators

Motivated by the need to integrate multiple disparately scaled ecological variables together in the absence of any normalizing information (such as benchmarks, guidelines or thresholds, see Section 5.1.2), a variety of predominantly multivariate analyses have been used in the generation of ecosystem health indices. However, Whittaker et al. (2012) cautioned that since the incorporated weights are all exclusively informed by the statistical properties of the constituent indicator data, if these statistical properties did not coincide with expert knowledge of the relative importance of the indicators, then the resulting indices are likely to be poor.

As an alternative, Whittaker et al. (2012) suggest the Malmquist index. The computational details of the Marlmquist index are rather complex and since this method does not appear to have been adopted by any report cards, we will restrict our description to just a brief overview. Whittaker et al. (2012)'s proposed version of the Malmquist index calculates pairwise ratios of indicator distances from a multivariate benchmark curve. The benchmark curve (a form of indifference curve), is a multivariate curve defined by the lower boundary of a convex hull of all indicator values and is thus derived entirely from the observed data. Using simulated data with manufactured statistical complications (heterogeneity and temporal autocorrelation), Whittaker et al. (2012) demonstrated that the Malmquist index out performs indices based on principal components analysis and suggested other statistical methods would have similar shortcomings.

5.1.2 Thresholds

The absolute value of an indicator is rarely a meaningful assessment of ecosystem health assessments. Nor are the statistical properties of a time series necessarily a good basis for normalizing indicators or representing the objectives. What constitutes a 'good' or 'poor' level is likely to vary according to indicator, the ecosystem (e.g. freshwater, estuarine or marine) as well as the geographical and temporal (e.g. pre-industrial or current, seasonal) context. Another way to normalize the location (center) of indicators (if not the scale as well) that incorporates both knowledge about the ecological basis of the indicator and the objectives that they address is to express the indicators relative to *benchmarks*.

Benchmarks are typically either reference or baseline conditions (sites or historic data representing relatively low disturbance 'healthy' conditions), threshold values (ecotoxicology tolerances representing the cusp of 'unhealthy conditions) or guideline values (derived from either historical quantiles or ecotoxicology). Thresholds and guideline values are typically peer reviewed and ecologically meaningful, yet their specificity varies from local to regional, national or international standards.

Whilst a 'distance to benchmark' approach does provides some level of standardization (Connolly et al., 2013), to be useful, not only should there be some form of homogenization in what the benchmark condition represents, the polarity of the distance should be well understood (Hijuelos and Reed, 2013) and the magnitude of the distance should be commensurate with position along a disturbance gradient. That is, there should be some consistency in what it means to be above or below a benchmark, and indeed what it means to be a certain distance from a benchmark. Ideally, benchmarks should also be locally relevant (Connolly et al., 2013) and consider seasonal variability (Coates et al., 2007; Hallett et al., 2012). Indeed, in a review of the methodologies used to set benchmarks, (Borja et al., 2012) demonstrated the importance of setting appropriate benchmarks from which to assess ecosystem quality by directly linking the inability of indices to detect impacts in ecosystems to inappropriate reference conditions.

It is also important that benchmarks align with objectives in order to ensure indicators are appropriate. For example, if an objective is to maintain sustainable stocks of a particular species of fish, a benchmarks that reflect either historical numbers or the numbers present at low pressure sites do not necessarily represent the level of sustainability.

Ecological monitors have long recognized the need to express ecosystem ratings as standardized scores and in terms that are more accessible to policy makers and the general public. Whilst initial applications focused on normalizing observed measures against subjective rating curves to yield dimensionless index values on the scale of [0,1] that could be readiby combined into a single understandable score or rating (e.g. Miller et al., 1986), more recent studies have explored formulations that compare observed measures to baseline, reference, objectives or guideline values (collectively, benchmarks) values (e.g. CCME, 2001; Hurley et al., 2012; Jones et al., 2013).

Connolly et al. (2013) reviewed the use of report cards for monitoring ecosystem health and tabulated the general properties of a range of methods employed across a many different monitoring programs. Rather than duplicate that information here, the current intention is to provide more specific details about the algorithms used across those programs.

5.1.3 Unifying indices

The Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI; CCME, 2001) incorporates comparisons to baseline based on *scope* (proportion of indicators that have one or more failures to meet objectives), *frequency* (proportion of all comparisons failing to meet objectives) and *amplitude* (the normalized degree to which failed comparisons exceed objectives).

$$F_1 = 100.\left(\frac{Number\ of\ failed\ indicators}{Total\ number\ of\ indicators}\right)$$

$$F_2 = 100.\left(\frac{Number\ of\ failed\ comparisons}{Total\ number\ of\ comparisons}\right)$$

$$F_3 = \frac{100.E}{1+E}; \qquad E = \frac{\sum_{i=1}^{n} e_i}{n}; \qquad e_i = z_i.\left[\left(\frac{x_i}{benchmark_i}\right)^{\lambda_i} - 1\right]$$

$$z_i = \begin{cases} 1 & \text{if ith comparison fails} \\ 0 & \text{otherwise} \end{cases}; \qquad \lambda_1 = \begin{cases} 1 & \text{If} < benchmark_i = \text{fail} \\ -1 & \text{If} > benchmark_i = \text{fail} \end{cases}$$

$$CCMEWQI = 100 - \left(\frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732}\right)$$

where n is the number of comparisons.

Whilst the CCME WQI might serve its purpose in the context to which it is applied, it is unlikely to be a useful metric for any indices involving remote sensing data or indeed any situation with a reasonable large amount of data or indicators. One-third of the weighting of the metric is calculated on the proportion of indicators that failed. The more observations are collected, the more likely at least one of them will exceed the benchmark. Hence, this one-third will quickly approach a constant of 1 thereby reducing overall sensitivity. In addition, the one-third of the method that weighting on amplitude only does so with respect to failure - there is no degree of how well the data recedes the benchmark. Finally, unifying indices have very limited scope for propagating any uncertainty. Consequently, this metric of index computation will not be explored in this project.

Rather than calculate the proportion of all comparisons failing to meet objectives across all indicators (as in the *frequency* component of the CCME WQI), we could perform the calculation separately for each variable (measure). Whilst this formulation (**Exceedence**), is characterised by the same limitations as the above *frequency* component, since it is calculated separately for each variable, when aggregated together to form an overall indicator, there is greater potential for improved resolution and granularity.

5.1.4 Hierarchical indices

The CCME WQI unifies all indicators into a single index as part of the calculations. However, most other indices involve aggregating across a sets of individual indicator scores. There are numerous ways to formulate indicator scores based on deviations from a benchmark (see Table 12).

Importantly, these scores are typically calculated at the level of the observations. Most of the index formulations are relatively robust to outliers (since the scores are either on a scale that reduces the magnitude of outliers or are capped to a range) and thus aggregating together indicies is likely to be more robust than calculating indices from aggregated raw data. An exception to this might be in situations where benchmarks are defined in the context of a specific spatial or temporal aggregation (such as annual mean or median value).

The Binary method expresses a comparison to benchmark values on a binary compliance scale (1: complies with benchmark, 0: fails to comply) and whilst simple to perform and understand, this method results in indices that have the potential to be either under or overly sensitive (depending on how far observed values typically are from the benchmark). For example, at one extreme (when values are close to benchmark), slight changes yield dramatic fluctuations in scores. However, when values are substantially above or below the benchmark, even modest improvements or deterioration will be undetected. This rapid 'switching' behaviour is depicted by the stepped response curve.

Note, when aggregated via means, the Binary method is identical to the Exceedence method, except that uncertaintly propagation is slightly more straight forward via the Binary method.

In the State of the Great Lakes Report (EPA/EC, 1995), greater granuality is achieved via a panel of experts who classify each of six health indicators (aquatic community health, human health, habitat, contaminants, nutrients and economy) into four categories: poor, mixed/deteriorating, mixed/improving, good/restored. Similar expert rating or multi-category exceedance grading systems are employed in other report cards (e.g Tamar estuary Report Card; Attard et al., 2012) and whilst probably reasonably accurate, they are nonetheless highly dependent on the ongoing availability of a reasonably stable panel of independent experts.

The Benchmark and Worst Case Scenario method (see Table 12) employed by the Fitzroy Basin Report Card (Jones et al., 2013) reflects the degree of failure by scaling the difference between the observed values and benchmarks (20_{th} or 80_{th} percentile of long term data for values above and below the benchmark respectively) to the Worst Case Scenario values (10_{th} or 90_{th} percentiles respectively). The associated response curve demonstrates a linear decline in Score with increasing distance from the benchmark.

The Modified Amplitude method calculates the distance to benchmark on a logarithmic (base 2) scale. The base 2 logarithm represents ratios on a symmetric scale such that value that are twice and half the benchmark yield scores of the same magnitude (yet apposing signs), and has some inbuilt capacity to accommodate skewed data. The Modified Amplitude response curve illustrates how this method can be simultaneously relatively insensitive to slight fluctuations around the benchmark as well as sensitive to changes further away from the benchmark.

Contrastingly, the Logistic Amplitude method operates on a logit scale such that it is very sensitive to slight fluctuations close to the benchmark and becomes progressively less sensitive with increasing distance. This method is also automatically scaled to the range [0,1]. The steepness of the Logistic Amplitude response can also be controlled by a tuning parameter (T).

Water Quality indices (which are standardized measures of condition) are typically expressed relative to a guideline, threshold (see Table A1 on page 177) or benchmark. Of the numerous calculation methods available, those that take into account the distance from the threshold (i.e. incorporate *difference-to-reference*) rather than simply an indication of whether or not a threshold value has been exceeded are likely to retain more information as well as being less sensitive to small changes in condition close to the threshold.

The challenging aspect of distance (or amplitude) based index methodologies is that determinination what constitutes a large deviation from a benchmark depends on the scale of the measure. For example, a deviation of 10 units might be considered relatively large of turbidity (NTU) or salinity (ppt), yet might be considered only minor for the Chlorophyll-a ($\mu g/L$). In order to combine a range of such metrics together into a meaningful index, the individual scores must be expressed on a common scale. Whilst this is automatically the case for Binary compliance, it is not necessarily the case for distance based indices.

Table 12 describes and compares the formulations and response curves of the Binary compliance method as well as a number of amplitude (distance based) indexing methods.

The Modified Amplitude and Logistic Modified Amplitude are both based on a base 2 logarithm of the ratio of observed values to the associated be benchmark (see Table 12). This scale ensures that distances to the benchmark are symmetric (in that a doubling and halving equate to the same magnitude - yet apposing sign). Furthermore, the logarithmic transformation does provide some inbuilt capacity to accommodate log-normality (a common property of measured values).

By altering the sign of the exponent, the Modified Amplitude methods can facilitate stressors and responses for which a failure to comply with a benchmark would be either above or below the benchmark (e.g. NTU vs Secchi depth). Further modifications can be applied to accommodate measures in which the benchmark represents the ideal and deviations either above or below represent increasingly poorer conditions (e.g. pH and dissolved oxygen).

The raw Modified Amplitude scores are relatively insensitive to small fluctuations around a benchmarks and sensitivity increases exponentially with increasing distance to the benchmark. The resulting scores can take any value in the real line $[-\infty, \infty]$ and hence are not bounded[6] There are two broad approaches to scaling (see Table 12):

1. Capping and scaling: The log_2 scale can be capped to a range representing either a constant extent of change (e.g. twice and half the benchmark - a cap factor of 2) or else use historical quantiles (10th and 90th percentiles) to define the upper and lower bounds to which to cap the scale. Note historical quantiles are unavailable for the current application[7]. Thereafter, either can be scaled to the range [0,1] via a simple formula (see Table 12 III.Scaled).

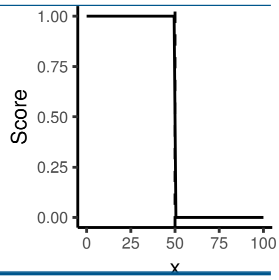
2. Logistic Modified Amplitude: By expressing the scores on a logistic scale, the range of scores can be automatically scaled to range [0,1]. Moreover, this method allows the shape of the response curve to be customized for purpose. For example, the relative sensitivity to changes close or far from the benchmarks can be altered by a tuning parameter.

Rather than aggregating across sites before calculating indices, we would suggest that indices should be calculated at the site level. This is particularly important when different measures are measured at different sites. Spatial variability can be addressed via the use of a bootstrapping routine (see below). We would recommend that measurements collected throughout the reporting year be aggregated together into a single annual value. This is primarily because most water quality thresholds pertain specifically to annual averages rather than single time samples. Although it is possible to incorporate uncertainty due to temporal variability, the low sparse temporal frequency of sample collection is likely to yield uncertainty characteristics that will swamp the more interesting spatial sources of uncertainty.

Alternatively, if we relax the application of thresholds to individual observations, annual indices can be generated by aggregating observations level indices. When doing so, the Binary Compliance formulation aggregated via means will yield identical outcomes to the Exceedence formulation.

A useful metric for comparing the sensitivity of one indexing method over another is to take some representative longitudinal data and calculate indices based on the actual data as well as data that introduces progressively more noise.

Table 12: Formulations and example response curves for a variety of indicator scoring methods that compare observed values (x_i) to associated benchmark, thresholds or references values (B_i and dashed line). The Scaled Modified Amplitude Method can be viewed as three Steps: I. Initial Score generation, II. Score capping (two alternatives are provided) and III. Scaling to the range [0,1]. The first of the alternative capping formulations simply caps the Scores to set values (on a log_2 scale), whereas the second formulation (Quantile based, where $Q1$ and $Q2$ are quantiles) allows thresholds quantiles to be used for capping purposes. Dotted lines represent capping boundaries. In the Logistic Scaled Amplitude method, T is a tuning parameter that controls the logistic rate (steepness at the inflection point). For the purpose of example, the benchmark was set to 50.

Method	**Formulation**	**Response curve**
Binary compliance	$Score_i = \begin{cases} 1 & \text{if } x_i \leq B_i \\ 0 & \text{if } x_i \text{ else} \end{cases}$	

[6]Unbounded indices are difficult to aggregate, since items that have very large magnitude scores will have more influence on the aggregation than those items with scores of smaller magnitude. Furthermore, unbounded scores are difficult to convert into alphanumeric Grades. Consequently, the Scores need to be scaled before they can be converted to alphabetical grading scale.

[7]The use of historical quantiles makes the explicit assumption that the domain of expectations (from very good to very poor) is encapsulated within the historical data. For the eReefs model data, only three years of historical data are available. This is unlikely to be sufficient to represent the full spread of what we should consider our expectations - particularly when we acknowledge that the eReefs model data do not extend back as far as the 2010-2011 floods during which water quality conditions might be expected to be lower than the years to follow.

Table 12: Report Card indexing methods, continued

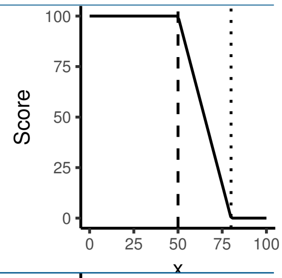
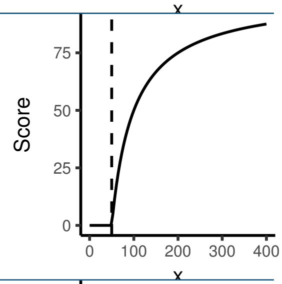
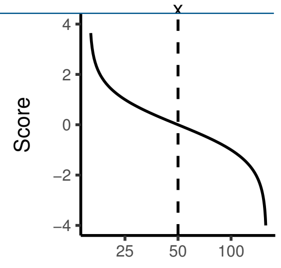
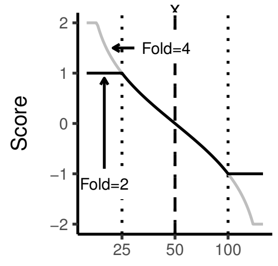
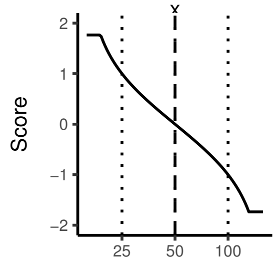
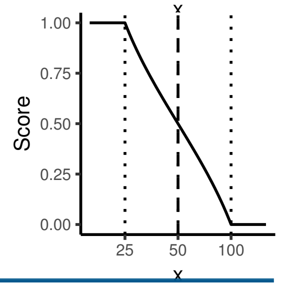
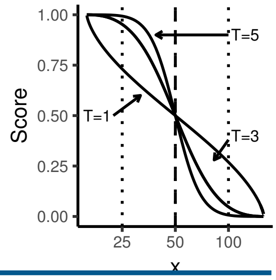
Method	**Formulation**	**Response curve**
Benchmark and WCS	$Score_i = \begin{cases} 100 & \text{if } x_i \leq B_i \\ 0 & \text{if } x_i \geq WCS_i \\ \left[1.0 - \left\lvert \frac{x_i - B_i}{WCS_i - B_i} \right\rvert \right].100 & \text{else} \end{cases}$	
Amplitude	$Score_i = \begin{cases} (\frac{x_i}{B_i})^{-1} & \text{if } > B_i = \text{fail} \\ (\frac{x_i}{B_i})^{1} & \text{if } < B_i = \text{fail} \end{cases}$ $Score_i = \frac{100 \times Score_i}{1 + Score_i}$	
Modified Amplitude	I. Raw (MAMP) $Score_i = \begin{cases} log_2(\frac{x_i}{B_i})^{-1} & \text{if } > B_i = \text{fail} \\ log_2(\frac{x_i}{B_i})^{1} & \text{if } < B_i = \text{fail} \end{cases}$	
	II. Fixed caps (Fold=2; [0.5,2]) (Fold=4; [0.25,4]) $Score_i = \begin{cases} log_2(1/2) & \text{if } Score_i < -1 \\ log_2(2/1) & \text{if } Score_i > 1 \\ Score_i & otherwise \end{cases}$	
	II. Quantile/extremes based caps ([15,170]) $Score_i = \begin{cases} log_2(\frac{Q1}{B_i})^{-1} & \text{if } x_i < Q1 \\ log_2(\frac{Q2}{B_i})^{-1} & \text{if } x_i > Q2 \\ Score_i & otherwise \end{cases}$	
	III. Scaled (Fixed: Fold=2) $Score_i = \frac{Score_i - min(Score_i)}{max(Score_i) - min(Score_i)}$	

Table 12: Report Card indexing methods, continued

Method	**Formulation**	**Response curve**
Logistic Scaled Modified Amplitude	Raw $Score_i = \begin{cases} log_2(\frac{x_i}{B_i})^{-1} & \text{if} > B_i = \text{fail} \\ log_2(\frac{x_i}{B_i})^{1} & \text{if} < B_i = \text{fail} \end{cases}$ $Score_i = \frac{1}{1+e^{Score_i \cdot -T}}$	
Logistic	Raw $Score_i = \begin{cases} \frac{1}{1+e^{T.(x_i/B_i)}} & \text{if} > B_i = \text{fail} \\ \frac{1}{1+e^{-T.(x_i/B_i)}} & \text{if} < B_i = \text{fail} \end{cases}$	

Whilst the state of the water (or other environmental condition) might be of interest in its own right, it might also be of interest from the perspective of the ecosystem supported by the water. For example, turbidity might be considered to provide important insights into the light availabity within the ecosystem. As such, the variability in light availability (turbidity) might be a more influential ecological driver/pressure than the exact light level within any given time frame. Furthermore, sustained conditions might be more influential than rapidly fluctuating conditions. For example, two time windows could experience the same turbidity average and variance, yet these summaries could manifest from very different fluctuation patterns (one experiencing rapid fluctuations, and the other experiencing sustained periods of contrasting conditions).

One index that captures the pattern of fluctuations could be based on a metric that expresses the number of consecutive days in which a threshold has been exceeded as a proportion of number of days in the time window (e.g. 365 days).

$$Score_i = 1 - (n_i/N_i)$$

where n_i is the maximum number of consecutive time units in which $x_i > B_i$ and N_i is the number of time units in the i^{th} spatio-temporal window.

Unfortunately, such a formulation imposes some relatively difficult requirements on the data. Firstly, the time series within each window must be complete (no gaps), otherwise it is difficult to asses N_i. This requirement limits its use to only the eReefs modelled data as the Satellite data, AIMS insitu and AIMS FLNTU data have substantial time gaps. Secondly, as the formulation is based on summing up exceedences, it is likely to be as susceptable to the recognised insensitivies associated with binary compliance. Indeed, these sensitivities may well be further amplified. Furthermore, it is not responsive to the magnitude of exceedence.

The next section will explore the performance of the following index formulations:

- Binary compliance (Binary)
- Exceedence - proportion of observations exceeding the threshold (on large datasets, this will converge with Binaary compliance (Exceed)
- Maximum duration of exceedence (Max_Duration)
- Modified Amplitude (MAMP)
- Fixed Modified Amplitude (fMAMP)
- Fixed Scaled (x2,1/2) Modified Amplitude (fsMAMP)
- Fixed Scaled (x4,1/4) Modified Amplitude (fsMAMP4)

5.2 Index sensitivity

The sensitivity of a metric can be gauged by either:

- Quantitative exploration of the relationships between the metric and gradients of the underlying conditions that the metric should respond to. This approach requires very well defined gradients as well as a clear understanding and measures of what constitutes a relationship. By optimizing the metric(s) to these gradients, this approach has the potential to bias outcomes towards these gradients at the expense of generality to other gradients.
- Have experts (or end users) qualitatively gauge the outcomes of different metrics against expected trends and patterns. That is, do the outcomes align with end user expectations. Although this approach is equally subjective and potentially biased as the quantitative exploration, it does not necessitate formulating statistical cutoffs and associated artifacts.
- Explore the behaviour and characteristics of the metric when calculated on data simulated to represent a range of scenarios (altering location and spread). Whilst this approach will not necessarily select the 'best' metric, it does permit identification of the limitations and assumptions associated with different metrics.

