Australian Government | Australian Institute of Marine Science

NESP.3.2.5

NESP.3.2.5

Author: Cedric Robillot, Britta Shaffelke, Murray Logan, Mark Baird, Katharine Martin

AIMS: Australia's tropical marine research agency

November 13, 2017

Australian Institute of Marine Science

PMB No 3 | PO Box 41775 | The UWA Oceans Institute (M096)
Townsville MC QLD 4810 | Casuarina NT 0811 | Crawley WA 6009

This report should be cited as:

Enquires should be directed to:

Murray Logan
m.logan@aims.gov.au

© Copyright: Australian Institute of Marine Science (AIMS) 2017

All rights are reserved and no part of this document may be reproduced, stored or copied in any form or by any means whatsoever except with the prior written permission of AIMS

DISCLAIMER

While reasonable efforts have been made to ensure that the contents of this document are factually correct, AIMS does not make any representation or give any warranty regarding the accuracy, completeness, currency or suitability for any particular purpose of the information or statements contained in this document. To the extent permitted by law AIMS shall not be liable for any loss, damage, cost or expense that may be occasioned directly or indirectly through the use of or reliance on the contents of this document.

Revision History

Version		Title	Name	Date	Comments
1	Author	Dr	Murray Logan	November 13, 2017	
	Approved by				
2	Author				
	Approved by				
3	Author				
	Approved by				
4	Author				
	Approved by				

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. EXECUTIVE SUMMARY

2. INTRODUCTION

3. DATA SOURCES

Report cards are typically compiled and communicated annually. However, the time window that constitutes a year differs from report card to report card. Many environmental report cards communicate on data collected within a financial year. This schedule provides a reporting window that is consistent with other management and governmental considerations. Others use a time window that naturally aligns with the cycle of some major underlying environmental gradient - such as wet/dry season. For this project, we are adopting using the same water year (1st Oct – 31 Sept) definition as the AIMS inshore Water Quality Marine Monitoring Program (Lønborg et al., 2016).

The Great Barrier Reef Marine Park (GBR)spans nearly 14° of latitude and covers approximately 344,400km^2.

- spanning multiple jurisdictions/pressures as well as distance offshore - more useful to partition the GBR into smaller more homogeneous zones representing combinations of region and water body. - Six regions (Cape York, Wet Tropics, Dry Tropics, Mackay Whitsunday, Fitzroy and Burnett Mary) - Four water bodies (Enclosed Coastal, Open Coastal, Midshelf and Offshore) - define each...

Table 1: Great Barrier Reef spatial Zones and associated Regions and Water bodies.

GBRMPA Zone	Zone	Region	Water body
Enclosed_Coastal_Cape_York	Enclosed_Coastal_Cape York	Cape York	Enclosed Coastal
Enclosed_Coastal_Terrain_NRM	Enclosed_Coastal_Wet Tropics	Wet Tropics	Enclosed Coastal
Enclosed_Coastal_Burdekin_Dry_Tropics_NRM	Enclosed_Coastal_Dry Tropics	Dry Tropics	Enclosed Coastal
Enclosed_Coastal_Mackay_Whitsunday_NRM_Group	Enclosed_Coastal_Mackay Whitsunday	Mackay Whitsunday	Enclosed Coastal
Enclosed_Coastal_Fitzroy_Basin_Association	Enclosed_Coastal_Fitzroy	Fitzroy	Enclosed Coastal
Enclosed_Coastal_Burnett_Mary_Regional_Group_for_NRM	Enclosed_Coastal_Burnett Mary	Burnett Mary	Enclosed Coastal
Open_Coastal_Cape_York	Open_Coastal_Cape York	Cape York	Open Coastal
Open_Coastal_Terrain_NRM	Open_Coastal_Wet Tropics	Wet Tropics	Open Coastal
Open_Coastal_Burdekin_Dry_Tropics_NRM	Open_Coastal_Dry Tropics	Dry Tropics	Open Coastal
Open_Coastal_Mackay_Whitsunday_NRM_Group	Open_Coastal_Mackay Whitsunday	Mackay Whitsunday	Open Coastal
Open_Coastal_Fitzroy_Basin_Association	Open_Coastal_Fitzroy	Fitzroy	Open Coastal
Open_Coastal_Burnett_Mary_Regional_Group_for_NRM	Open_Coastal_Burnett Mary	Burnett Mary	Open Coastal
Midshelf_Cape_York	Midshelf_Cape York	Cape York	Midshelf
Midshelf_Terrain_NRM	Midshelf_Wet Tropics	Wet Tropics	Midshelf
Midshelf_Burdekin_Dry_Tropics_NRM	Midshelf_Dry Tropics	Dry Tropics	Midshelf
Midshelf_Mackay_Whitsunday_NRM_Group	Midshelf_Mackay Whitsunday	Mackay Whitsunday	Midshelf
Midshelf_Fitzroy_Basin_Association	Midshelf_Fitzroy	Fitzroy	Midshelf
Midshelf_Burnett_Mary_Regional_Group_for_NRM	Midshelf_Burnett Mary	Burnett Mary	Midshelf
Offshore_Cape_York	Offshore_Cape York	Cape York	Offshore
Offshore_Terrain_NRM	Offshore_Wet Tropics	Wet Tropics	Offshore
Offshore_Burdekin_Dry_Tropics_NRM	Offshore_Dry Tropics	Dry Tropics	Offshore
Offshore_Mackay_Whitsunday_NRM_Group	Offshore_Mackay Whitsunday	Mackay Whitsunday	Offshore
Offshore_Fitzroy_Basin_Association	Offshore_Fitzroy	Fitzroy	Offshore
Offshore_Burnett_Mary_Regional_Group_for_NRM	Offshore_Burnett Mary	Burnett Mary	Offshore

Table 2: Summary of used data sources.