The above approaches are not mutually exclusive. The current project will explicitly explore sensitivity via a simulation approach, yet will also encourage feedback as to whether final outcomes align with expectations. It should be noted that the current project is limited in sources of data and measured properties. A metric is purely a re-expression of data in order to enhance or highlight a signal. If the underlying data do not contain the expected signal, a signal will likewise be absent from any metrics.

To explore the performance and sensitivity of the various index computations for a range of data scenarios, data were simulated from Gamma distributions varying in mean (relative to a threshold) and variance and sample size. The Gamma distribution is parameterized by two shape parameters that can be expressed in terms of mean and variance ($Gamma(\mu^2/\sigma^2, \mu/\sigma^2)$).

For each threshold value (GL = 0.1,0.2,0.5,1,1,10,100) and sample size (R=10,100,1000), a set of 28 data scenarios where simulated (see Table 13 so as to represent a full spectrum of possible sampling outcomes. For each threshold/sample size and set combination, indices were calculated and aggregated for the simulated data. The extremes of these combinations are presented in Figures 33, 36 and 37, a more extensive set of Figures are in Appendix **??**. For the set of simulations, the smaller the threshold, the more variable the samples relative to the threshold. Within each threshold, the set of 28 scenarios thereby represent combinations of varying mean and relative variability.

Table 13: Index performance and sensitivity data scenarios. Data in each group are drawn from Gamma distributions whose parameterizations are based on a mean and variance. In each case the mean is some multiple of the threshold (GL) value. Multiples of threshold that are less than 1 result in data with greatest density below the threshold value. Lower variances result in less varied data.

Grp	Mean	SD	Grp	Mean	SD	Grp	Mean	SD	Grp	Mean	SD
1	$\mu = 0.2GL$	$\sigma^2 = 0.1$	9	$\mu = 0.75GL$	$\sigma^2 = 0.1$	17	$\mu = 1.5GL$	$\sigma^2 = 0.1$	25	$\mu = 4GL$	$\sigma^2 = 0.1$
2	$\mu = 0.2GL$	$\sigma^2 = 0.2$	10	$\mu = 0.75GL$	$\sigma^2 = 0.2$	18	$\mu = 1.5GL$	$\sigma^2 = 0.2$	26	$\mu = 4GL$	$\sigma^2 = 0.2$
3	$\mu = 0.2GL$	$\sigma^2 = 0.3$	11	$\mu = 0.75GL$	$\sigma^2 = 0.3$	19	$\mu = 1.5GL$	$\sigma^2 = 0.3$	27	$\mu = 4GL$	$\sigma^2 = 0.3$
4	$\mu = 0.2GL$	$\sigma^2 = 0.5$	12	$\mu = 0.75GL$	$\sigma^2 = 0.5$	20	$\mu = 1.5GL$	$\sigma^2 = 0.5$	28	$\mu = 4GL$	$\sigma^2 = 0.5$
5	$\mu = 0.5GL$	$\sigma^2 = 0.1$	13	$\mu = 1GL$	$\sigma^2 = 0.1$	21	$\mu = 2GL$	$\sigma^2 = 0.1$			
6	$\mu = 0.5GL$	$\sigma^2 = 0.2$	14	$\mu = 1GL$	$\sigma^2 = 0.2$	22	$\mu = 2GL$	$\sigma^2 = 0.2$			
7	$\mu = 0.5GL$	$\sigma^2 = 0.3$	15	$\mu = 1GL$	$\sigma^2 = 0.3$	23	$\mu = 2GL$	$\sigma^2 = 0.3$			
8	$\mu = 0.5GL$	$\sigma^2 = 0.5$	16	$\mu = 1GL$	$\sigma^2 = 0.5$	24	$\mu = 2GL$	$\sigma^2 = 0.5$			

Figure 33: Simulated data and associated indices for threshold of 0.1 and very large sample sizes (R=1000). Samples represent high variability relative to threshold.

As expected, indices decline with increasing values relative to the threshold (as would be the case for Chl-a or TSS) with a generally linear response being the attribute sought in our specific context. Testing the responses of indices to various combinations allowed the identification of the most appropriate and robust index calculation method.

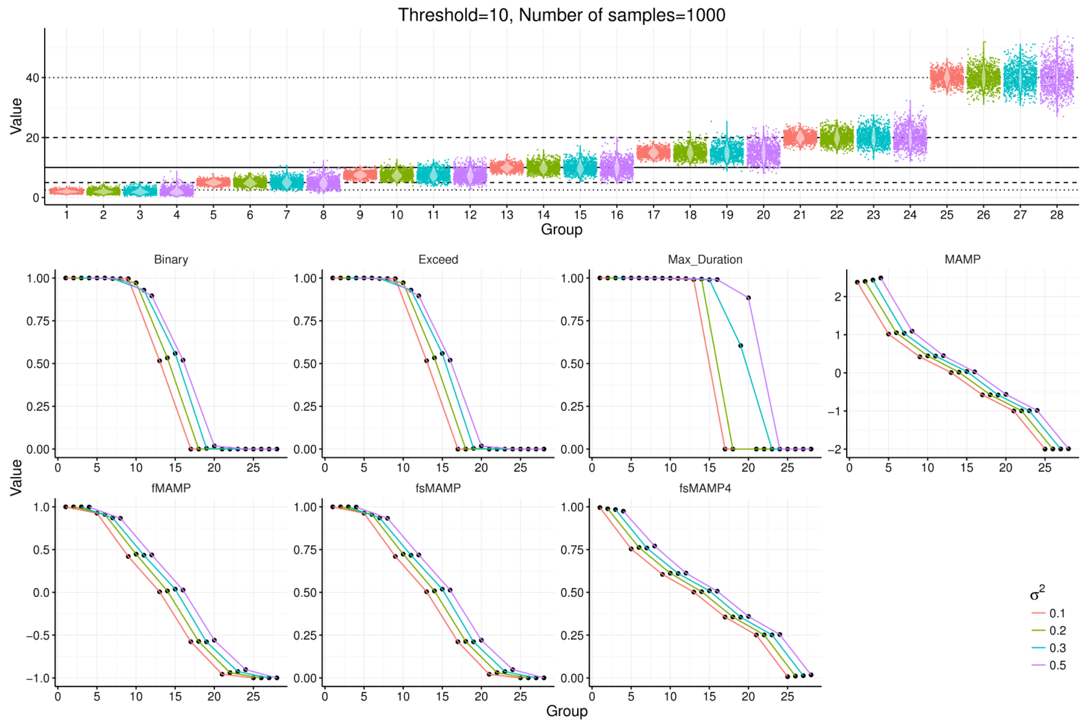


Figure 34: Simulated data and associated indices for threshold of 10 and very large sample sizes (R=1000).

When the number of samples and the relative sample variability is very large (e.g. fig. 33), with the exception of the the maximum duration of exceedance and the uncapped and unscaled modified amplitude (MAMP) methods, the different index calculation methods behave very similarly. However, as the variability of the samples declines relative to the threshold (e.g. compare figs. 33, 34 and 35), such that observations are predominantly within twice/half the threshold value, and data is predominantly distributed between the threshold value binary or frequency of exceedance methods both increasingly become simultaneously overly and under sensitive. The response curve of these metrics becomes less linear, whereas the linearity of the other metrics is maintained for a greater span of observation means. This is further exacerbated by small sample sizes (see fig.37).

Over all of the scenarios, the fsMAMP4 (Modified Amplitude capped at four times/quarter of threshold values) appears to be as linear or more linear than the fsMAMP (Modified Amplitude capped at twice/half), particularly as relative variability declines. However, the cost of this extended range of sensitivity, is that it is predominantly more sensitive at the extremes and less so (at least compared to fsMAMP) towards the mid-region (corresponding to values close to the threshold). Arguably, it is more desirable for an index to be most sensitive around the threshold (unless there is substantial uncertainty about the threshold value) and become progressively less sensitive at increasing distance from the threshold - the binary and exceedence metrics are the extreme cases of this.

The fixed capped modified amplitude (fsMAMP) index was considered the 'best' compromise between consistent sensitivity throughout the range of scenarios and the nature of data presented in exploratory data analyses (see Section **??**). It should be noted that it is possible to modify the fsMAMP index metric to facilitate caps based on historical, biological or ecological parameters. It is also possible to define these parameters (an upper and lower capping) at any spatial/temporal/measure level so as to potentially build indices that are optimized for each measure. Such an exercise requires extensive expert knowledge to define and justify each of the parameters and is beyond the scope of the current project.

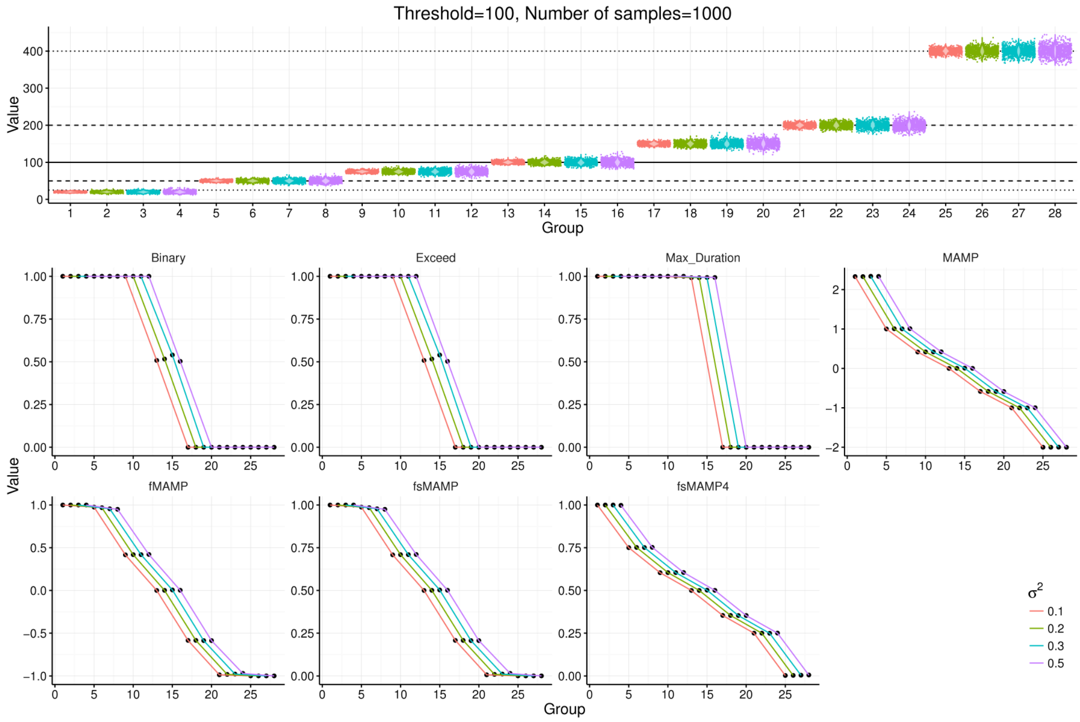


Figure 35: Simulated data and associated indices for threshold of 100 and very large sample sizes (R=1000).

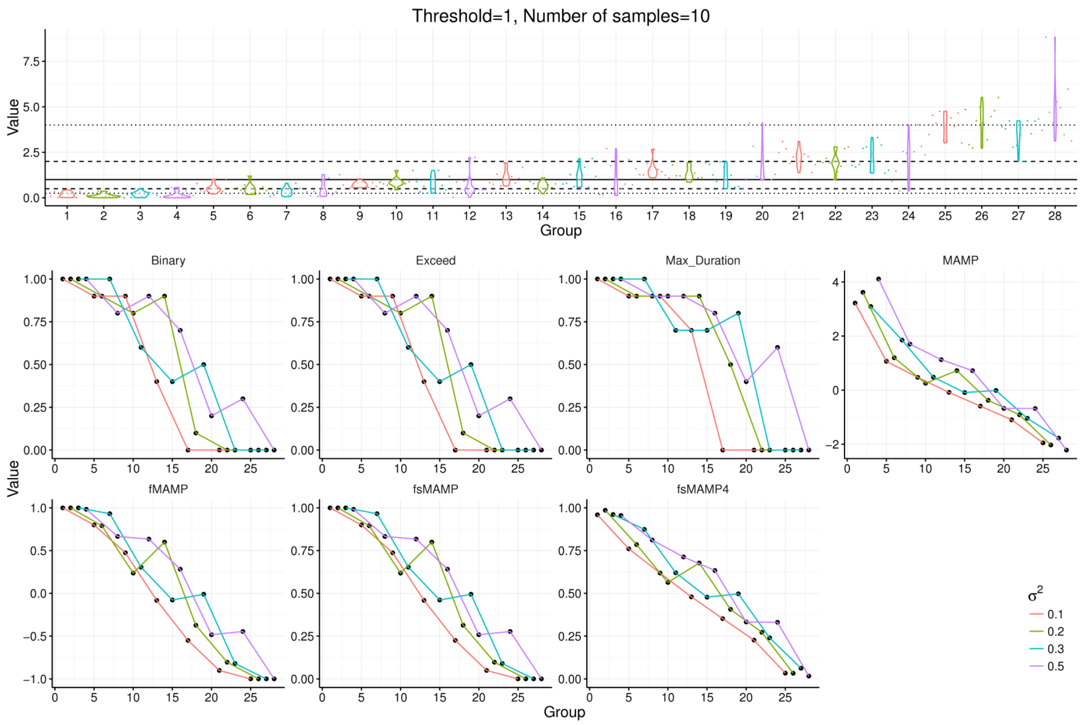


Figure 36: Simulated data and associated indices for threshold of 1 and large sample sizes (R=100).

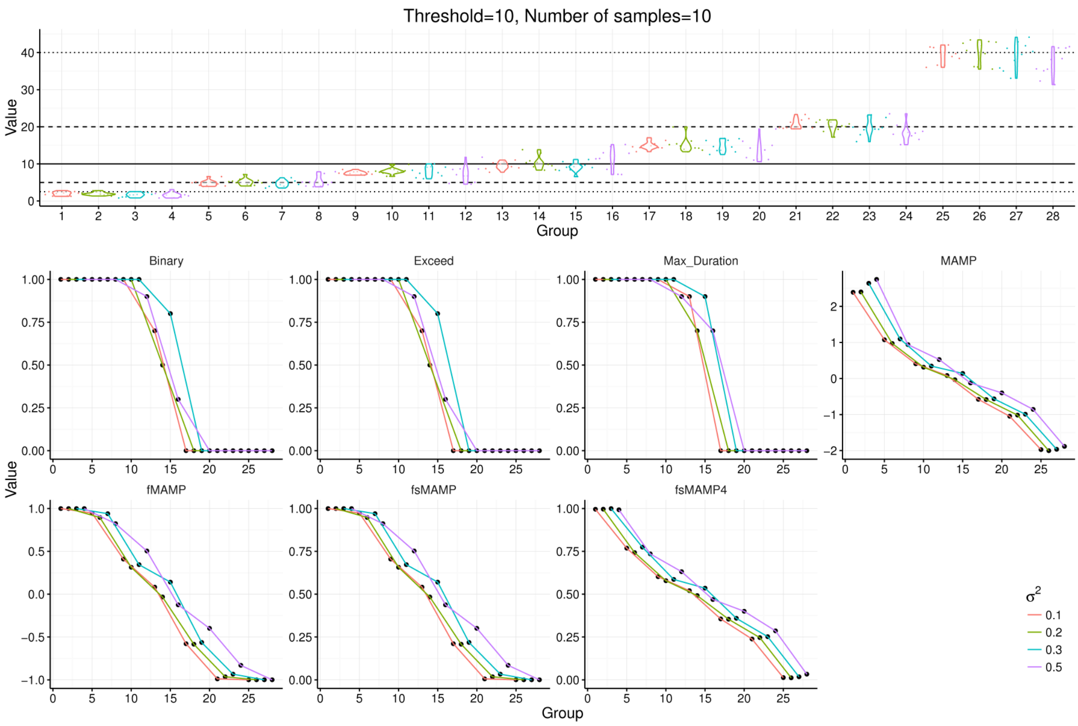


Figure 37: Simulated data and associated indices for threshold of 10 and small sample sizes (R=10).

5.2.0.1 Summary of simulation index sensitivity exploration

- Indices decline with increasing values relative to the thresholds (and for a given variability)
- Indices increase with increasing variability (since in Gamma distributions, this results in more values towards lower end)
- when R is very large, the different indicators behave similarly (except Max_Duration and MAMP)
- MAMP is more susceptable to outliers

5.3 Index explorations

Before data can be combined and aggregated across the various Sources (AIMS insitu, AIMS FLNTU, Satellite, eReefs and eReefs926) and Measures (Chlorophyll, TSS, Secchi depth and NOx), it is important that we evaluate the likely usefulness of each Source/Measure combination. For example, a Measure or Source that does not vary in both time and space is not considered very informative parameter.

Although an exploration of the patterns of spatial and temporal variation of the raw data does offer some insights into the usefulness of a parameter, it is variation in relation to expectations (thresholds) that are likely to be of greatest utility. For example, a parameter might vary substantially in time and or space and yet always be well above (or below) the threshold. In this situation (despite the apparent variability), with respect to the expectation domain, there is very little (if any) variability and thus the realised utility of the parameter is low (or else the threshold is inappropriate for the particular measure to which it is being applied).

Different parameters are measured on different scales or else have different natural background levels. Since variability (for example variance) is dependent on scale, parameters measured in larger units will typically exhibit more variability in absolute terms. Hence, in order to compare the relative utility of different parameters, it is necessary to either express variation relative to scale (such as coefficient of variation) or standardize the parameters. The scaled hierarchical index formations of Section 5.1.4 (such as Binary, fsMAMP, fsMAMP4 and logistic MAMP) are all a form of standardization which yeild scores on scales that are all bound [0,1].

The following three subsections will provide information to assist in the selection of:

- which Index formulation to adopt
- which Sources of data to use
- which Measures to include

5.3.1 Indices

Theoretical sensitivity investigation suggested that the fixed capped (half/twice threshold) Modified Amplitude (fsMAMP) is likely to be the best compromise between under and over sensitivity given the patterns of variance observed across and between the various Sources (AIMS insitu, AIMS FLNTU, Satellite, eReefs and eReefs926) and Measures (Chlorophyll, TSS, Secchi depth and NOx). The alternate approach is to explore and compare the patterns of the various index formulations in the context of both the raw collected data and expert expectations. Broadly speaking, we might expect that many water Quality parameters improve across the shelf with increasing distance from coastline. We might also expect some latitudinal patterns in which water quality generally improves along a south-north gradient with interruptions coinciding with outflow of major rivers.

To explore how the raw data are transformed into the various indices, it is useful to pair up 'before' and 'after' figures. Again, for the sake of brevity, we will focus on the same data that featured in Figure 9 (Chlorophyll-a from Wet Tropics, Open Coastal). Figures 38 – 42 illustrate the associations between the site means (subfigure a) and three of the major index candidates (b: Binary, c: fsMAMP and d: fsMAMP4) for each of the Sources of data (AIMS insitu, AIMS FLNTU, Satellite, eReefs and eReefs926). In these figures, purple and blue lines represent annual means and within year Generalized Additive Model (Wood, 2006) respectively and help highlight inter- and intra-annual variation[8].

Similar figures for the other Measures (Total Suspended Solids, Secchi Depth and NOx) for the Dry Tropics Midshelf zone are presented in Appendix Figures C97–C142.

Inter and Intra annual variation is greatest in the Binary index method for each data Source[9]. Whilst this method does illustrate sensitivity, the values of the index do not contain any context about the magnitude of values relative to the threshold. That is, it is not possible to distinguish situations in which all observations are just under (or over) the threshold from when they are substantially under (or over) the threshold. In this way, the index has the potential to be under-sensitive to magnitude, yet very sensitive to change around the threshold. For each of the Sources (except AIMS insitu for which data are too sparse), the relative magnitude of fluctuations in the Binary index (subfigure b) appears to be substantially greater than the relative magnitude of fluctuation in the observed data (subfigure a). These patterns of relative variability might imply that the Binary index is over-sensitive.

[8] GAMs not performed for AIMS insitu data due to a lack of data over which to estimate splines

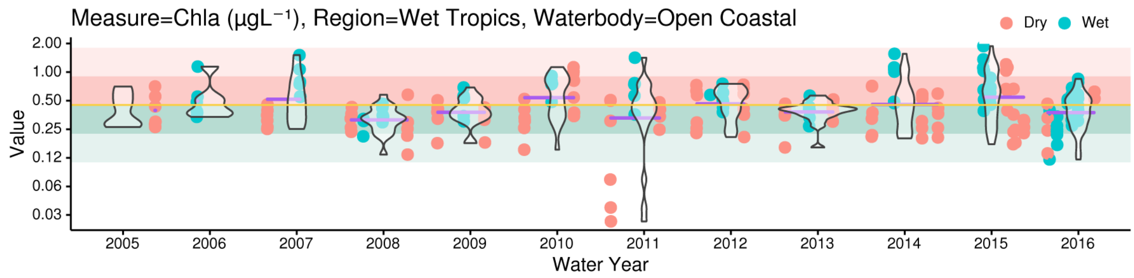
[9] this pattern also persists across all Zones (Region/Water body) and Measures - although other Measures and Zones not provided here to reduce space.