Source	Custodian	Description
AIMS Insitu	AIMS	AIMS inshore monitoring program Niskin data
AIMS FLNTU	AIMS	AIMS inshore monitoring program FLNTU logger data
Satellite	BOM	BOM: Catalog http://ereeftds.bom.gov.au/ereefs/tds/catalog/ereef/mwq/P1D/2002/catalog.html
eReefs	eReefs	provide a description in ../parameters/sources.csv
eReefs926	eReefs	eReefs: http://dapds00.nci.org.au/thredds/catalog/fx3/gbr4_bgc_926/catalog.html

3.1 Indicators

One of the biggest challenges of report card development is the selection of appropriate indicators from amongst a potentially very large candidate pool. Since the outcomes, conclusions and implications are all dependent on the indicators selected, the selection process is one of the most influential steps and has justifiably received a great deal of attention.

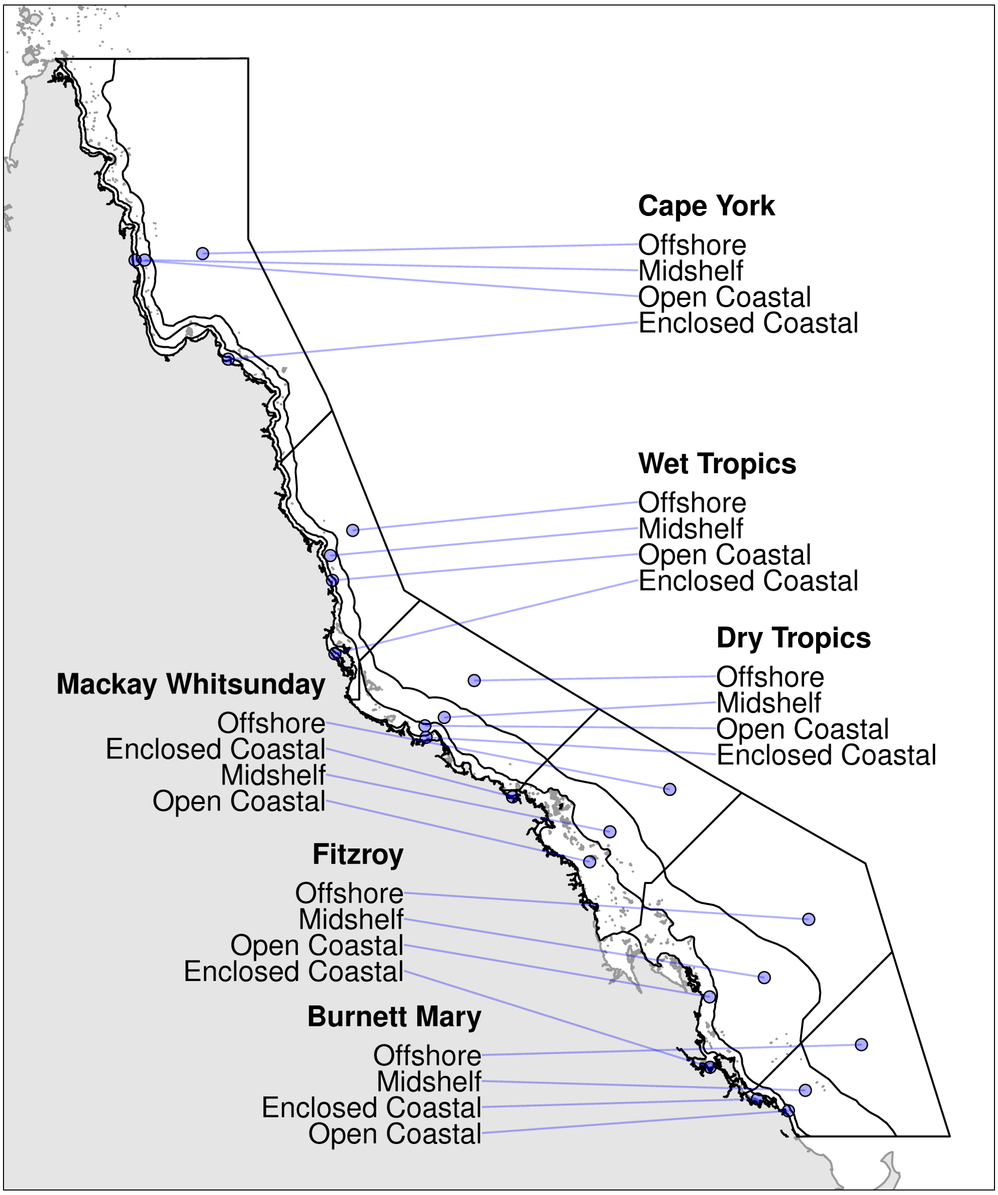


Figure 1: Great Barrier Reef Zones (Regions and Water Bodies).

As part of their ecosystem report card framework, Harwell et al. (1999) urged that the alignment of scientific information with societal goals and objectives should be the guiding principal of indicator selection. In their framework, clearly articulated societal goals and objectives (a combination of societal values and scientific knowledge, such as restored and sustainable wetland system) are translated into Essential Ecosystem Characteristics (EECs) that represent a set of generic attributes that further refine the broad goals (such as water quality, sediment quality, habitat quality, ecological processes). The EEC's are then further translated into a set of scientific informed indicators that are measured or monitored to indicate the status of trends or states associated with the EEC's.

There have since been numerous studies that have focused on providing more formal, objective criterion for indicator selection (Dauvin et al., 2008; Emerson et al., 2012; Flint et al., 2012; James et al., 2012). Whilst the specifics vary, most can be broadly encapsulated by a Dauvin et al. (2008)'s contextual implementation of the Doran (1981)'s SMART (Simple, Measurable, Achievable, Realistic, and Time limited) principle. A 'good' indicator should be representative, easily interpreted, broadly comparable, sensitive to change and have a reference or guideline value. To be 'useful', an indicator must be approved by international consensus, be well grounded and documented, have a reasonable cost/benefit ratio and have adequate historical and on-going spatial-temporal coverage. Flint et al. (2012) and James et al. (2012) further developed numerical scoring systems to help evaluate indicators objectively. Nevertheless, (Neary, 2012) warned against the potential to manipulate an index by saturating with inappropriate or biased indicators and whilst recommending that an index comprise of at least seven indicators, they did advocate that the type of indicator is more important than the number of indicators.

Since final outcomes are likely to be highly influenced by indicator choice, the robustness and sensitivity of both indicators and final outcomes to changes in ecosystem health should be understood if not formally investigated as part of the indicator selection process (Dobbie and Dail, 2013). Sensitivity analyses can involve:

- simulating changes in the underlying data of different magnitudes and estimating the resulting sensitivity (percentage or probability of change) expressed by the indicator
- estimating the effect of past perturbations on the indicator hindcasted from on historical data

As stressed above, indicators should align intimately with report card objectives. Yet in the more broad ecosystem report card frameworks, such indicators are often too general to be measurable. Therefore, in such cases, the indicators are further sub-divided into progressively more specific measures. For example, an indicator of water quality might comprise sub-indicators of nutrients, metals and physico-chemistry which in turn might be represented by more specific measures such as total nitrogen, mercury, dissolved oxygen, pH etc.

The resulting design is a hierarchical structure in which sub-indicators (etc) are nested within indicators and spatial scales are nested from entire regions, sub-regions or zones down to individual sites or sampling units. One of the strengths of such a hierarchical report card framework is that the inherent inbuilt redundancy allows for the addition, deletion or exchange of finer scale items (sites and actual measured variables) with minimum disruption to the actual report indicators. That is, the indicator is relatively robust to some degree of internal makeup. Furthermore, by abstracting away the fine details of an indicator, similar indicators from different report cards (each potentially comprising different sampling designs) are more directly comparably. For example, in different report cards that include water quality, a water quality indicator of 'water clarity' might comprise different Measures (e.g. suspended solids, NTU, Secchi depth etc) collected from different sources (e.g. satellite, in situ loggers or hand samples), yet provided each of these water clarity indicators are well calibrated, it should be possible to compare state and trend across the report cards.

Table 3: Water Quality Measure hierarchy specifying which Measures contribute to which Subindicators and which Subindicators contribute to which Indicators.

Indicator	Subindicator	Measure	Label	Units
Water Quality	Productivity	chl	Chlorophyll	μgL^{-1}
Water Quality	Water Clarity	nap	TSS	mgL^{-1}
Water Quality	Water Clarity	ntu	NTU	NTU
Water Quality	Water Clarity	sd	Secchi	m
Water Quality	Nutrients	NOx	NOx	μgL^{-1}

3.2 AIMS insitu samples

The AIMS component of MMP inshore water quality monitoring sampling program has been designed to quantify spatial and temporal patterns in inshore water quality, particularly in the context of catchment loads. Details of the sampling design are outlined in (Lønborg et al., 2016). From 2006–2014, AIMS visited 20 sites, three times per year (roughly corresponding to wet, early and late dry seasons), see Figures 2 and 3. The sites were largely selected along approximate north-south transects proximal to major rivers so as to provide samples along an expected water quality gradients (exposure to runoff). Following a review in 2014, the design was modified to intensify the spatial (32 sites) and temporal (typically between 5 and 10 samples per year) coverage of the sampling program. In particular, additional sampling effort was applied around three priority focal areas (Russell-Mulgrave, Tully and Burdekin).