By contrast, the fsMAMP4 (capped at four times and one-forth threshold, subfigures d) could be interpreted as under-sensitive - particularly for the Satellite data (which has highly variable observations). The fsMAMP (twice/half threshold) appears to in between these two extremes and thus could be considered a reasonable comprimise between over and under sensitivity.

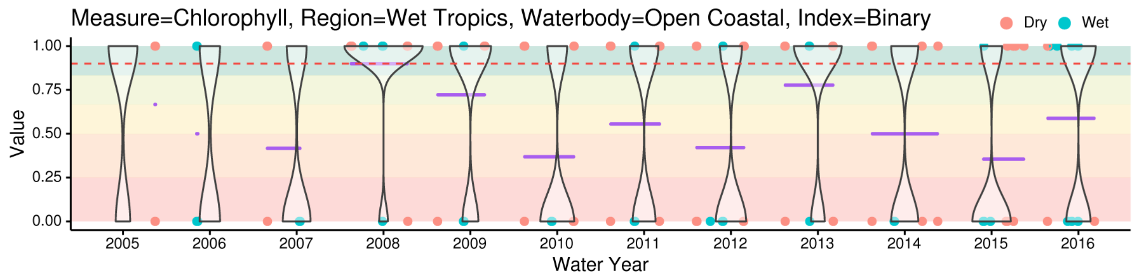
Spatial representations for Wet Tropics Open Coastal Chlorophyll-a (figs. 43 – 46) and Dry Tropics Midshelf Chlorophyll-a (figs. 47 – 51) offer similar assessments - that fsMAMP provides a reasonable compromise between the potentially under and over sensitive fsMAMP4 and Binary formulations. Similar representations for Total Suspended Solids, Secchi Depth and NOx are presented in Appendix Figures C122 – C143

Time series of annually aggregated observations and associated annually aggregated indices (figs. 52 – 56) provide simplified representations of the overall spatio-temporal patterns. As with the temporal and spatial representations, the fsMAMP index consistently manifests between the Binary and fsMAMP4 formulations.

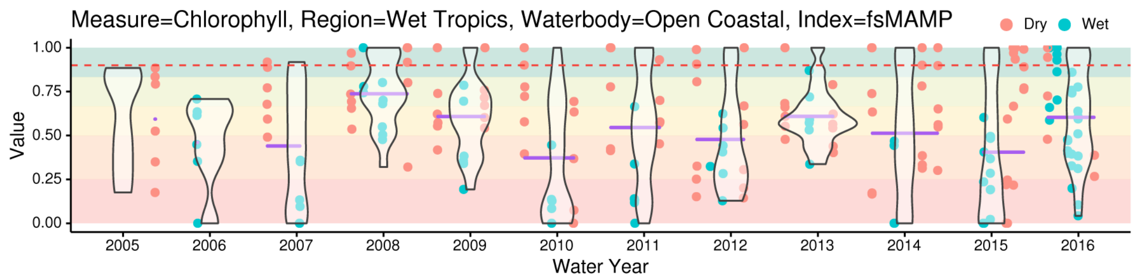
a) AIMS insitu site means



b) AIMS insitu site mean Binary



c) AIMS insitu site mean fsMAMP



d) AIMS insitu site mean fsMAMP4

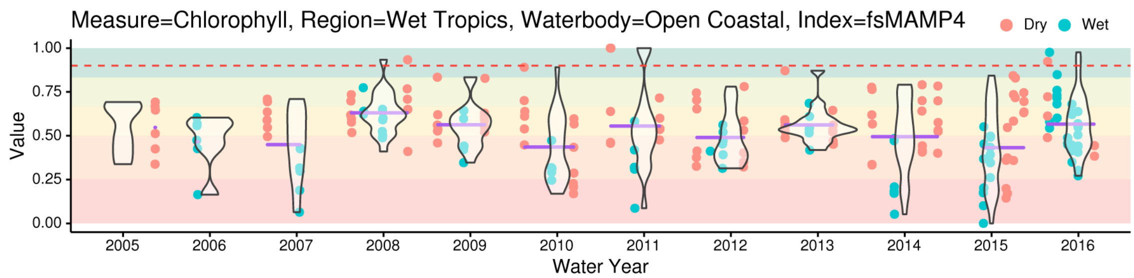
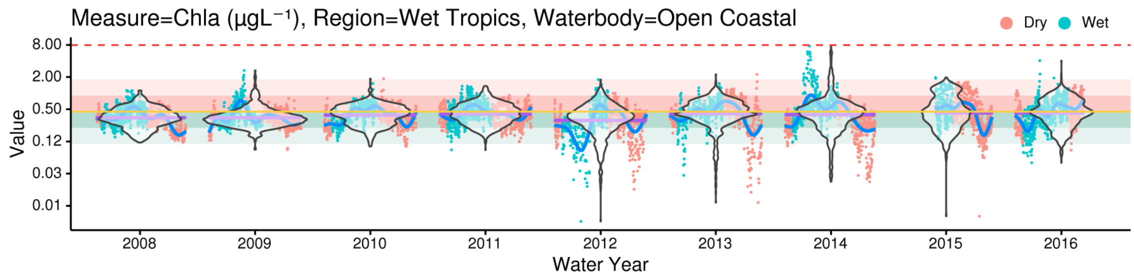
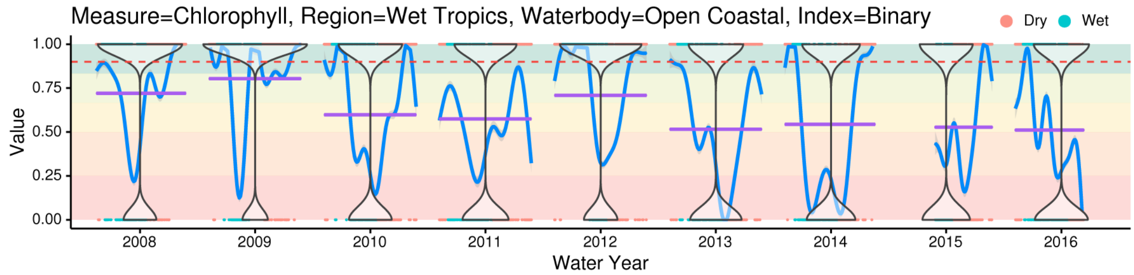


Figure 38: Temporal distribution of AIMS insitu Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) respectively represent half and twice threshold value (50% shading) and one-forth and four times threshold value (30% shading). Traffic-light banding on b-d) indicates simple 5-level color scheme. Purple lines represent annual means.

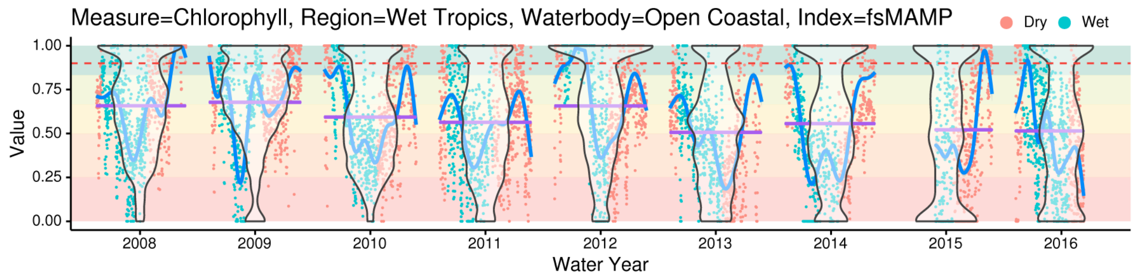
a) AIMS FLNTU raw site means



b) AIMS FLNTU site mean Binary



c) AIMS FLNTU site mean fsMAMP



d) AIMS FLNTU site mean fsMAMP4

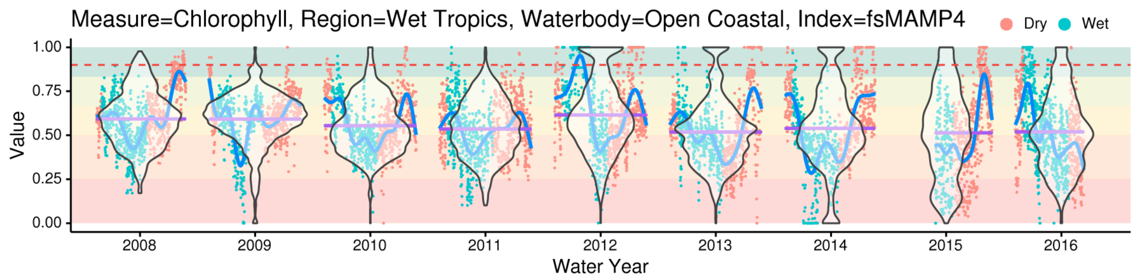
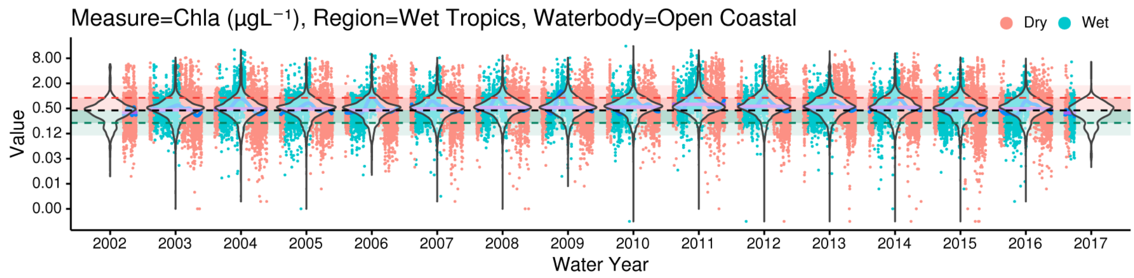
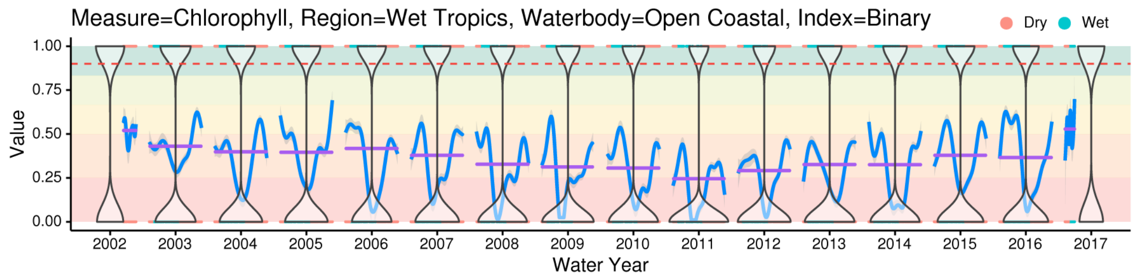


Figure 39: Temporal distribution of AIMS FLNTU Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) respectively represent half and twice threshold value (50% shading) and one-forth and four times threshold value (30% shading). Traffic-light banding on b-d) indicates simple 5-level color scheme. Purple lines represent annual means.

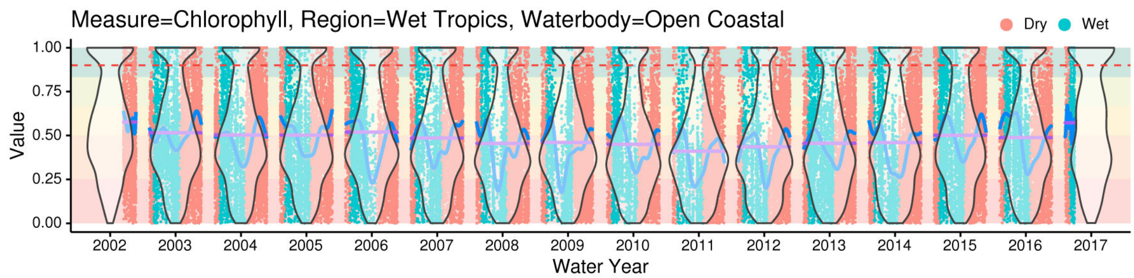
a) Satellite raw site means



b) Satellite site mean Binary



c) Satellite site mean fsMAMP



d) Satellite site mean fsMAMP4

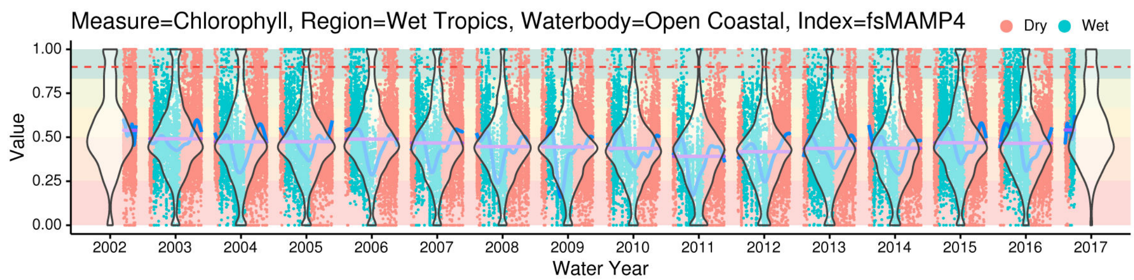
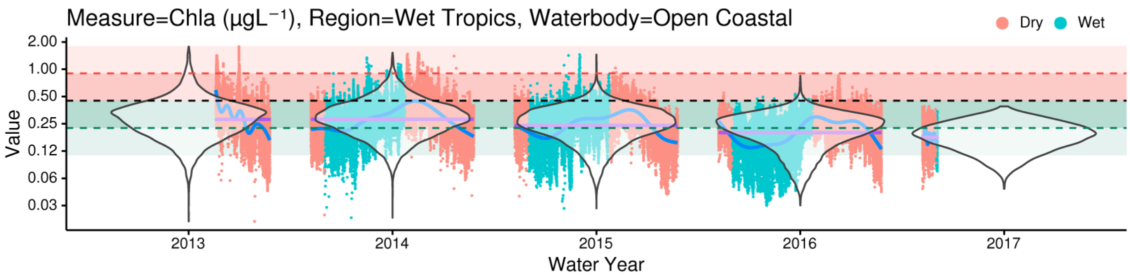
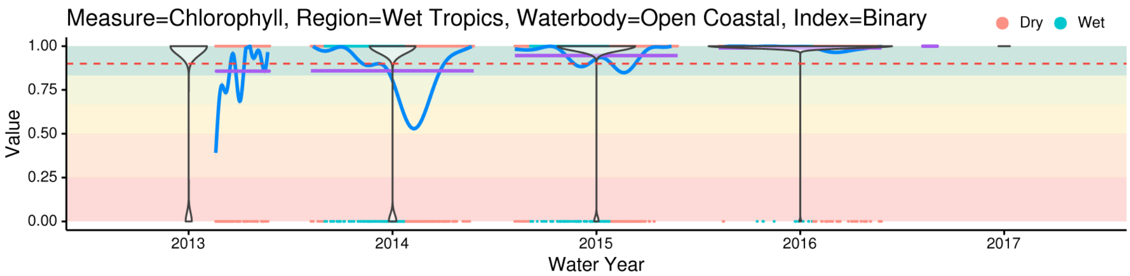


Figure 40: Temporal distribution of Satellite Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) respectively represent half and twice threshold value (50% shading) and one-forth and four times threshold value (30% shading). Traffic-light banding on b-d) indicates simple 5-level color scheme. Purple lines represent annual means.

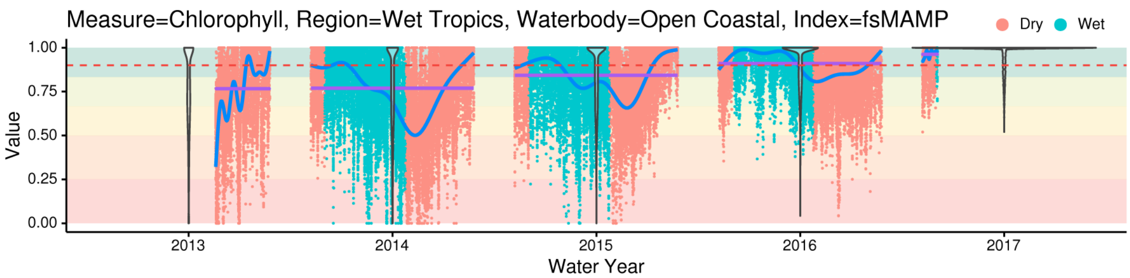
a) eReefs raw site means



b) eReefs site mean Binary



c) eReefs site mean fsMAMP



d) eReefs site mean fsMAMP4

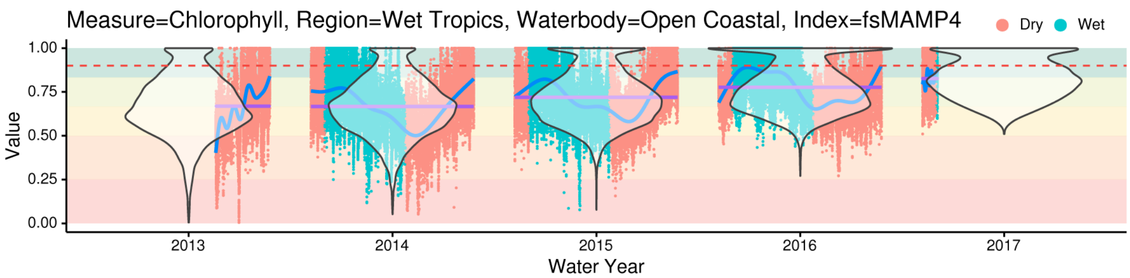
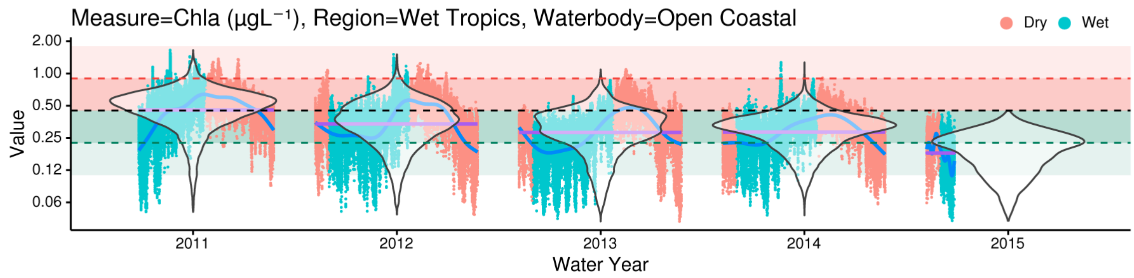
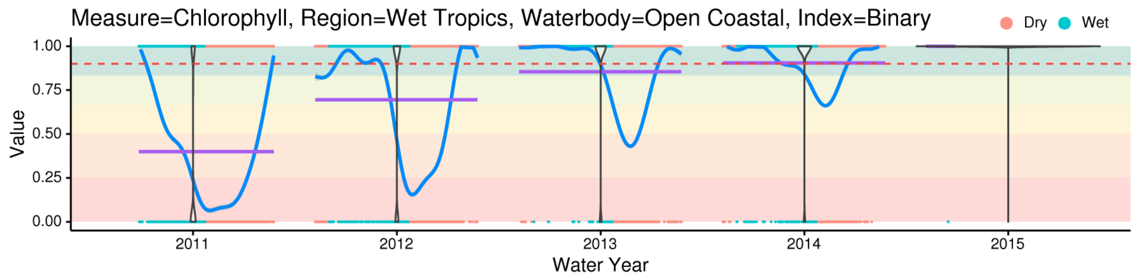


Figure 41: Temporal distribution of eReefs Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) respectively represent half and twice threshold value (50% shading) and one-forth and four times threshold value (30% shading). Traffic-light banding on b-d) indicates simple 5-level color scheme. Purple lines represent annual means.

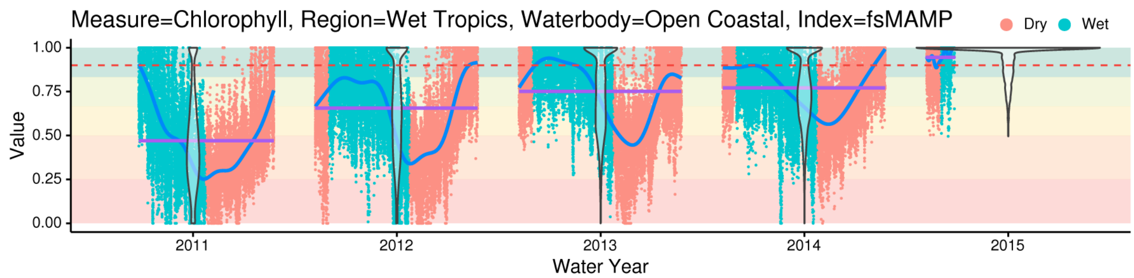
a) eReefs926 raw site means



b) eReefs926 site mean Binary



c) eReefs926 site mean fsMAMP



d) eReefs926 site mean fsMAMP4

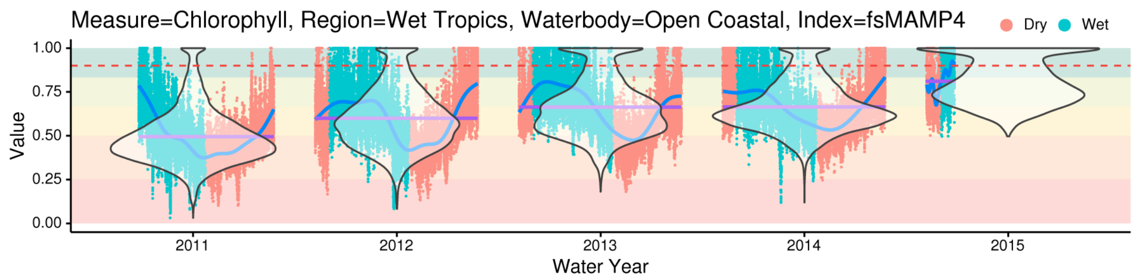


Figure 42: Temporal distribution of eReefs926 Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) respectively represent half and twice threshold value (50% shading) and one-forth and four times threshold value (30% shading). Traffic-light banding on b-d) indicates simple 5-level color scheme. Purple lines represent annual means.