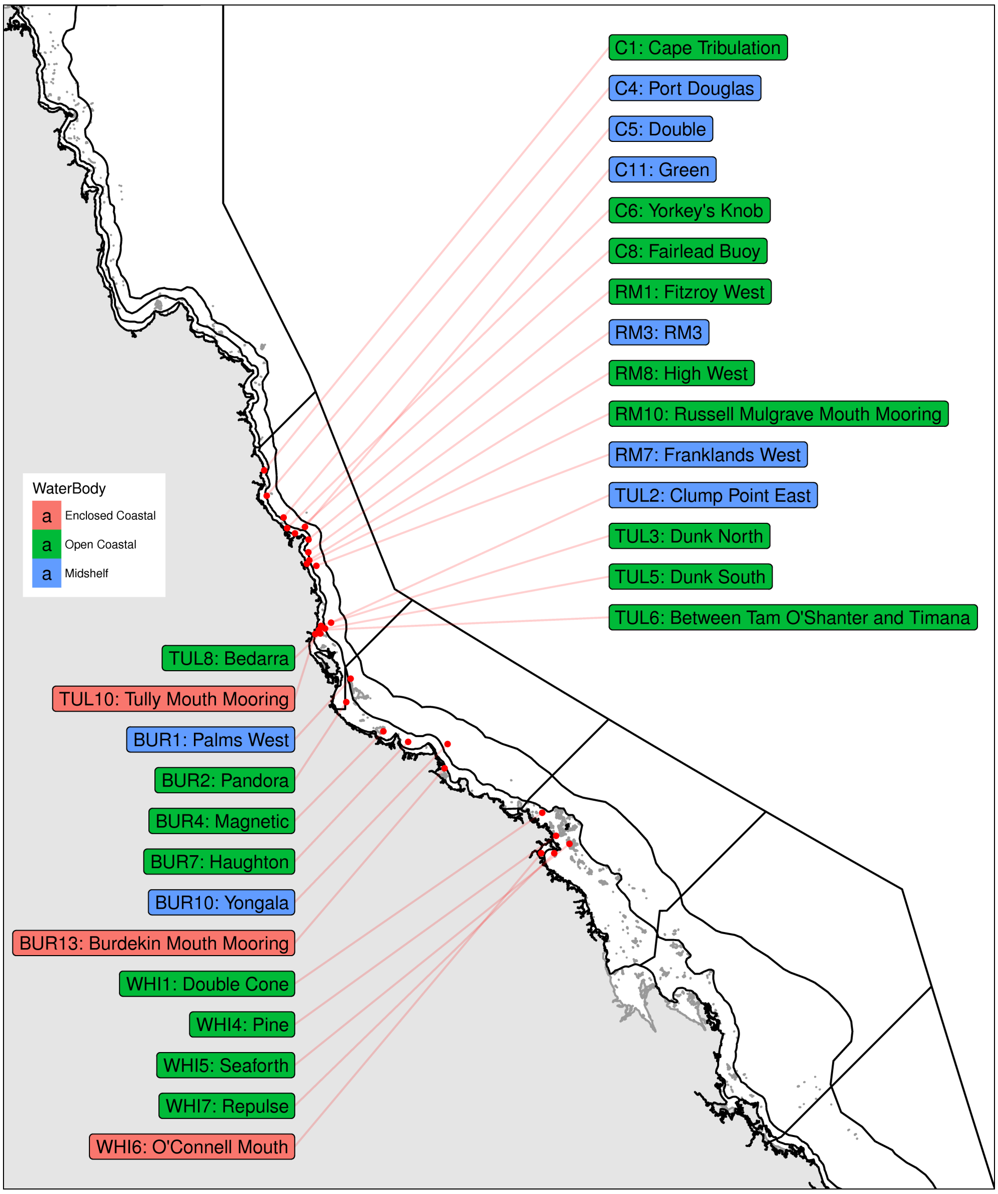


Figure 2: Map of AIMS in situ samples.

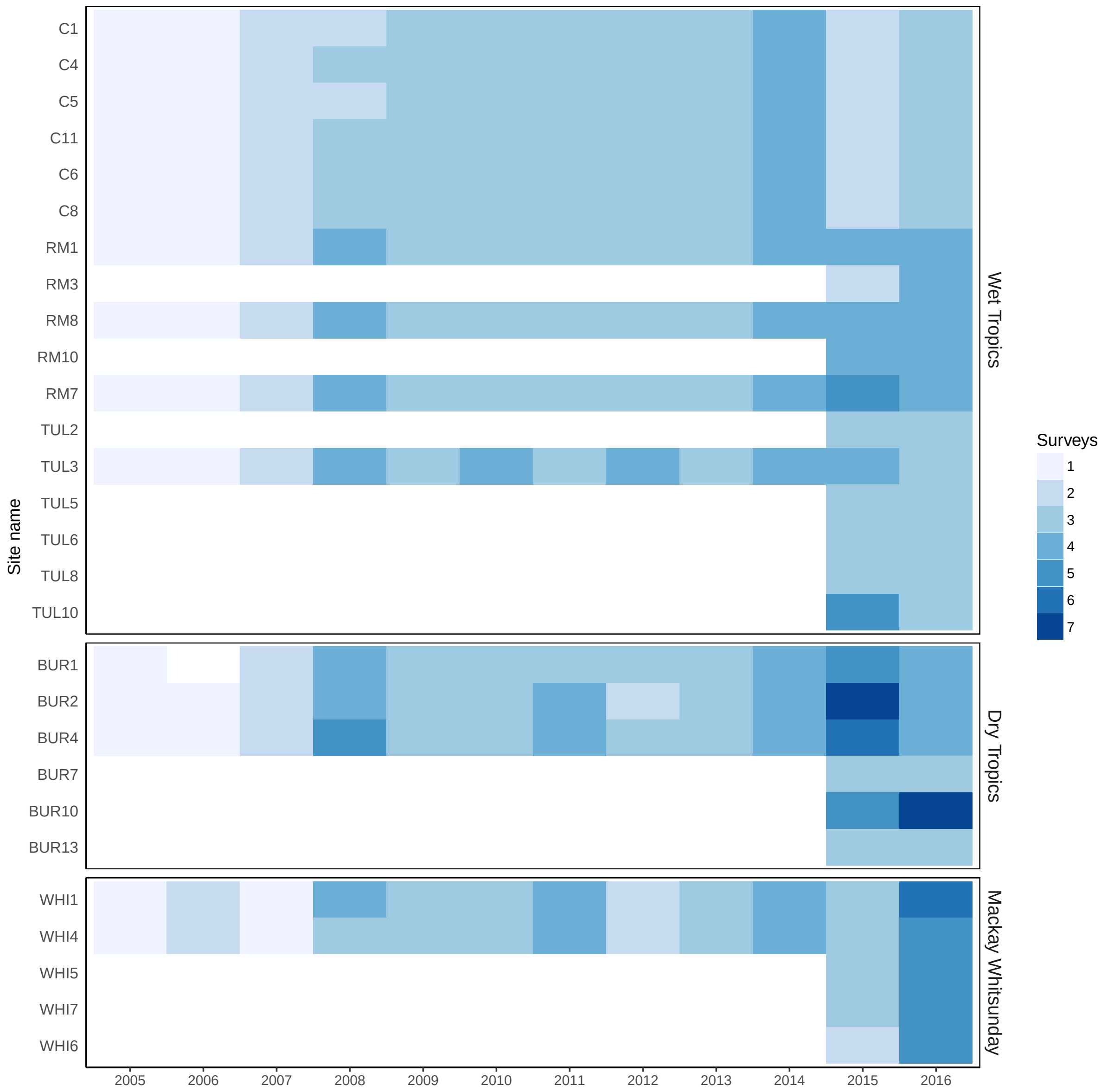


Figure 3: Spatial and temporal distribution of AIMS insitu samples. Sites names follow Great Barrier Reef Marine Park Authority (GBRMPA) and sites are arranged north to south into the focal Regions. Blue shading of tiles denotes the number of surveys conducted in the year at each site.

Table 4: Measures collected in AIMS MMP insitu inshore water quality monitoring program. NOx is the sum of NO_2 and NO_3. Data used are annual means of depth weighted averages per site.

Measure	Variable	Description	Abbreviation	Conversion	Units
Chlorophyll-a	DRIFTCHL_UGPERL.wm	Chlorophyll-a (µg/L)	chl	x1	μgL^{-1}
Total Suspended Solids	TSS_MGPERL.wm	Suspended solids (mg/L)	nap	x1	mgL^{-1}
Secchi Depth	SECCHI_DEPTH.wm	Secchi depth (m)	sd	x1	m
NOx	NOX.wm	Nitrite and Nitrate measured by microanalyser (µM/L)	NOx	x14	μgL^{-1}

3.3 AIMS FLNTU samples

Combination continuous Flourometer and Turbidity Sensors (hereafter FLNTU) loggers were deployed at 15 of the AIMS MMP inshore water quality monitoring sites.

Table 5: Measures collected in AIMS MMP flntu inshore water quality monitoring program. Data used are daily means per site.

Measure	Variable	Description	Abbreviation	Conversion	Units
Chlorophyll-a	CHL_QA_AVG	??	chl	CHL_QA_AVG x1	μgL^{-1}
NTU	NTU_QA_AVG	??	ntu	NTU_QA_AVG x1	NTU