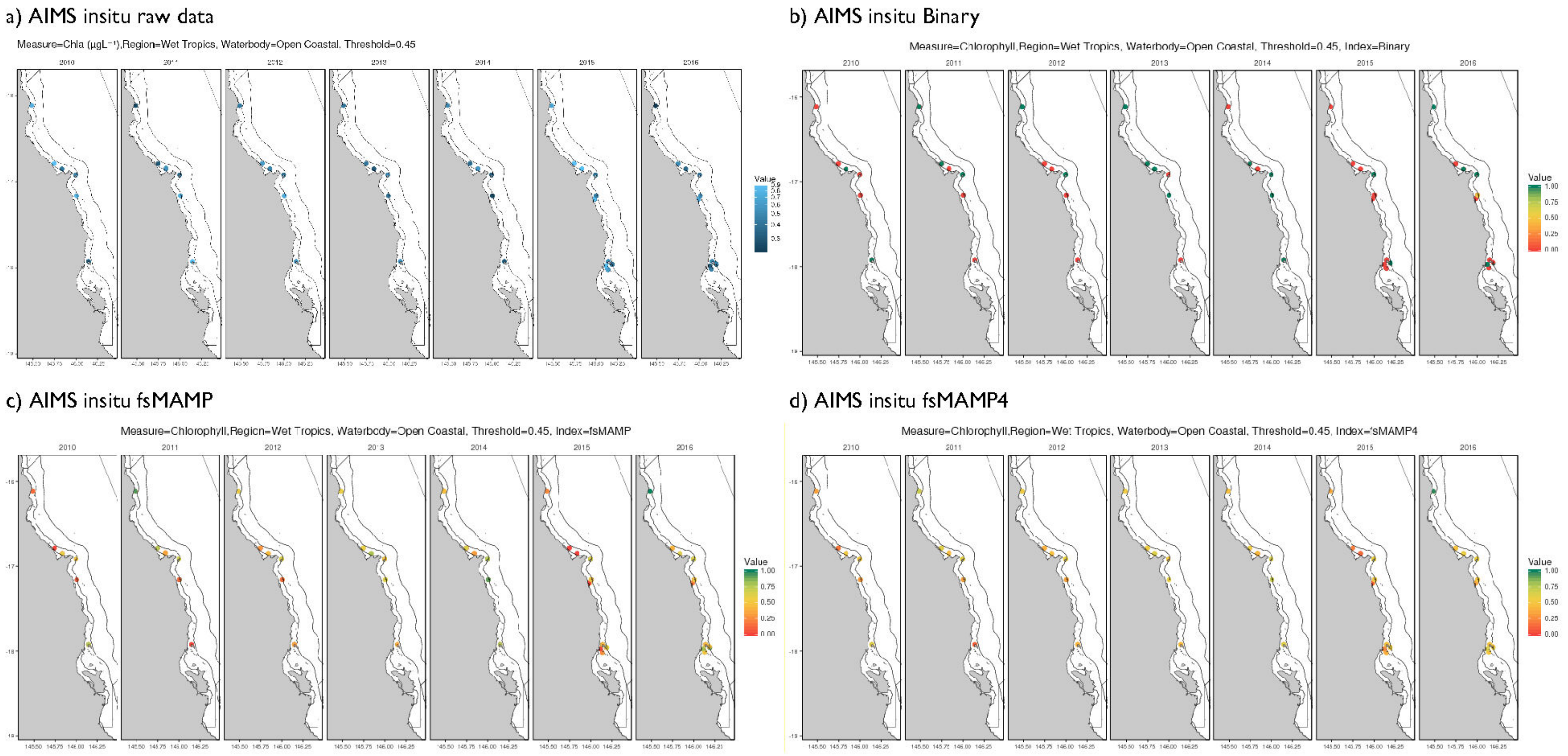


Figure 43: Spatial distribution of AIMS in situ Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

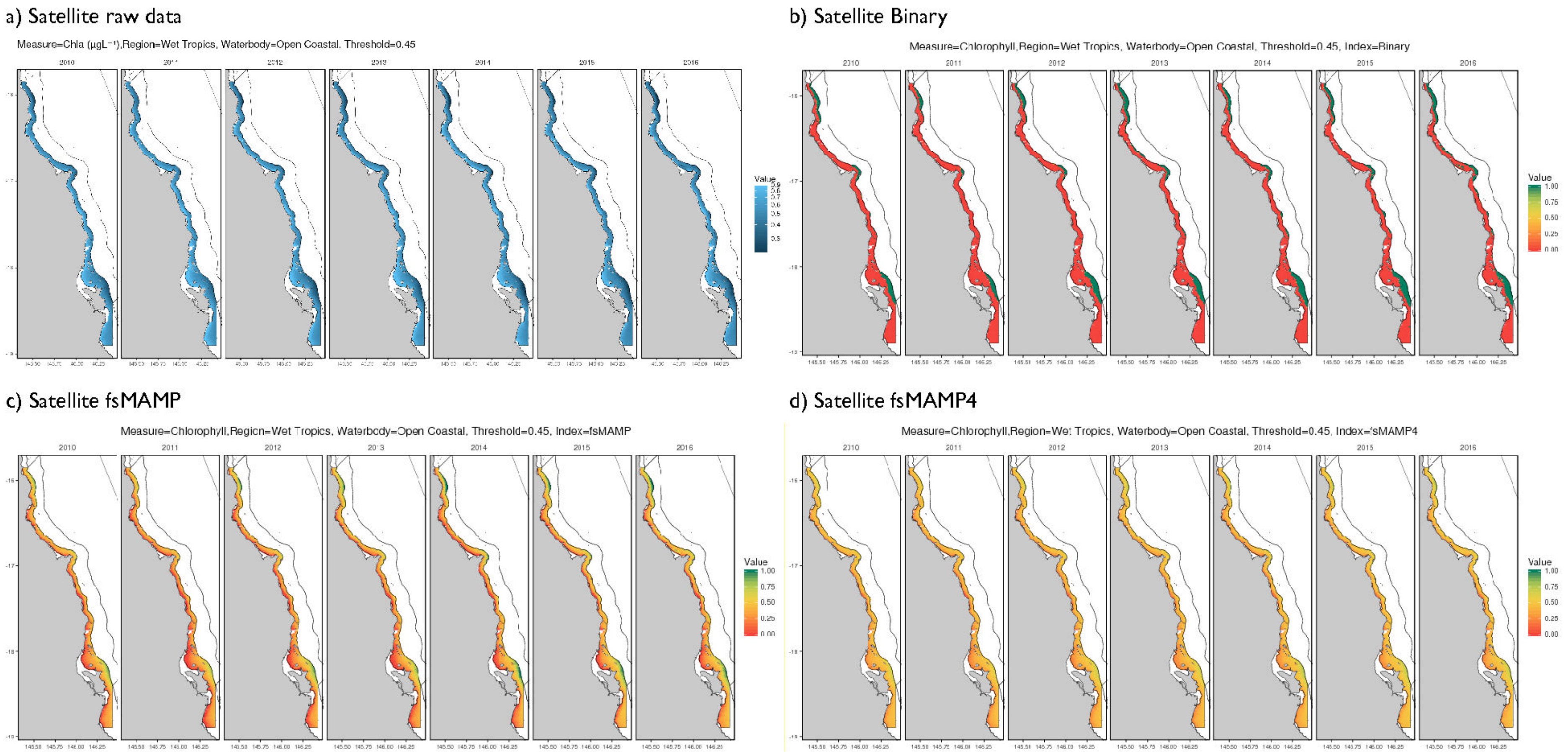


Figure 44: Spatial distribution of Satellite Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

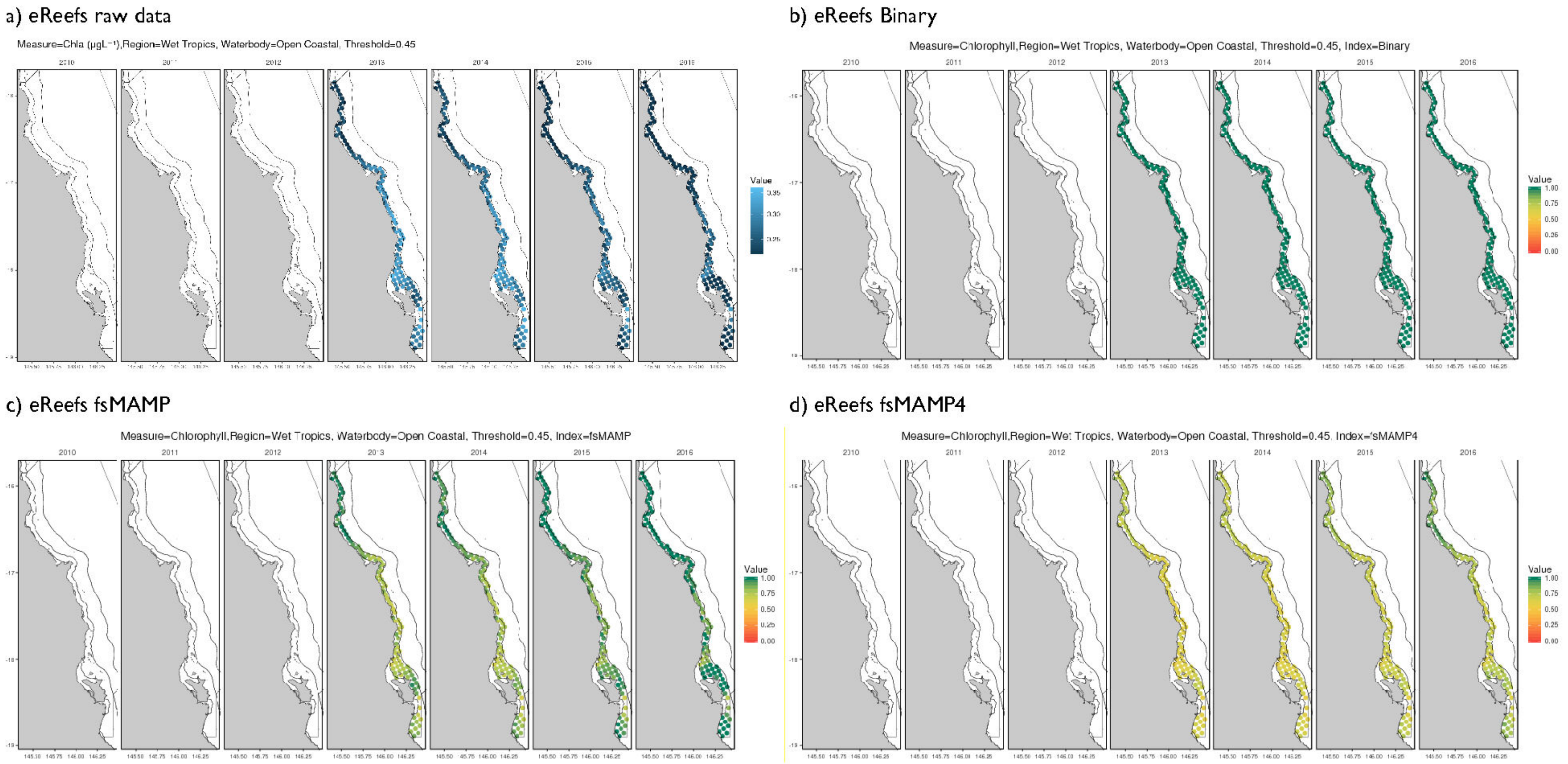


Figure 45: Spatial distribution of eReefs Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

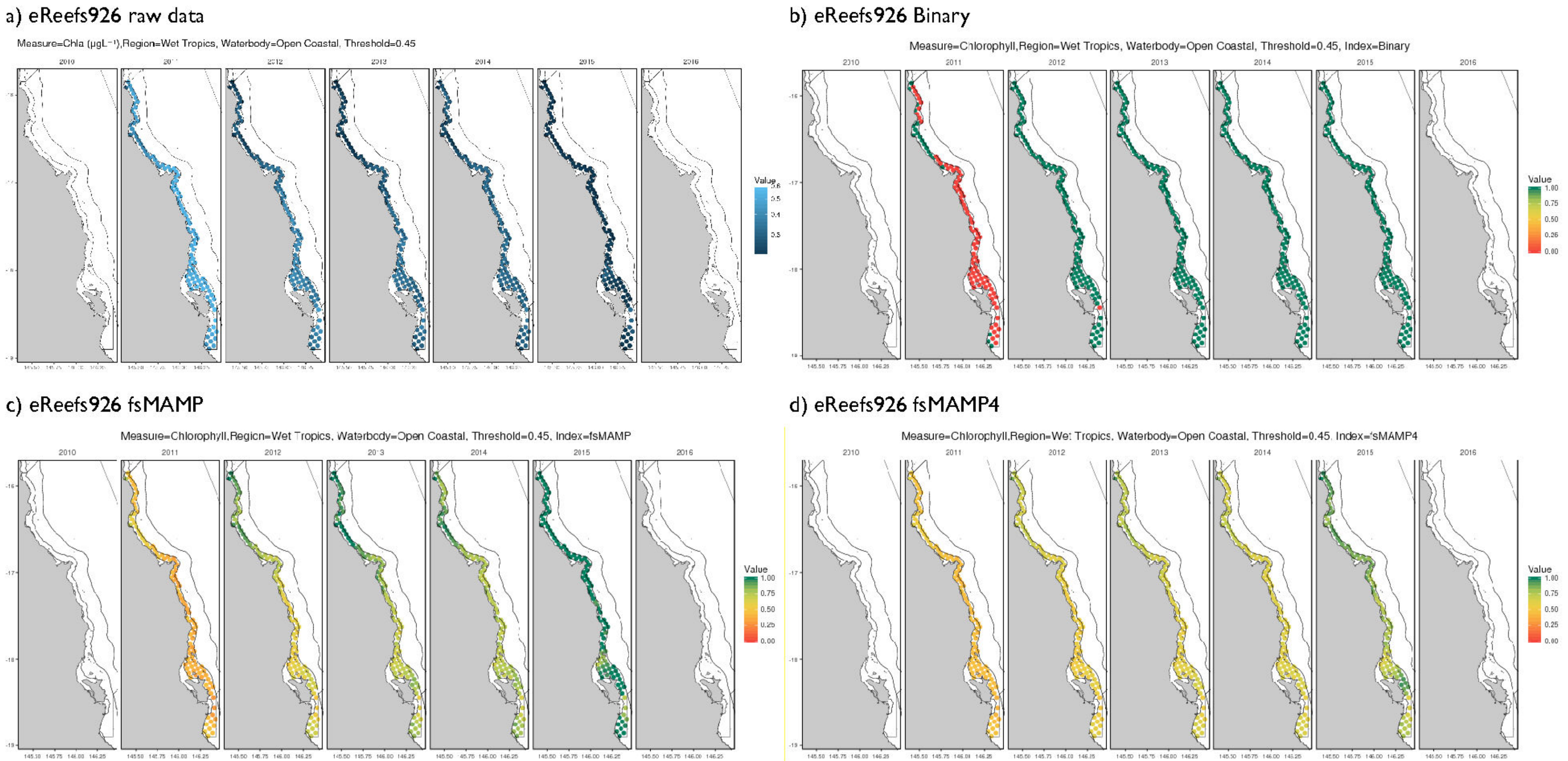
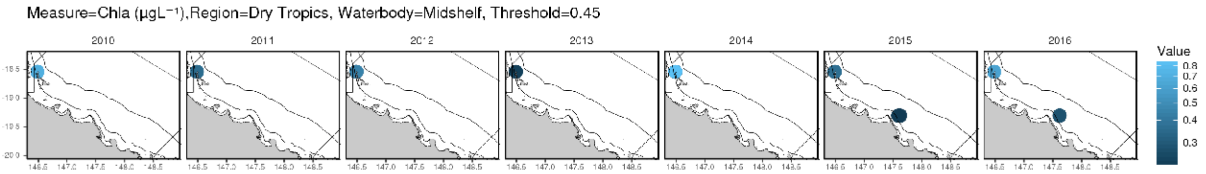
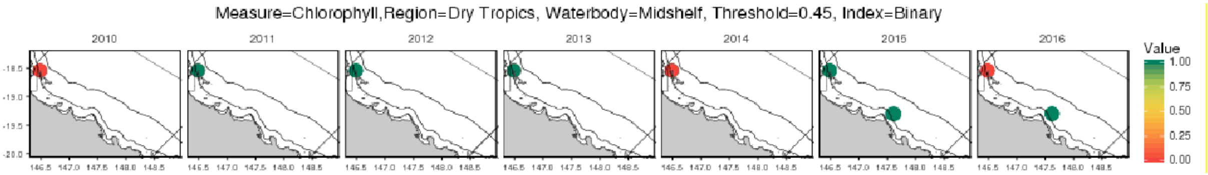


Figure 46: Spatial distribution of eReefs926 Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

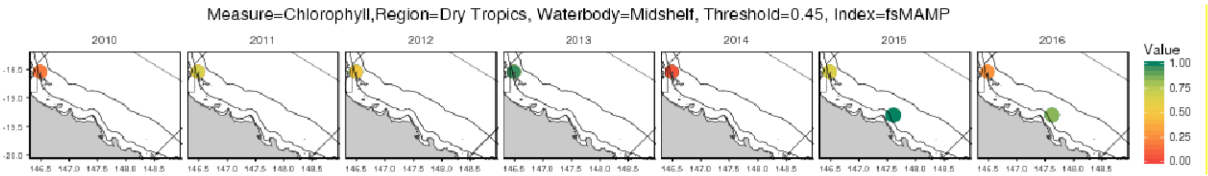
a) AIMS insitu raw data



b) AIMS insitu Binary



c) AIMS insitu fsMAMP



d) AIMS insitu fsMAMP4

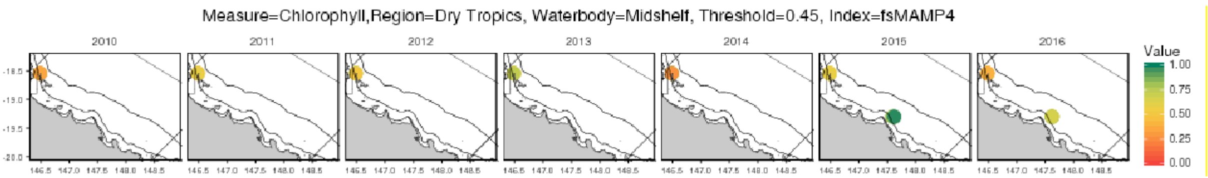
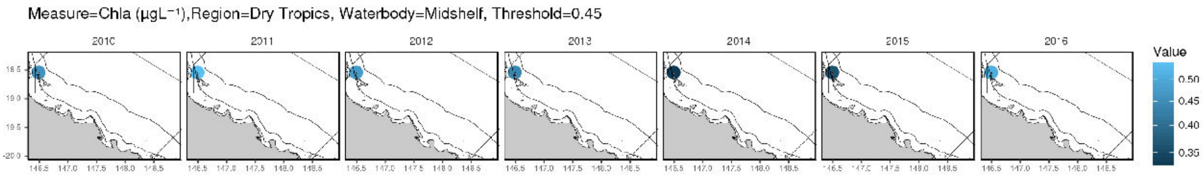
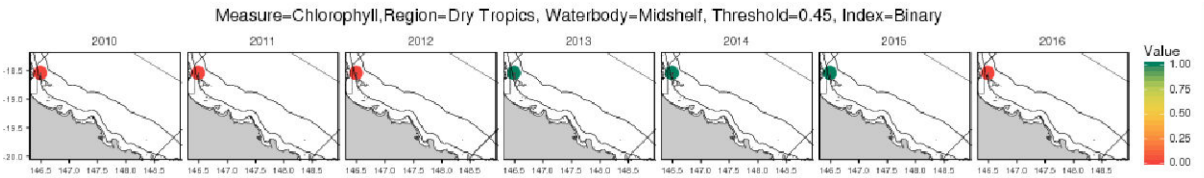


Figure 47: Spatial distribution of AIMS insitu Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