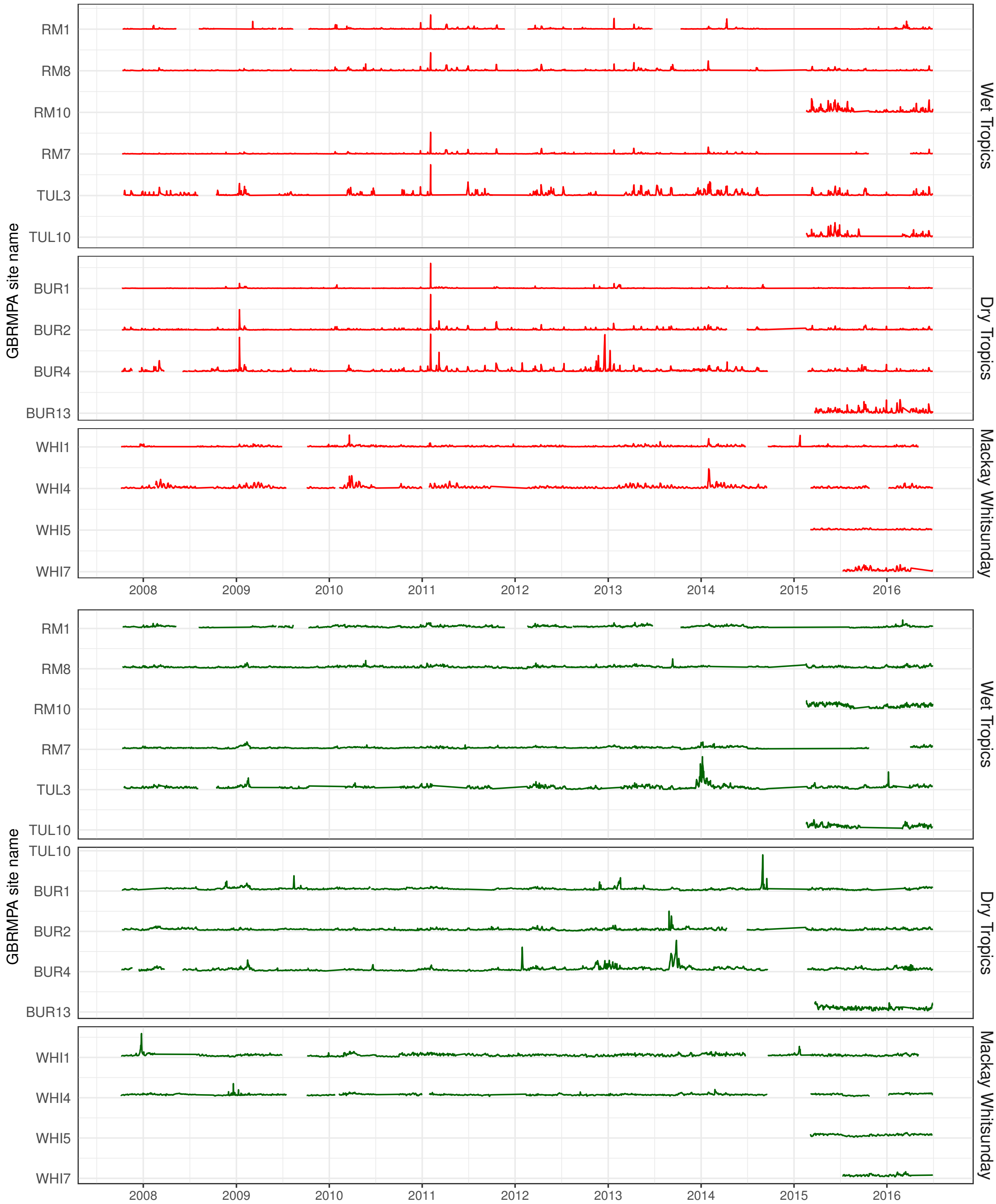


Figure 4: Spatial and temporal distribution of AIMS FLNTU samples (Red: NTU, Green: Chlorophyll-a). Sites names follow Great Barrier Reef Marine Park Authority (GBRMPA) and sites are arranged north to south into the focal Regions.

3.4 Remote sensing (BOM satellite)

Daily (July 2002–Dec 2016, $1 \times 1km^2$ resolution) Moderate Resolution Imaging Spectroradiometer (MODIS satellite) imagery (hereafter referred to as Satellite) data were obtained by downloading NETCDF files from the thredds server.

Table 6: Measures collected from MODIS satellite imaging. Data used are daily means per pixel. Variable and Description pertain to the eReefs source. Conversion indicates the conversion applied on data to conform to threshold Units. Abbreviation provides a consistent key accross data.

Measure	Variable	Description	Abbreviation	Conversion	Units
Chlorophyll-a	Chl_MIM	??	chl	Chl_MIM x1	µgL^{-1}
Non-Algal Particles	Nap_MIM	??	nap	Nap_MIM x1	mgL^{-1}
Secchi Depth	SD_MIM	??	sd	SD_MIM x1	m

3.5 eReefs assimilated model

Mark to provide a brief description. In this context, the eReefs model refers to the gbr4_bgc_?? model (see Table?? for the catalog and model descriptions).

This source of data only extends back to 2014. Whilst the eReefs GBR4_BGC_? model technically does contain 2013 calendar year data, the current project partitions time into water years in which the full 2013 water year starts in October 2012. Therefore as the 2013 is not a complete 12 months of data, it is excluded from analyses. Unfortunately, this means that any signals associated with the 2010-2011 floods are unavailable.

Table 7: Measures collected from eReefs assimilated model. Data used are daily means per pixel. Variable and Description pertain to the eReefs source. Conversion indicates the conversion applied on data to conform to threshold Units. Abbreviation provides a consistent key accross data.