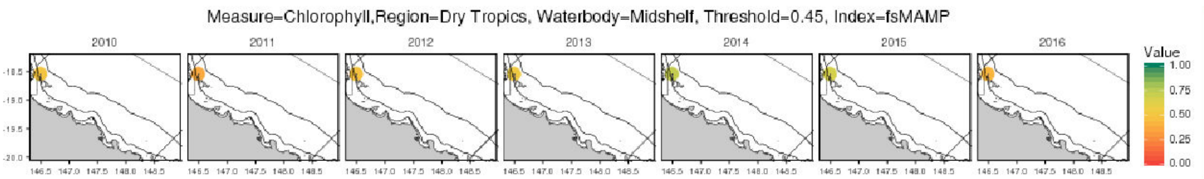
a) AIMS FLNTU raw data



b) AIMS FLNTU Binary



c) AIMS FLNTU fsMAMP



d) AIMS FLNTU fsMAMP4

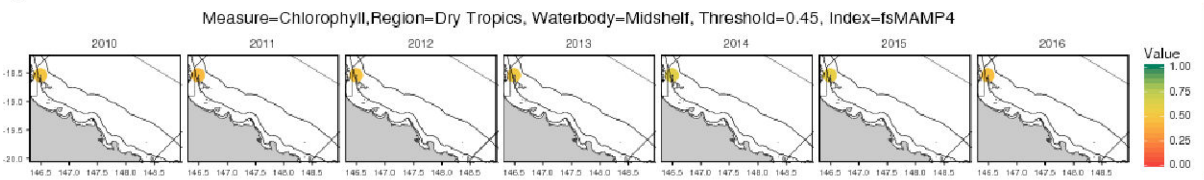


Figure 48: Spatial distribution of AIMS FLNTU Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

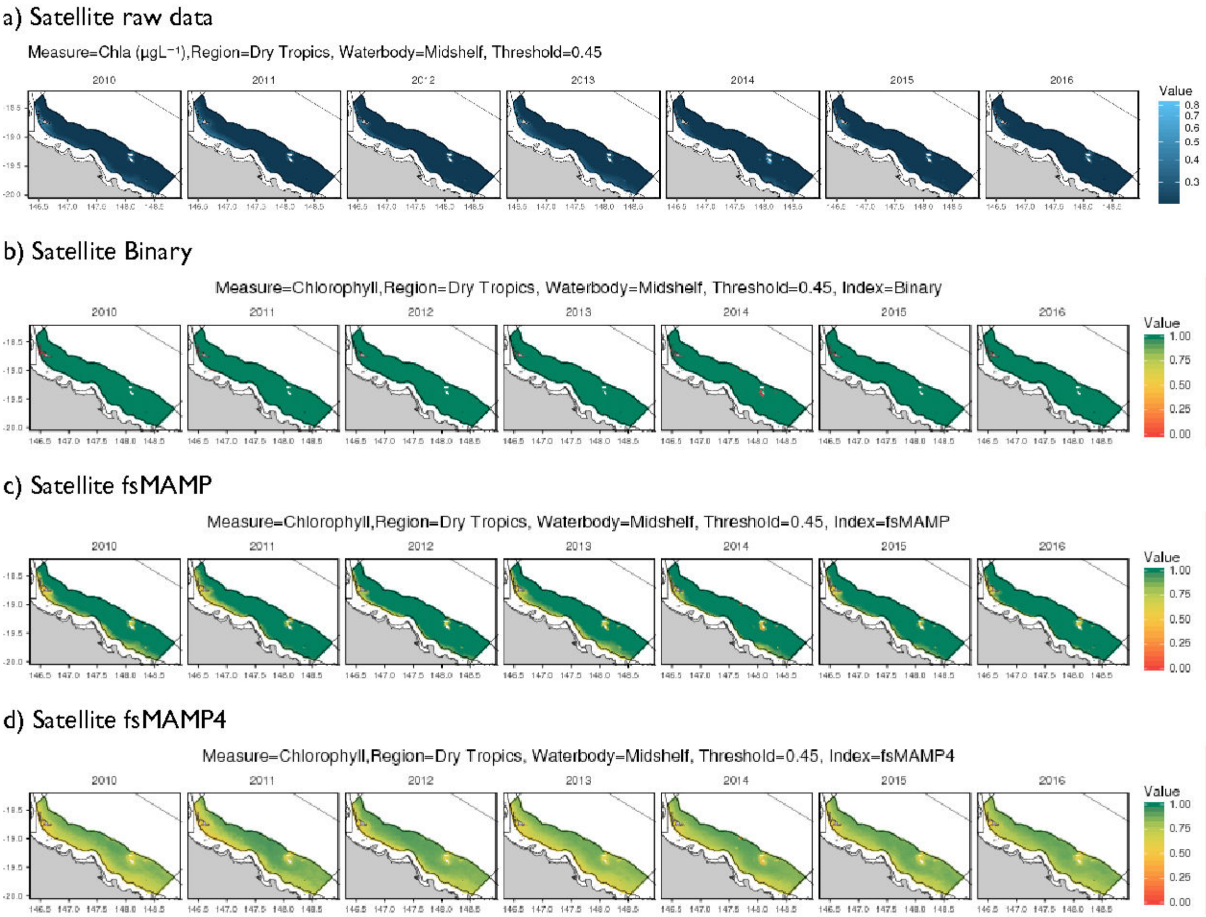
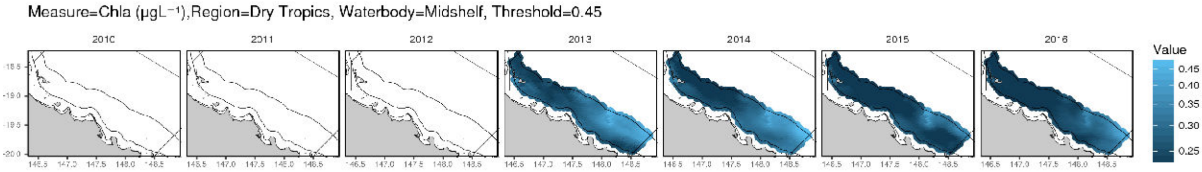
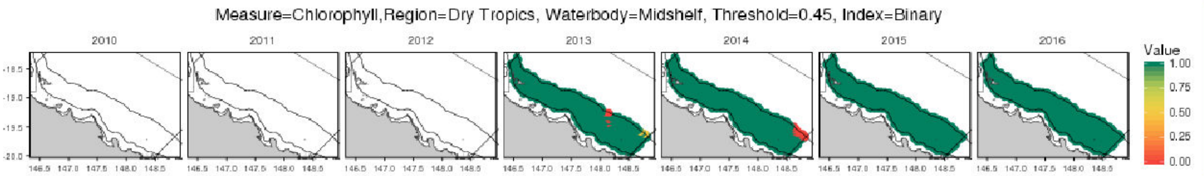


Figure 49: Spatial distribution of Satellite Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

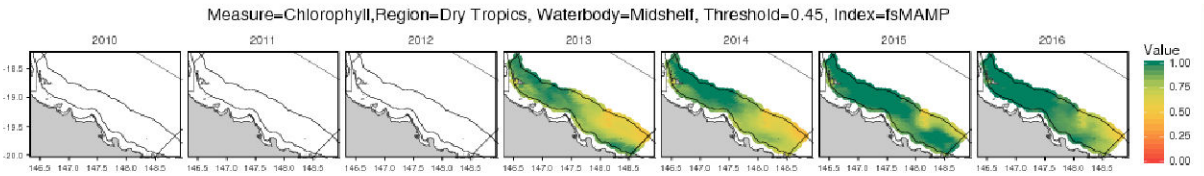
a) eReefs raw data



b) eReefs Binary



c) eReefs fsMAMP



d) eReefs fsMAMP4

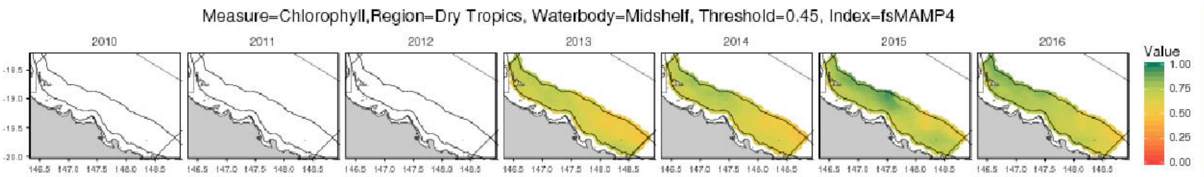
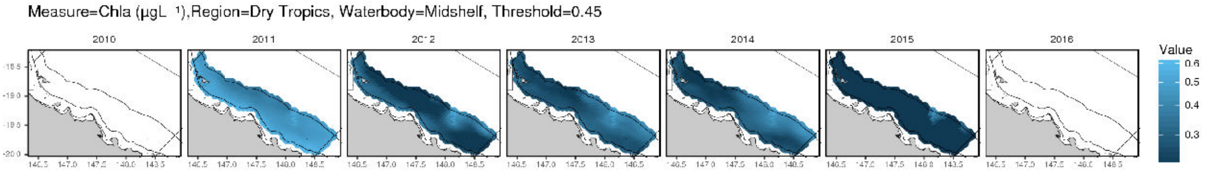
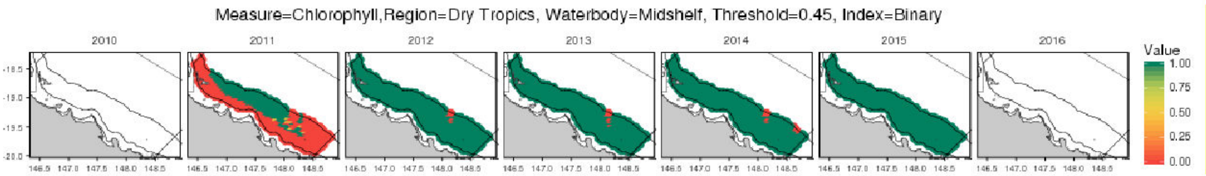


Figure 50: Spatial distribution of eReefs Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

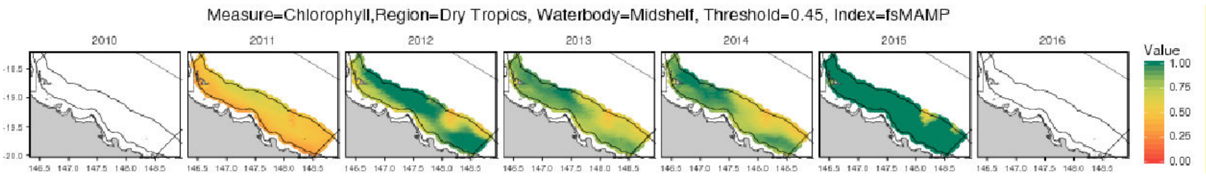
a) eReefs926 raw data



b) eReefs926 Binary



c) eReefs926 fsMAMP



d) eReefs926 fsMAMP4

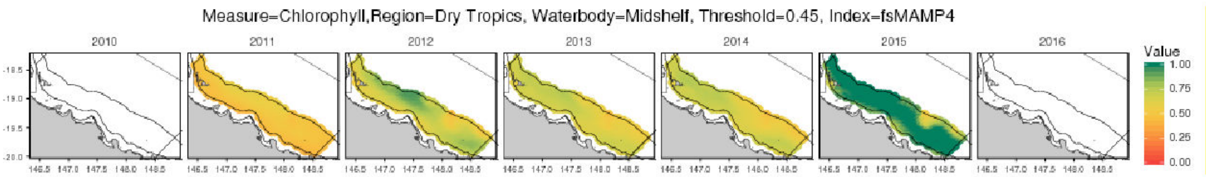


Figure 51: Spatial distribution of eReefs926 Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

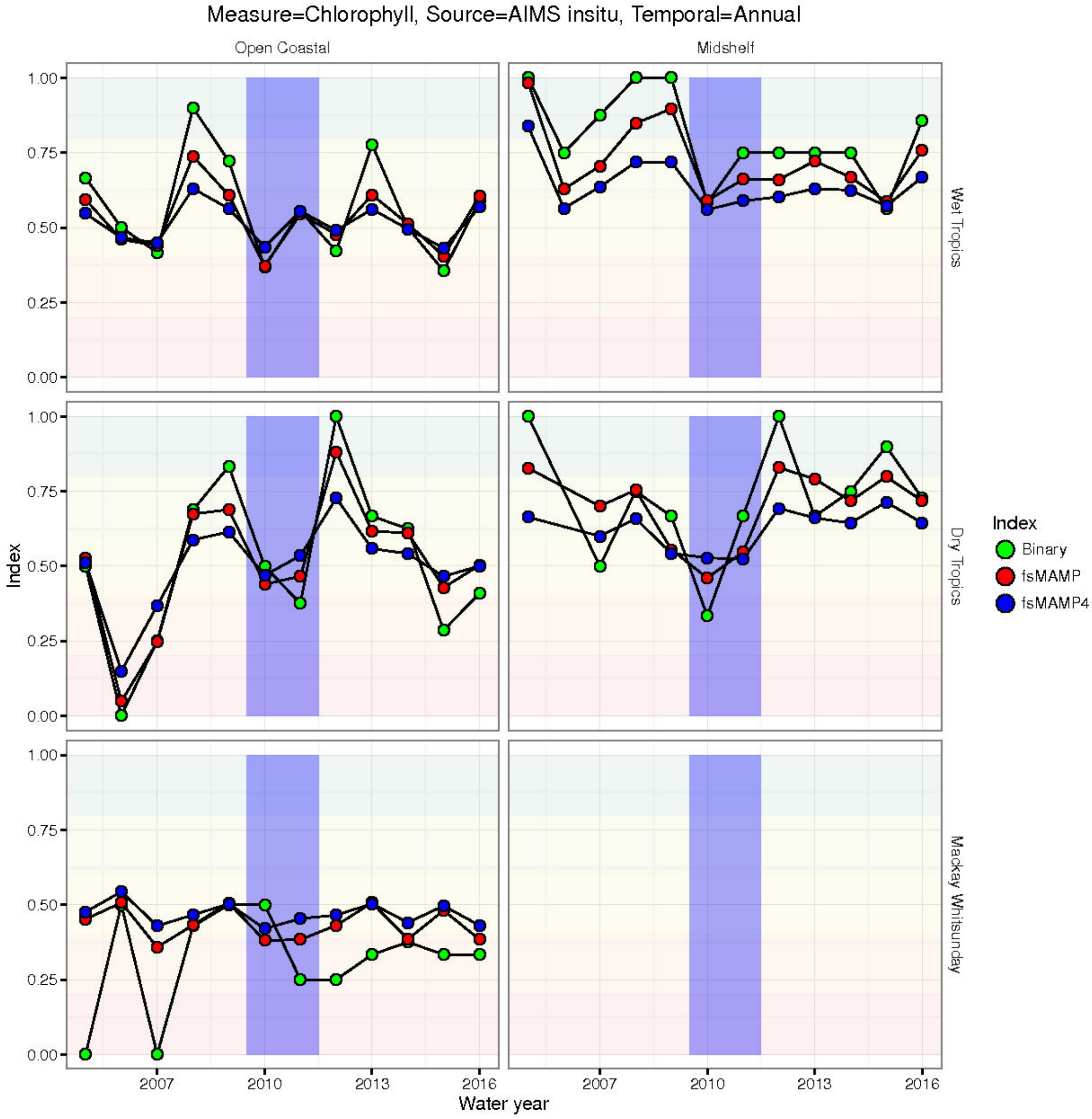


Figure 52: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for AIMS insitu Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

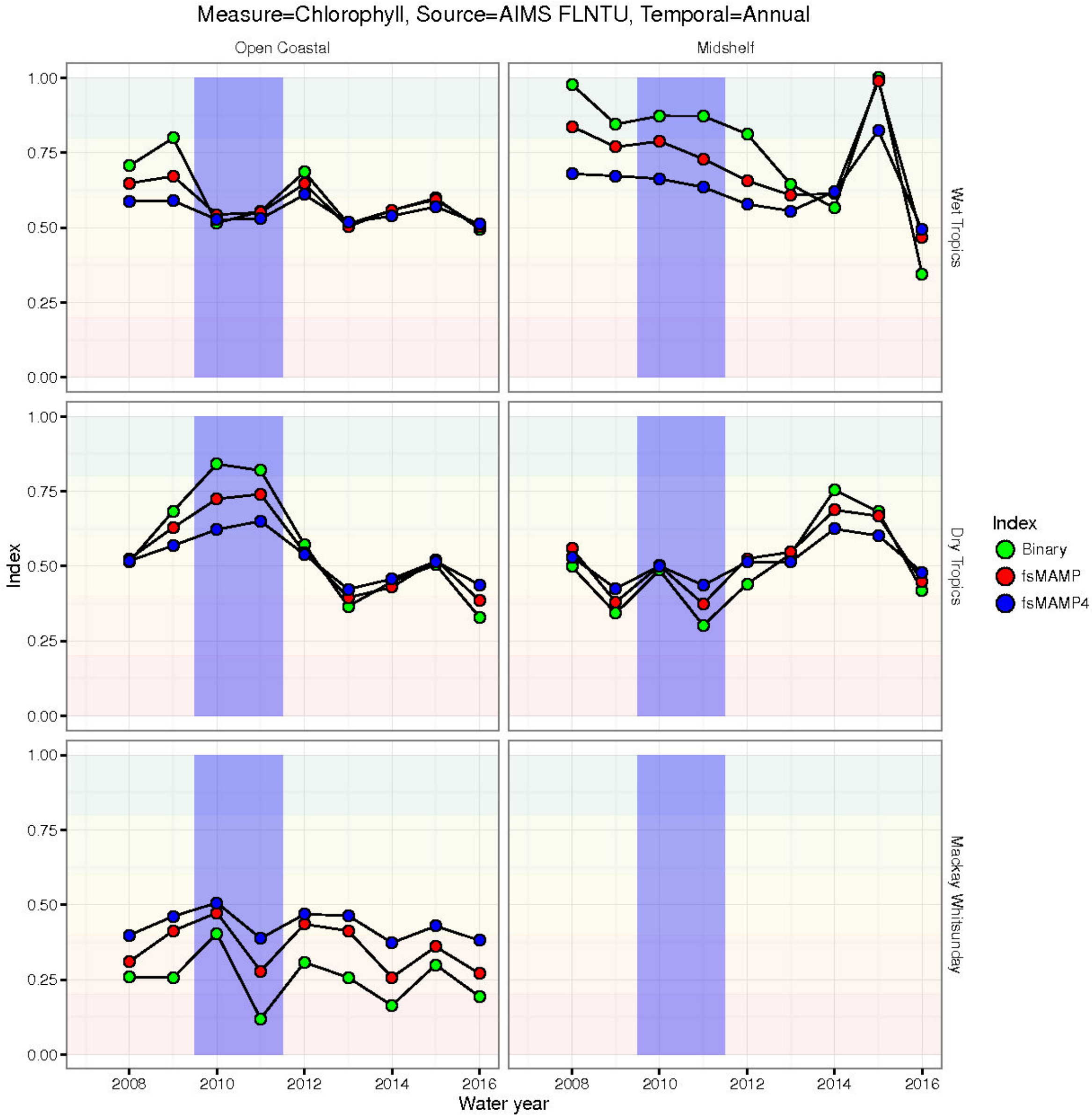


Figure 53: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for AIMS FLNTU Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

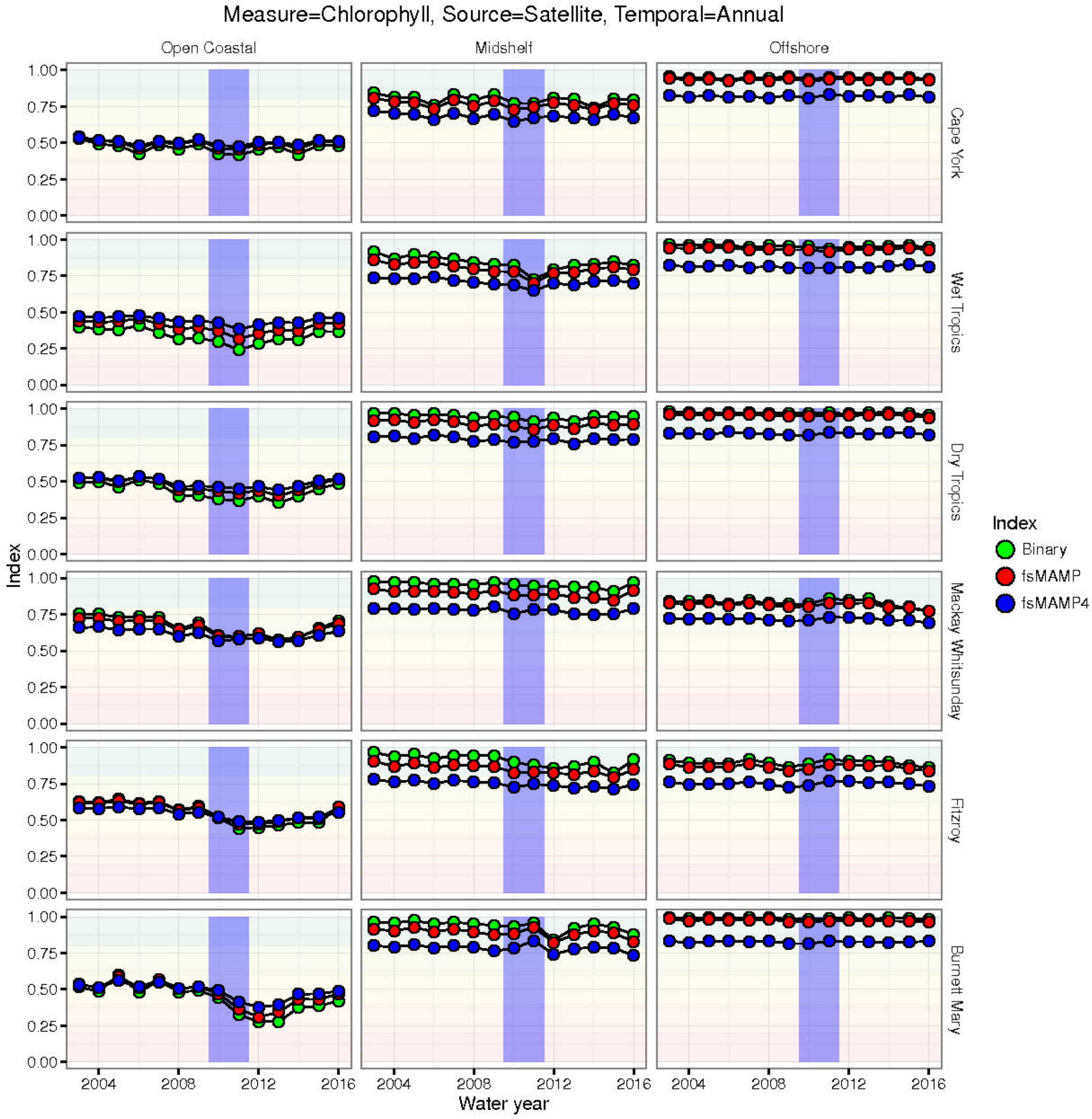


Figure 54: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for Satellite Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

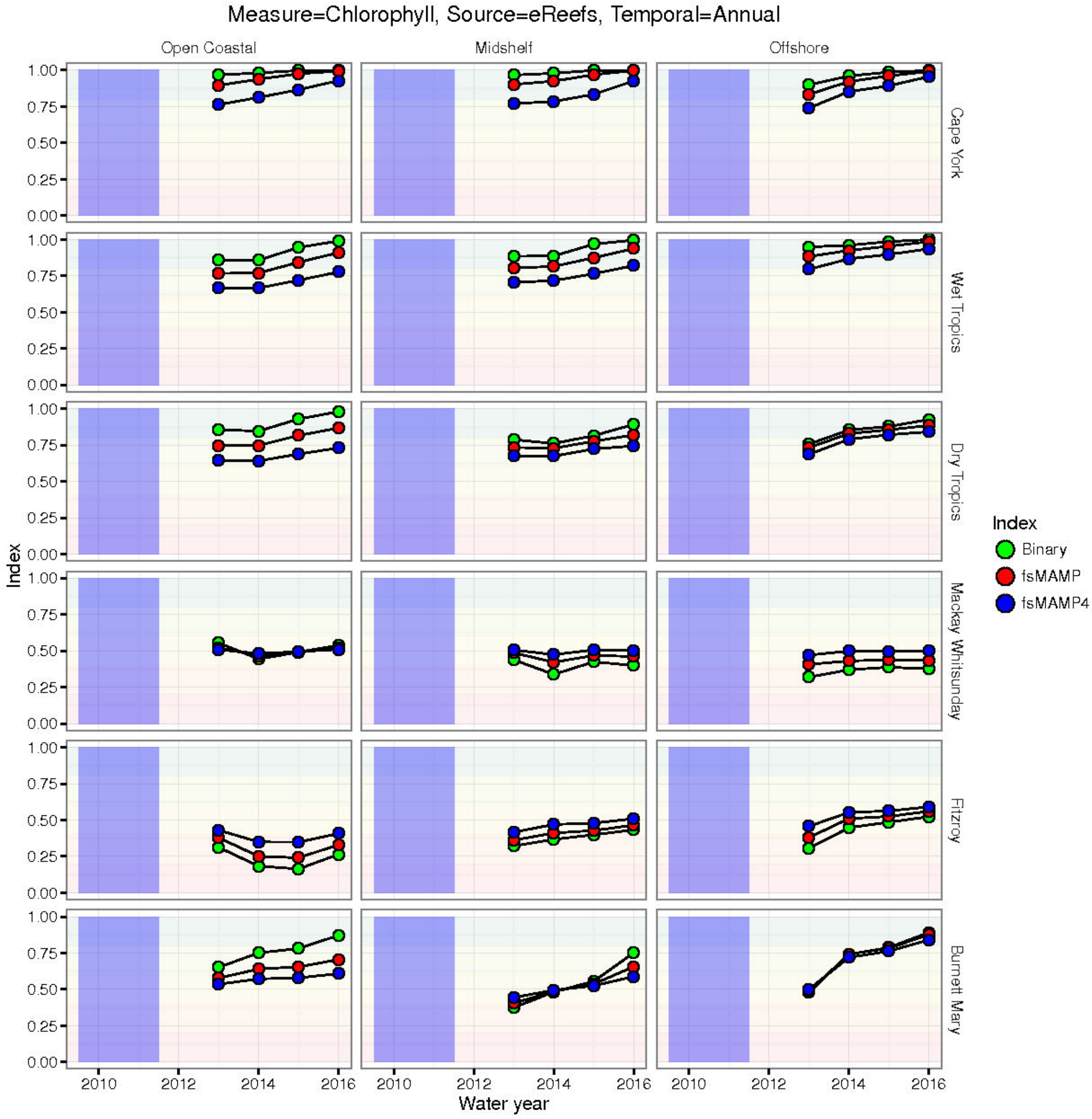


Figure 55: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for eReefs Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

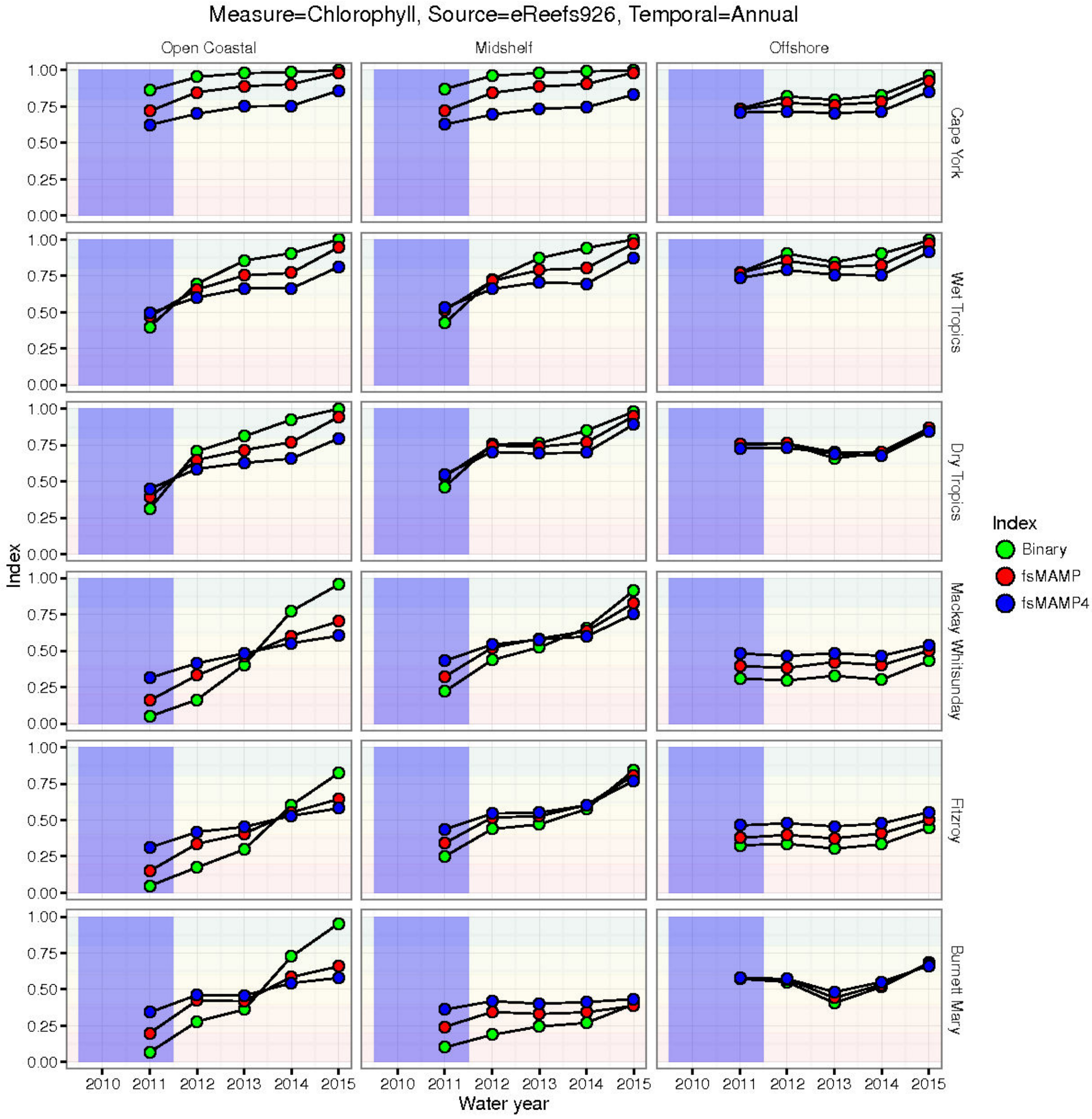


Figure 56: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for eReefs926 Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

5.3.2 Sources

Typically, the major aspects of a property like water quality are not directly measurable. Properties such as productivity, water clarity, nutrients, pesticides etc encapsulate a set of underlying conditions and yet themselves are not directly measurable. Directly measurable properties (such as Chlorophyll-a, total suspended solids etc) thus act as proxies for the more broader properties. As directly measurable entities, many of these measures have long monitoring histories and there are at least some understanding of the ecological role of these measures.

A major advantage of remote sensing and modelling products in the context of environmental monitoring is that they provide substantially greater spatial and temporal coverage. However, the majority of the parameters yielded from these tools are algorithmic approximations of traditional measures. Consequently, in the context of water quality, they produce proxies of proxies.

The current project has access to a variety of sources of water quality monitoring data (see Section **??**) ranging from sparse, yet vigorous direct in situ measurements (AIMS insitu) and temporally rich, spatially sparse AIMS FLNTU logger data through to spatio-temporally extensive, yet patchy remote sensing MODIS Satellite data and multiple versions of eReefs modelled data. These different sources of data are likely to provide estimates of the parameters that differ in both location (such as mean) as well as scale (variability).

Whilst it is beyond the scope of the current project to undertake a full evaluation of the accuracy, robustness and reliability of each of these sources, the indexed data permit us to explore and compare the spatio-temporal patterns of each data source. In particular, we can focus on sensitivity as suggested by variability in spatio-temporal patterns of indices of each data source and whether these patterns are consistent with expert expectations.

It is reasonable to expect that since the AIMS insitu data are the most direct measures, they would be the most accurate of all the sources, however it is also likely that these observations only represent conditions over a very restricted space and time. The AIMS insitu data are predominantly the limited spatial coverage of the AIMS insitu data that limits its utility as input into a water quality metric for the entire Great Barrier Reef.

A motivating inspiration for this project was the perceived insensitivity of the Satellite data source and aspirations to improve the sensitivity of the water quality metric as a whole. It was hoped that the introduction of eReefs modelled data would result in a metric that yields patterns that are more consistent with assumed trends.

Figures 57 – 60 contrast the broad spatial and temporal patterns in aggregated fsMAMP Chlorophyll-a, TSS, Secchi depth and NOx indices. Within a zone (Region/Water body), the Satellite data (Remote sensing) are substantially less varied than the other sources. Obvious deviations in trajectory are only really apparent for the Open Coastal areas (although not for Cape York). Moreover, while the Satellite indices are suggestive of a cross-shelf (West to East) increase in water quality, this mainly occurs between Open Coastal and Midshelf and there is little (if any) consistent South-North water quality increase.

The AIMS insitu data result in the most sensitive metrics. However, the temporal deviances in data (and thus indices) could be exaggerated by the the proximal location of AIMS insitu sites relative to sources of major river discharge. Thus, this sensitivity could be artificially inflated and is unlikely to be unrepresentative. Moreover, the AIMS insitu data are restricted to just a subset (5/18) of the zones of interest.

Surprisingly, there is relatively little correspondence in trajectories between AIMS insitu and AIMS FLNTU logger data. These differences could be due either to differences in sampling designs (AIMS insitu have additional sites and thus represent a different spatial domains, AIMS FLNTU have substantially greater temporal coverage and thus are potentially more representative over time) and could also reflect direct (AIMS insitu) vs indirect (AIMS FLNTU) nature of the measurements. Either way, it is difficult endorse either of these sources as a primary data source on which to construct GBR wide Water Quality metrics.

The broad spatial pattern of both eReefs and eReefs926 appear to follow the overall expectations of South - North and West - East gradients[10], with Chlorophyll-a typically increasing from S to N and W to E - more so for eReefs926 than eReefs. Unfortunately it is difficult to assess the sensitivity of temporal patterns in eReefs and eReefs926 data sources due to their relatively short availability windows. In particular, it is inconvenient that neither eReefs source extend back to the 2010–2011 wet years to provide some form of qualitative calibration.

Underlying alterations in the eReefs biogeochemical model have resulted in some relatively large changes for each of Chlorophyll-a, Secchi depth and NOx and evaluating the causes of these differences is beyond the scope of the current study.

[10]less obvious for TSS and NOx

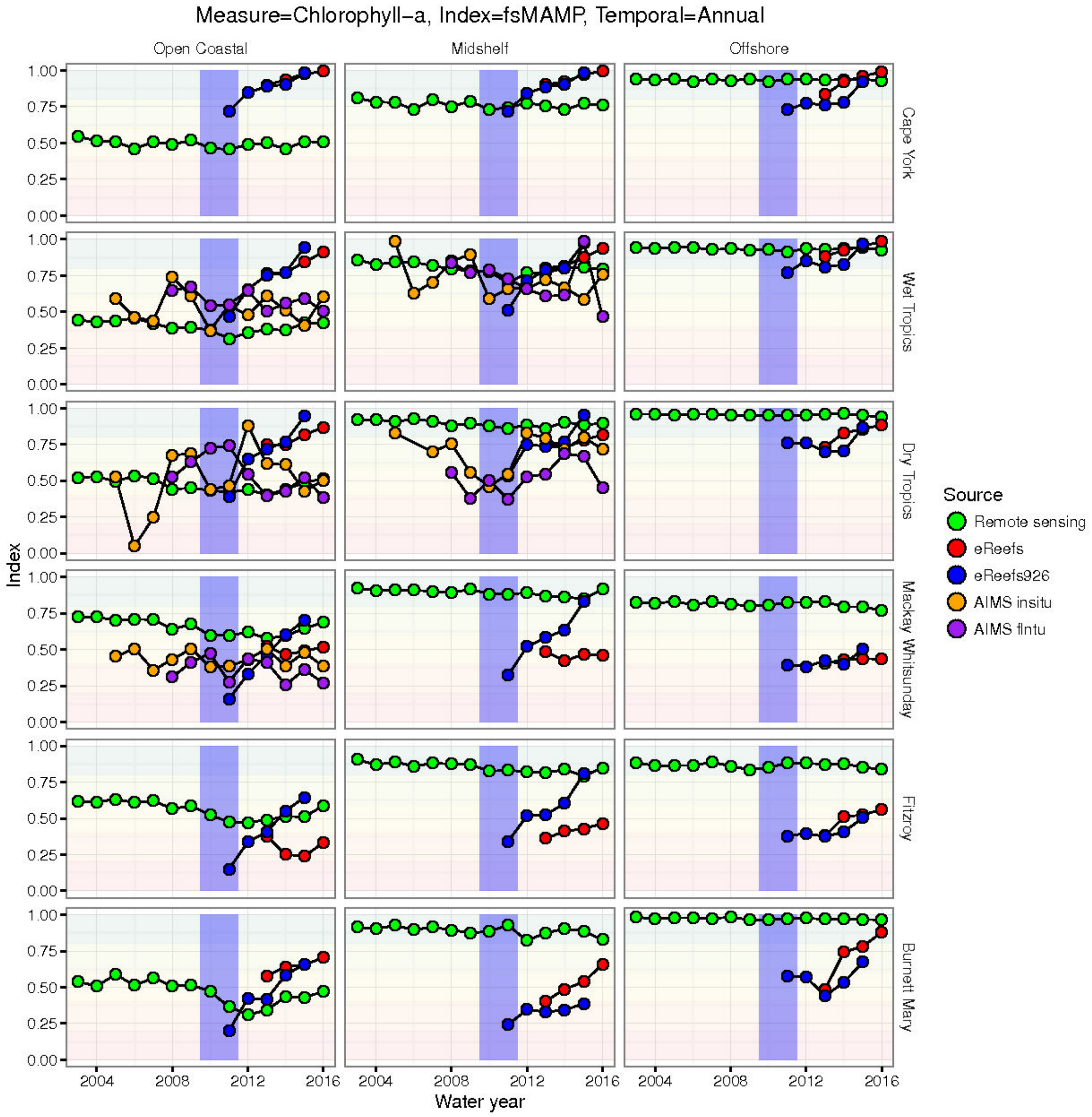


Figure 57: Time series of fsMAMP Chlorophyll-a index scores by zone for each data source. The blue vertical bar spans from mid 2009 to mid 2011.

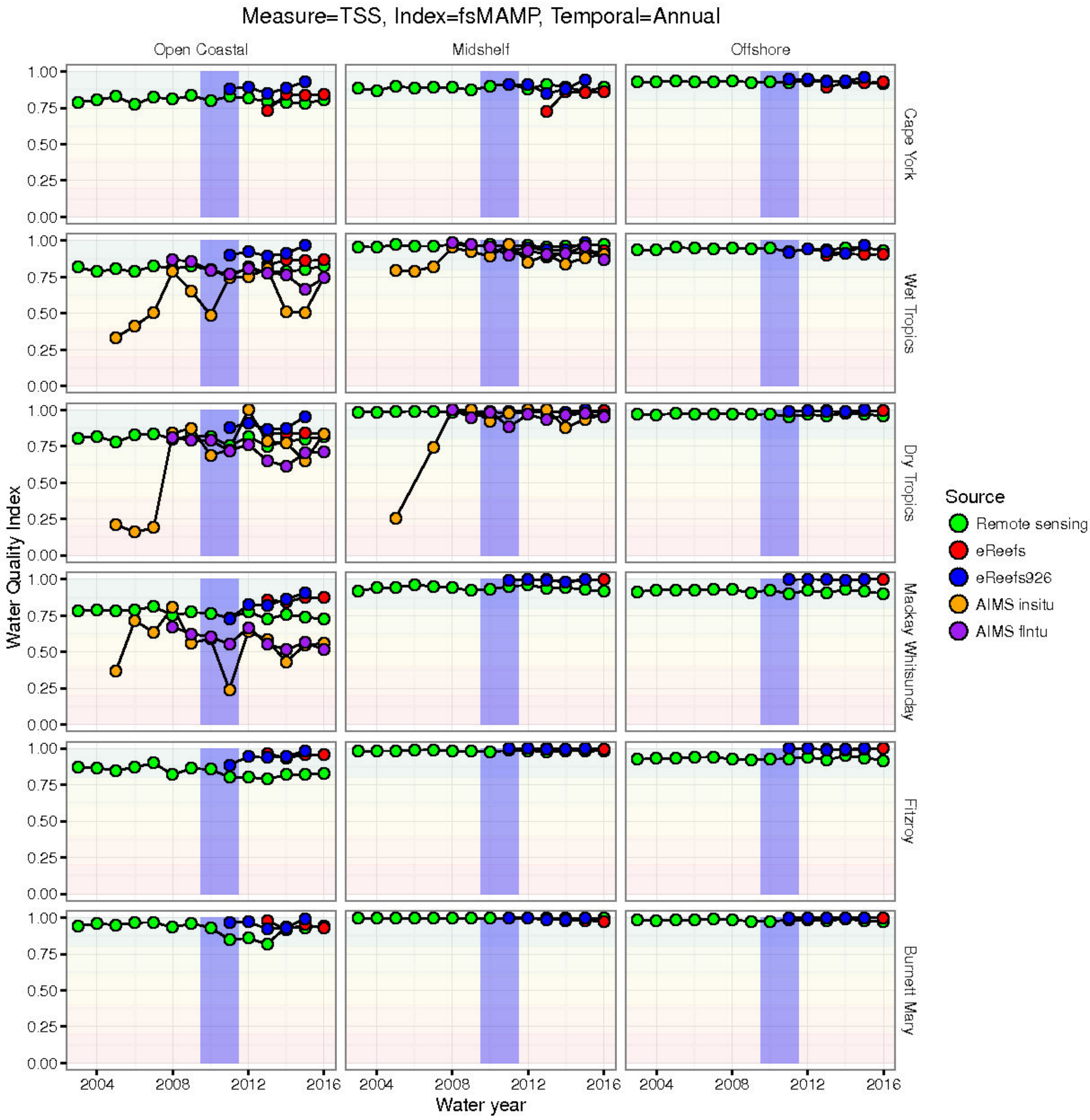


Figure 58: Time series of fsMAMP TSS index scores by zone for each data source. The blue vertical bar spans from mid 2009 to mid 2011.