Measure	Variable	Description	Abbreviation	Conversion	Units
Chlorophyll-a	Chl_a_um	Sum of Chlorophyll concentration of four microalgae types (mg/m^3)	chl	Chl_a_um x1	µgL^{-1}
Non-Algal Particles	EFI	??	nap	EFI x1000	mgL^{-1}
Secchi Depth	Kd_490	??	sd	1/Kd_490	m
NOx	NO3	Concentration of Nitrate. As Nitrite is not represented in the model, NO3 = $[NO_3^-] + [NO_2^-]$ (mg/m^3)	NOx	NO3 x1	µgL^{-1}

3.6 eReefs926

Mark to provide a brief description. In this context, the eReefs926 model refers to the gbr4_bgc_926 model (see Table?? This model provides alternative formulation and importantly does extend back to the full 2013 water year thereby providing some coverage closer to the 2010-2011 flood period.

Variables used as per Table 7

3.7 Thresholds

An environmental health metric represents the state or condition relative to some reference, threshold or expectation. Most of the current water quality indices compare values to a set of specifically selected guidelines. These guidelines are either formulated specifically from long-term historical data appropriate to the spatial and temporal domain of interest or else are based on ANZEC guidelines (Australian and New Zealand Environment and Conservation Council, 2000).

Typically there are strict guidelines on how these guidelines should be applied. In particular, the guidelines associated with various measures used in various report cards throughout the Great Barrier Reef should be applied to annually aggregated data - not individual observations. Since this project intends to generate indices on the scale of individual observations, we have decided to refer to the guidelines as thresholds so as to avoid contradicting the terms of use of guidelines..

The thresholds used for each Measure within each Region and Water body are indicated in Table A1 (page 128). Note, that whilst the application of seasonal thresholds could potentially remove some uncertainty, in the absence of clear consensus on how to define wet and dry seasons and what the associated set of thresholds would be, seasonal thresholds are not used in this project.

4. EXPLORATORY DATA ANALYSIS

Exploratory data analysis is vital for informing data processing and analysis as well as establishing assumptions and limitations. Of particular importance for the current project is the spatial and temporal distribution and variability of the various data Measures and Sources. As such, a series of exploratory plots have been generated (see Appendix ?? begging on page ??). In the interest of keeping the main text free of copious graphics, we have elected to present only a small fraction of the exploratory data analyses figures here. The figures presented will act as exemplars of general format and predominant features or patterns.

4.1 All data

Figures 5 – 8 display the temporal distribution of Chlorophyll-a, TSS, Secchi depth and NOx observations for the Wet Tropics Open Coastal Zone from AIMS insitu, AIMS FLNTU, Satellite, eReefs and eReefs926 sources. All combinations of Measure/Zone/Source can be found in Figures ??–??.

All of the figures are presented with log-transformed y-axes as the data are typically skewed to the right. This is expected for parameters that have a natural minimum (zero), yet no theoretical maximum. It does however mean that these distributional properties should be considered during the analyses. In particular, for mean based aggregations, outliers and skewed distributions can impart unrepresentative influence on outcomes.

Each of the data sources presents different variability characteristics. The scale of the range of AIMS insitu data is predominantly and approximately less than or equal to the scale of the half/twice the associated threshold value (Fig. 5a). The AIMS FLNTU logger data (Fig. 5b) have a larger range than the AIMS insitu data - presumably because the former data collection frequency captures most of the peaks and trophs whereas the later is unlikely to do so. Similarly, the scale of the range eReefs and eReefs926 data (Fig. 5d-e) is approximately equal to the scale of the range of the span from half/twice the threshold value. This reflects both a more complete time series and broader spatial extent represented in the data. In contrast to the AIMS insitu and to a lesser extent the AIMS FLNTU and eReefs data, the scale of the range of the Satellite is relatively large - typically a greater span than the range of half/twice threshold value (Fig. 5c).

The Satellite, eReefs and eReefs926 data series all start and end part of the way through a water year. For annually aggregated data, this is likely to result in unrepresentative estimates and thus only full water years will be analysed.