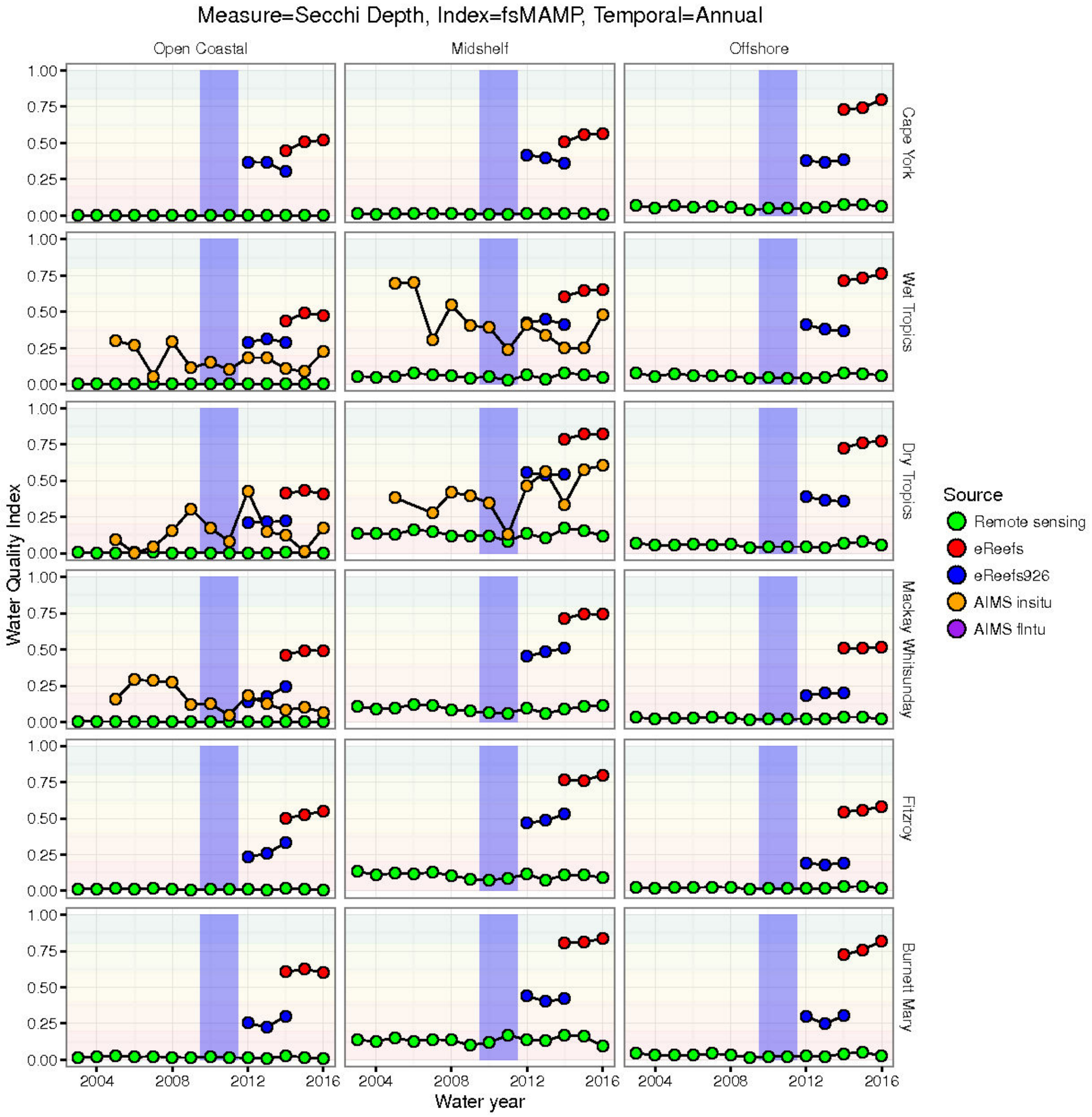


Figure 59: Time series of fsMAMP Secchi depth index scores by zone for each data source. The blue vertical bar spans from mid 2009 to mid 2011.

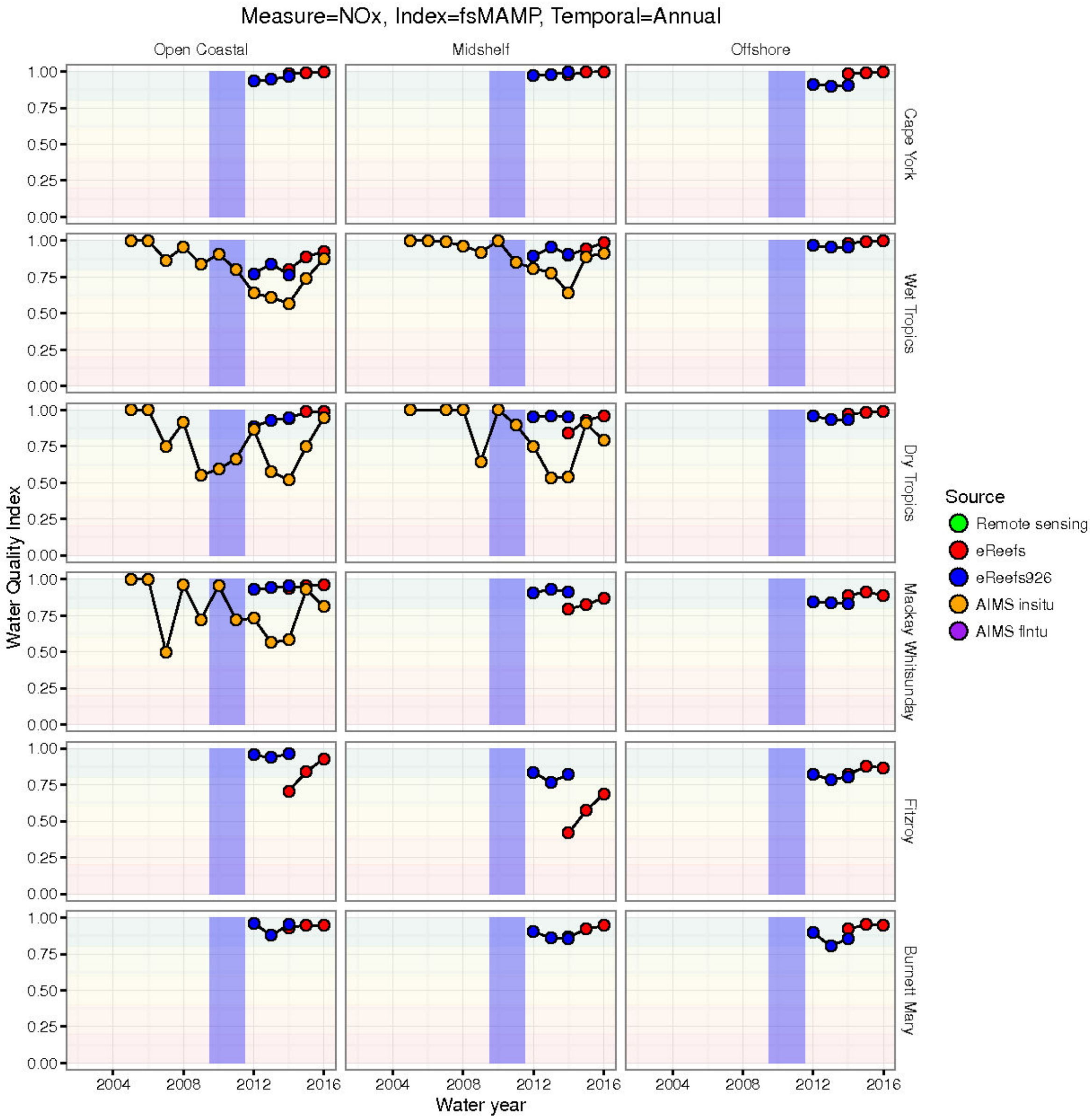


Figure 60: Time series of fsMAMP NOx index scores by zone for each data source. The blue vertical bar spans from mid 2009 to mid 2011.

5.3.3 Exploration of Measures

A Water Quality Index should attempt to reflect multiple properties of the underlying water bodies. For example, Water Quality could be characterized by combinations of Productivity, Water clarity, Nutrients, Toxicants etc. In turn, each of the above Sub-indicators, can be characterized by actual measureable properties (such as Chlorophyll-a, Total Suspended Solids, Total Nitrogen etc).

Typically, a Water Quality index is limited to what measureable properties are available and have appropriate guidelines (thresholds). The spatial extent of the current application of Water Quality metrics limits the Measures to Chlorophyll-a, Total Suspended Solids, Secchi Depth and NOx (Nitrite + Nitrate). Temporal series of the individual Measures for each Zone (based on fsMAMP of eReefs data) are presented in Figure 61.

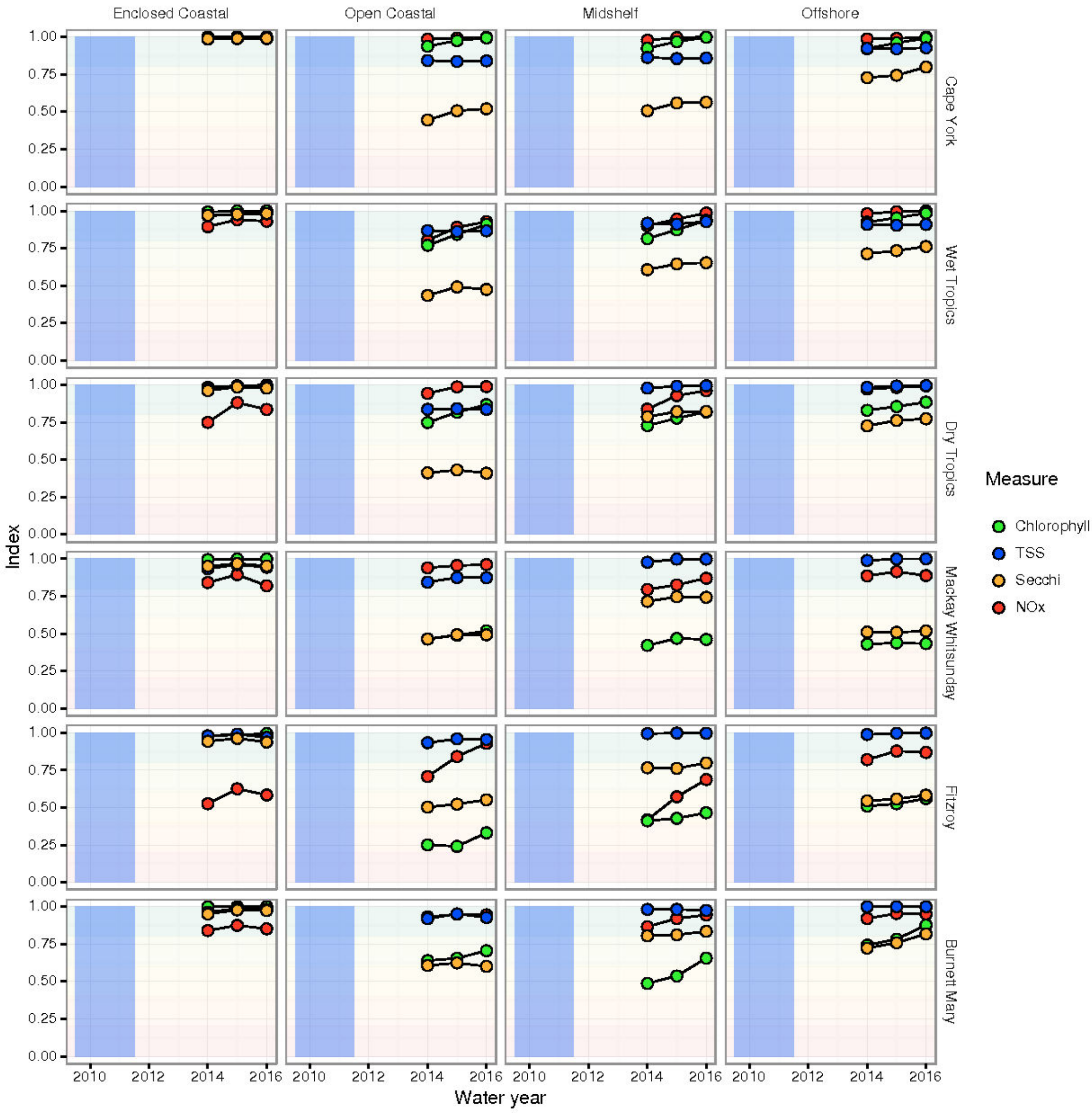


Figure 61: Time series of eReefs fsMAMP index scores by zone. The blue vertical bar spans from mid 2009 to mid 2011.

These four Measures can be placed in a aggregation hierarchy such as depicted in Table 14.

Table 14: Hierarchical association between Measures, Sub-indicators and Indicators.

Measure	Sub-indicator	Indicator
Chlorophyll-a	Productivity	Water Quality
Total Suspended Solids	Water Clarity	Water Quality
Secchi Depth	Water Clarity	Water Quality
NOx	Nutrients	Water Quality

Nevertheless, the reliability and utility of each of these Measures are not necessarily equal. A number of candidate Measure combinations[11] are considered (see below). The contributions of each Measure to the corresponding Water Quality Indicator Scores (based on the hierarchy presented in Table 14) are:

- Chlorophyll-a $(1/3)$, TSS $(1/2 \times 1/3 = 1/6)$, SD $(1/2 \times 1/3 = 1/6)$ and NOx $(1/3)$
- Chlorophyll-a $(1/3)$, TSS $(1/2 \times 1/2 = 1/4)$, SD $(1/2 \times 1/2 = 1/4)$
- Chlorophyll-a $(1/2)$, SD $(1/2)$
- Chlorophyll-a $(1/2)$, TSS $(1/2)$

For each candidates, eReefs data with fsMAMP formulations are presented (see Figure 62).

Water Quality Indicator Scores based on candidate combinations that include either all of Chl, TSS, SD and NOx or just Chl and TSS are considered very similar. Generally, Water Quality Indicator Scores are substantially lowered by the inclusion of Secchi Depth, the severity of which depends on the degree of dilution by other Measures.

[11] These effectively act as weights

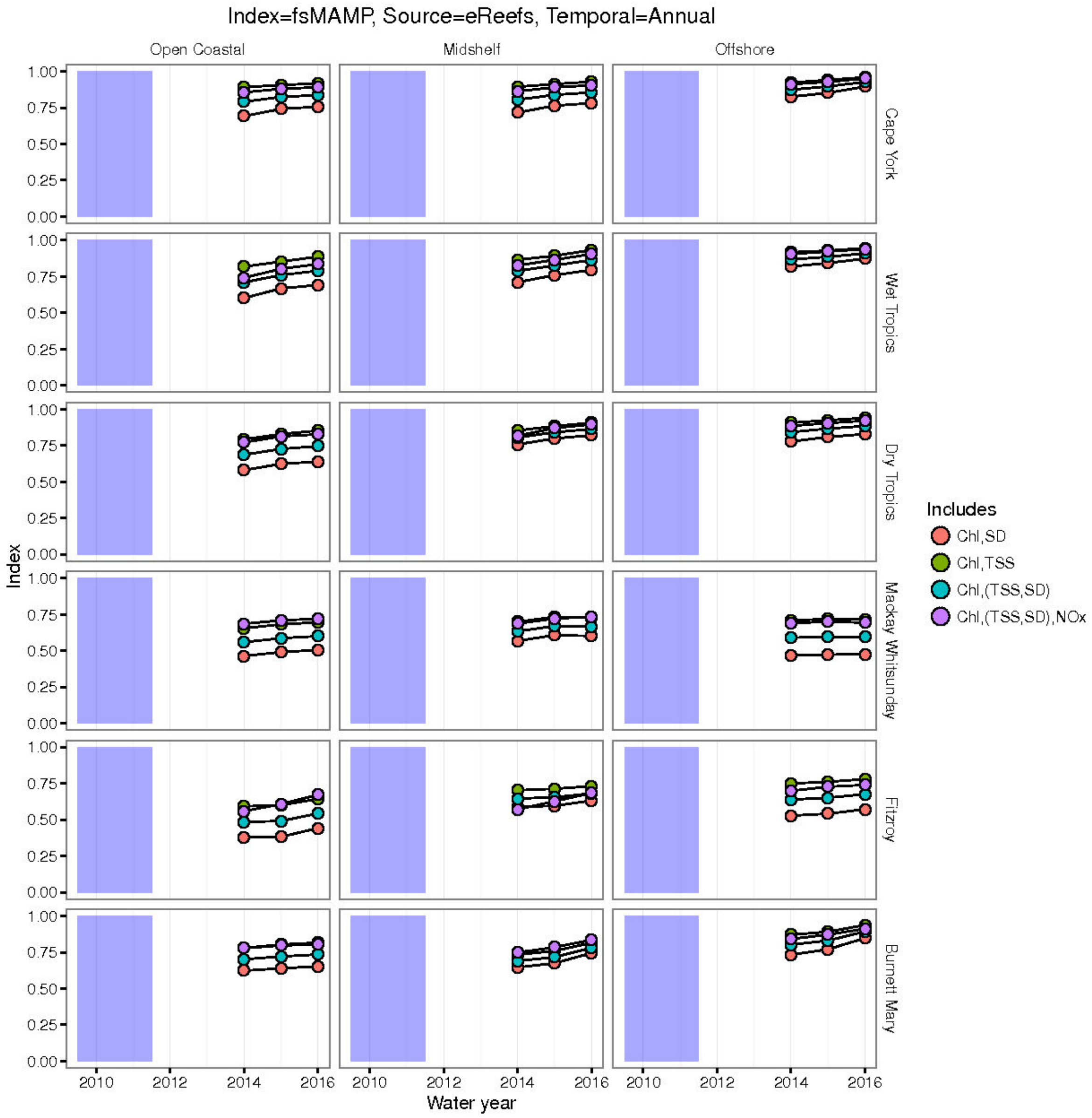


Figure 62: Time series of eReefs fsMAMP Measure Index Scores by zone. The blue vertical bar spans from mid 2009 to mid 2011.

5.3.4 Measure/Site

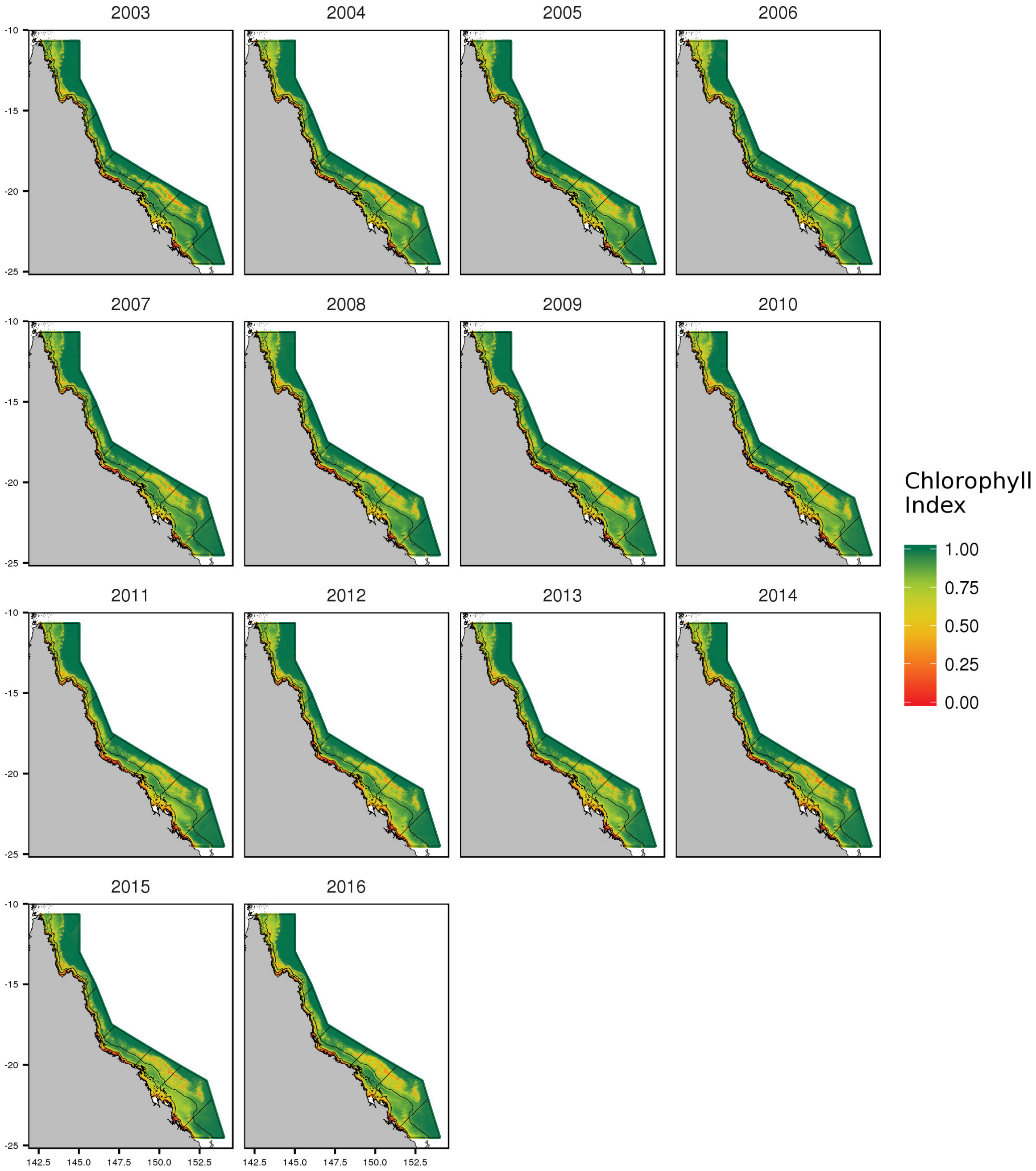


Figure 63: Spatio-temporal Satellite fsMAMP Chlorophyll-a index scores.

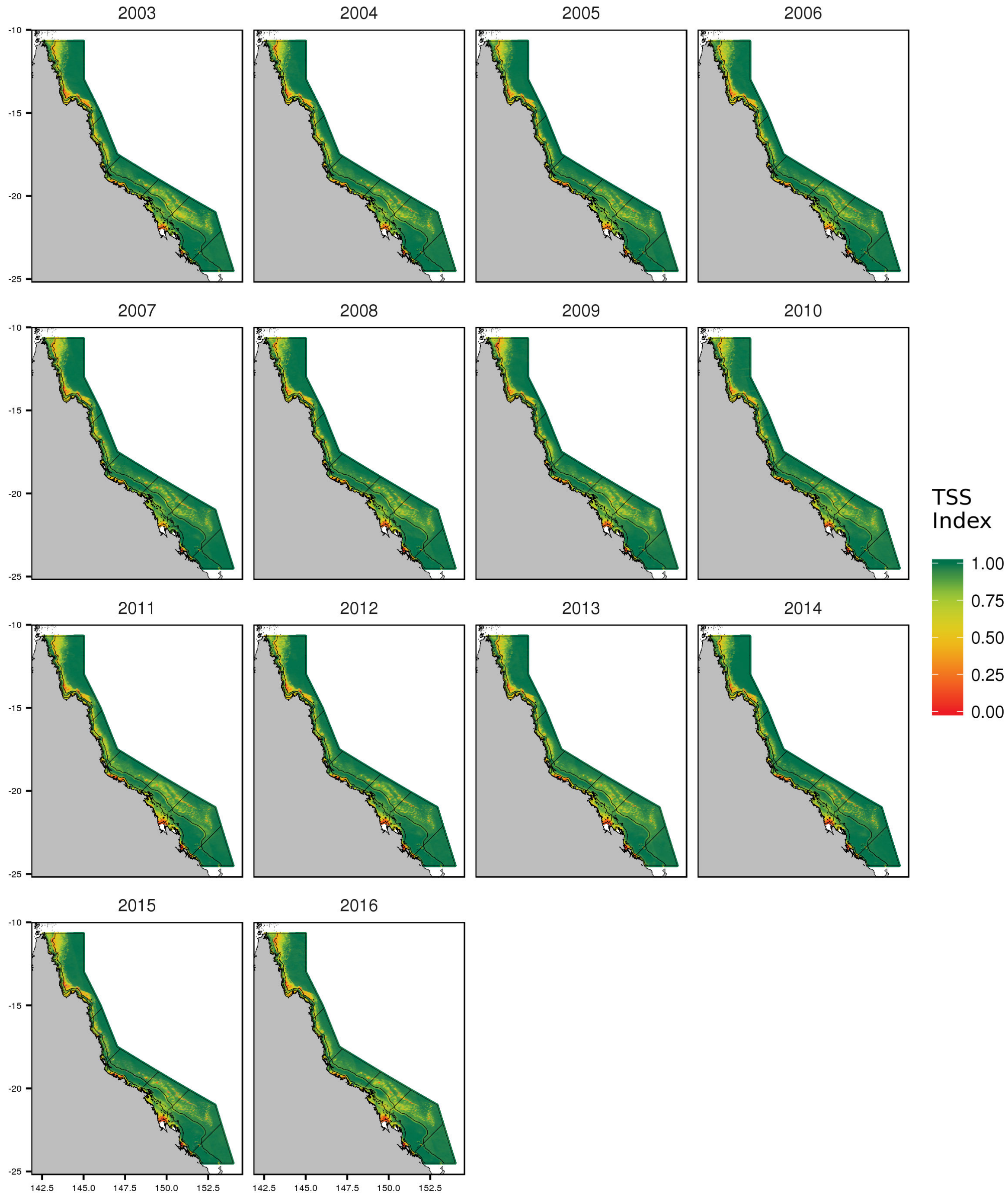


Figure 64: Spatio-temporal Satellite fsMAMP TSS index scores.

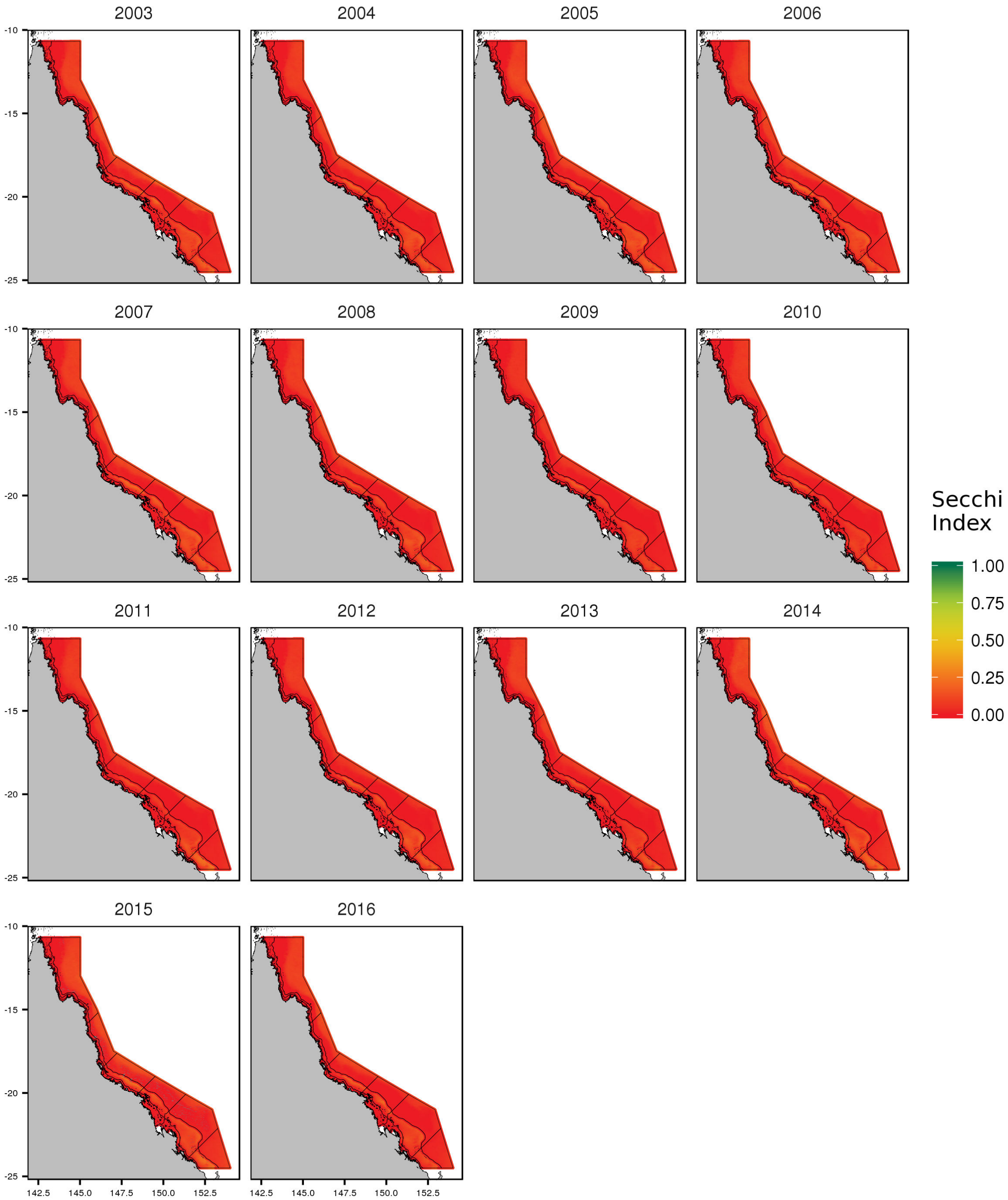


Figure 65: Spatio-temporal Satellite fsMAMP Secchi depth index scores.

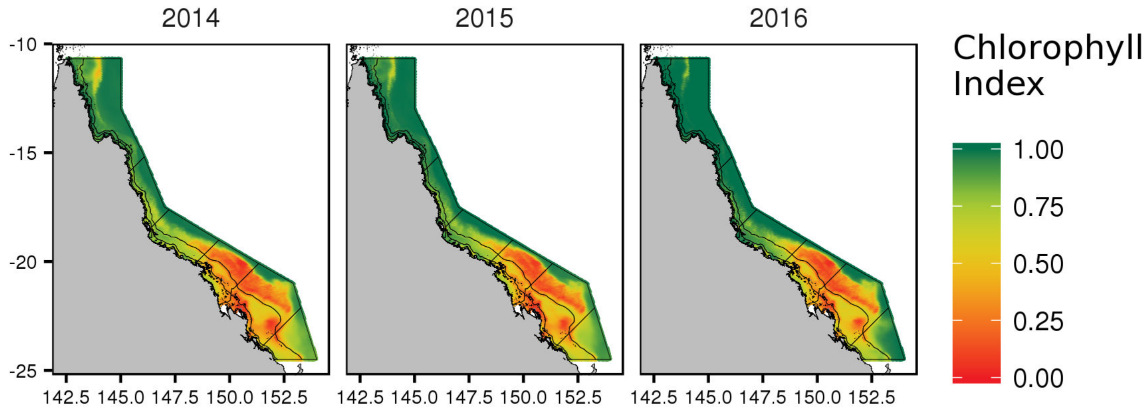


Figure 66: Spatio-temporal eReefs fsMAMP Chlorophyll-a index scores.

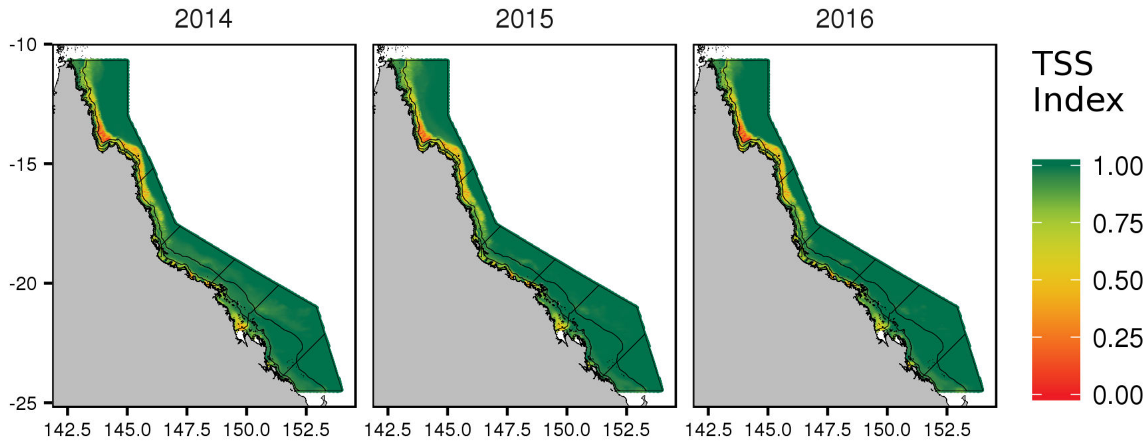


Figure 67: Spatio-temporal eReefs fsMAMP TSS index scores.

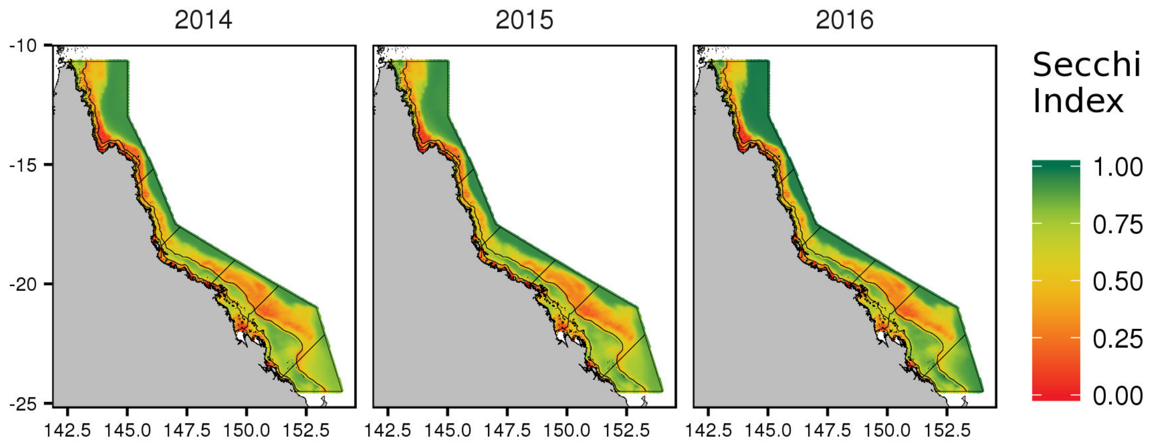


Figure 68: Spatio-temporal eReefs fsMAMP Secchi depth index scores.

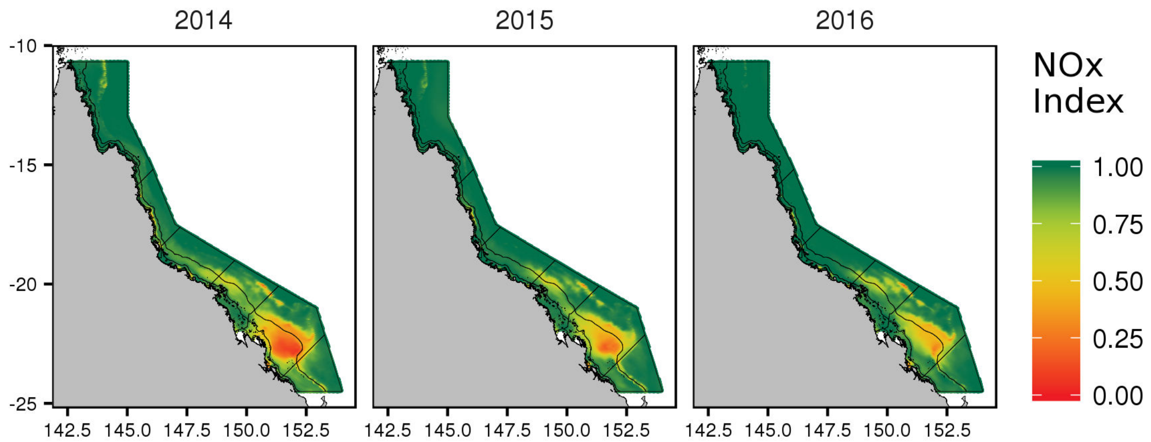


Figure 69: Spatio-temporal eReefs fsMAMP NOx index scores.

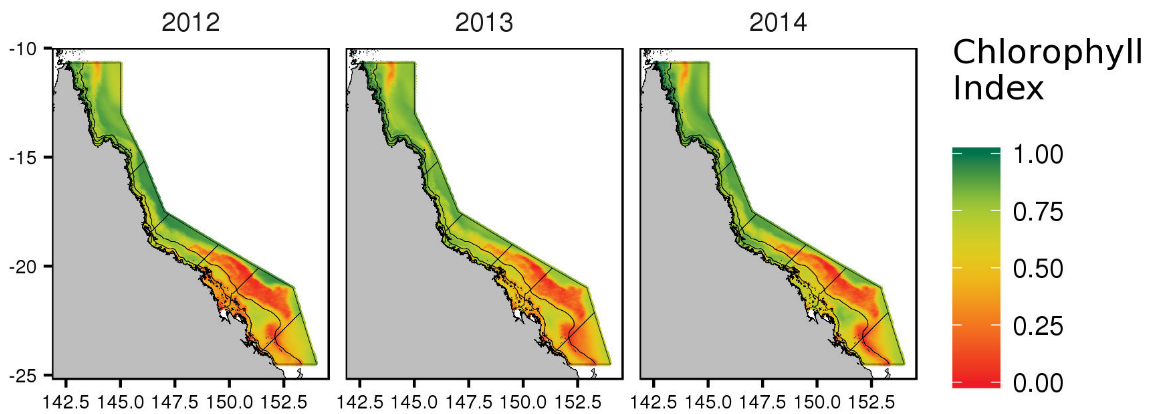


Figure 70: Spatio-temporal eReefs926 fsMAMP Chlorophyll-a index scores.

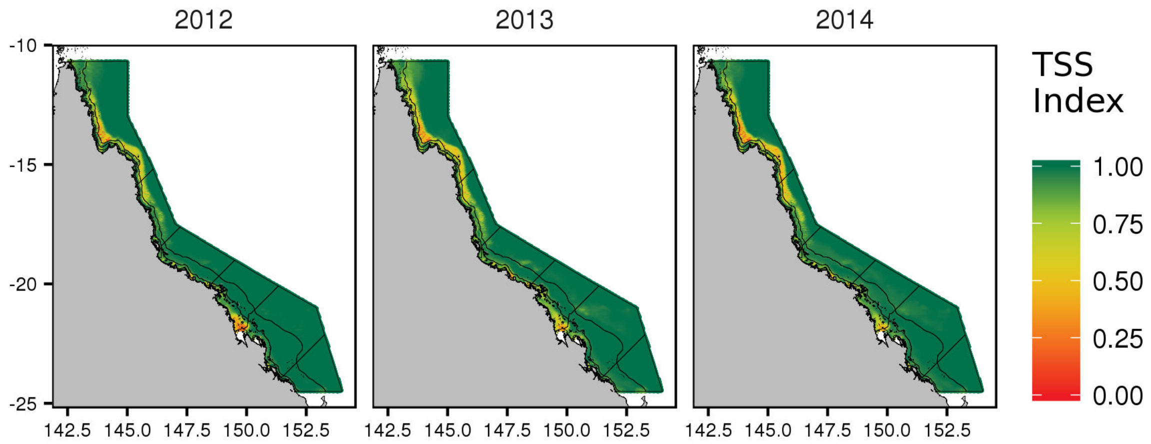


Figure 71: Spatio-temporal eReefs926 fsMAMP TSS index scores.

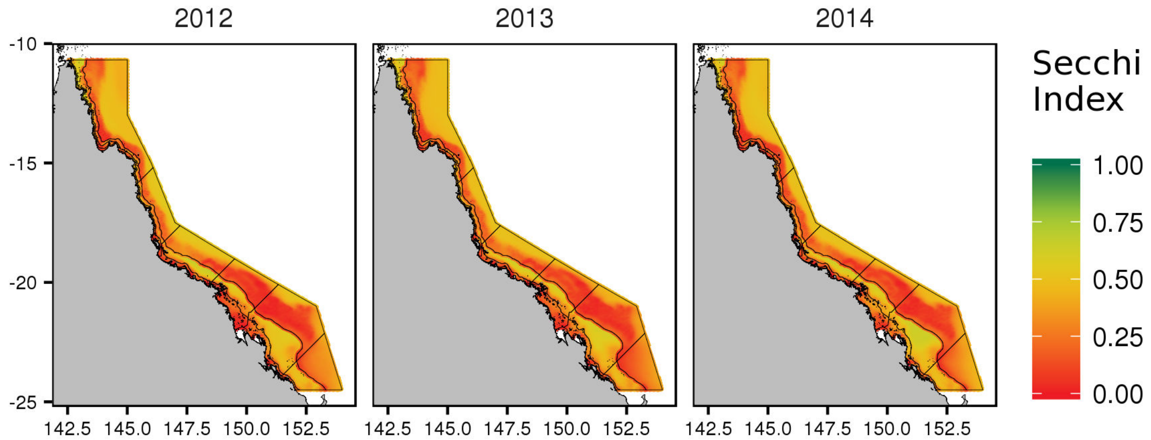


Figure 72: Spatio-temporal eReefs926 fsMAMP Secchi depth index scores.

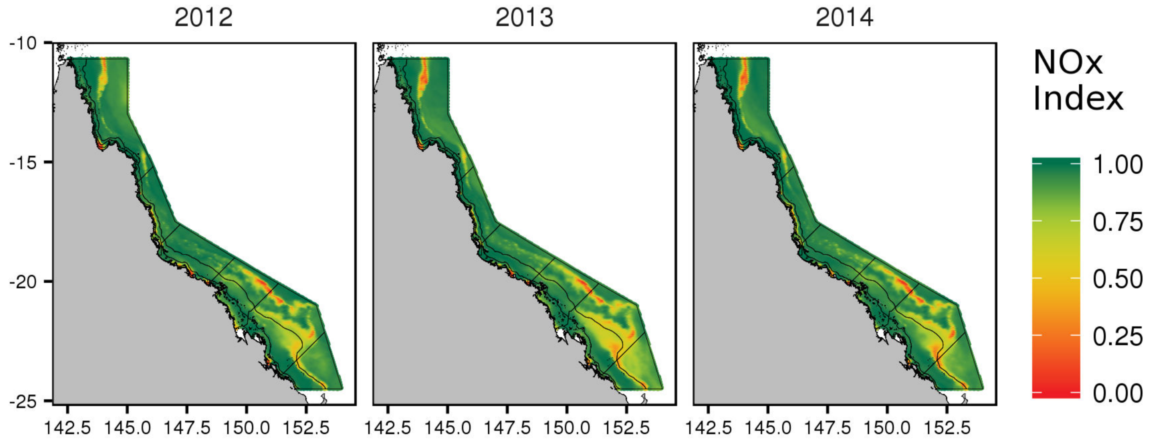


Figure 73: Spatio-temporal eReefs926 fsMAMP NOx index scores.

5.4 Summary of recommendations