4.2 Annual data

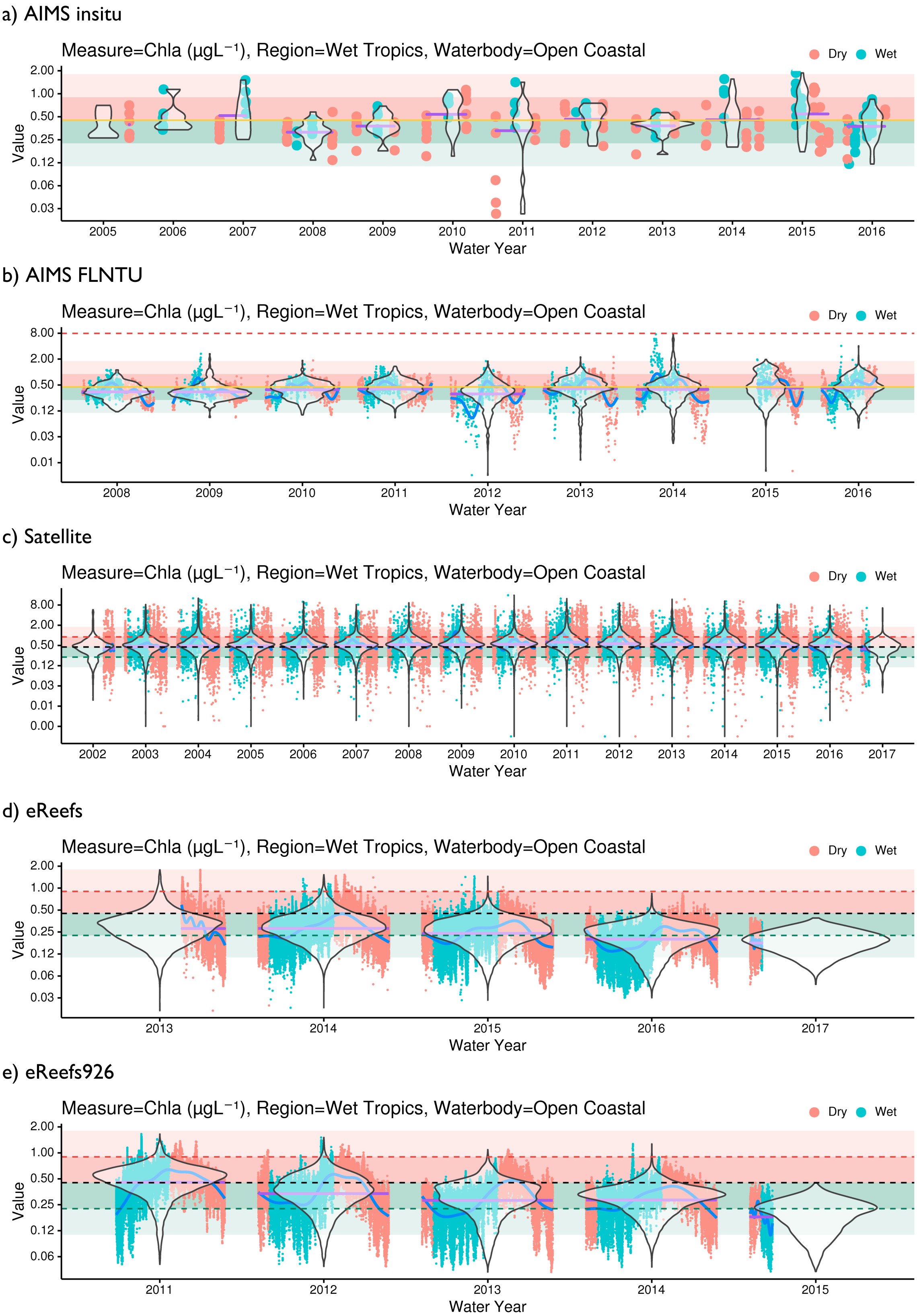
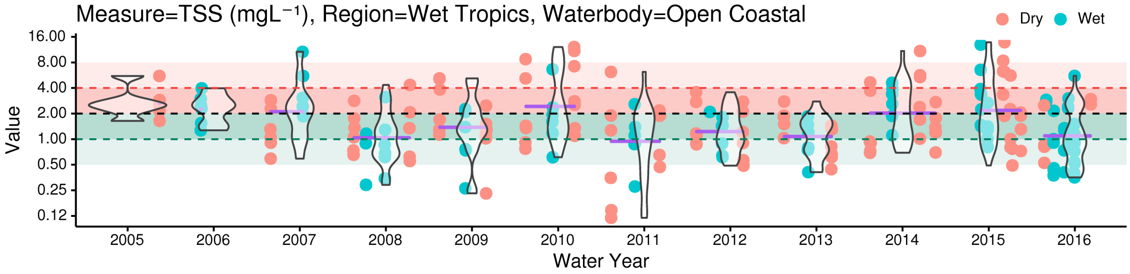
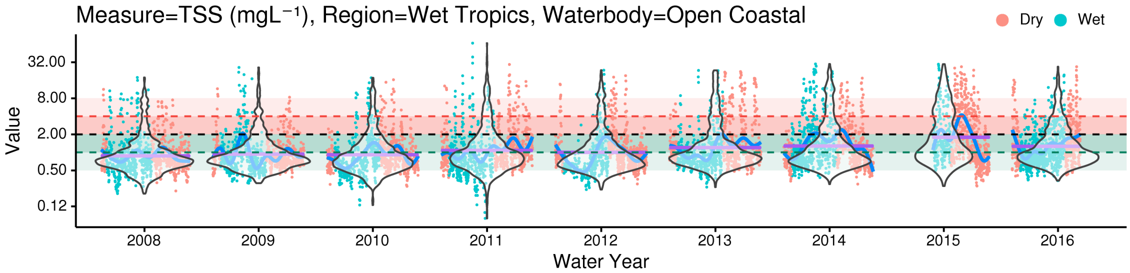


Figure 5: Observed (logarithmic axis with violin plot overlay) Chlorophyll-a data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU, c) Satellite, d) eReefs and e) eReefs926. Observations are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Blue smoother represents Generalized Additive Mixed Model within a water year and purple line represents average within the water year. Horizontal red, black and green dashed lines denote the twice threshold, threshold and half threshold values respectively. Red and green background shading indicates the range (10% shade: $\times 4, /4$, 30% shade: $\times 2, /2$) above and below threshold respectively.

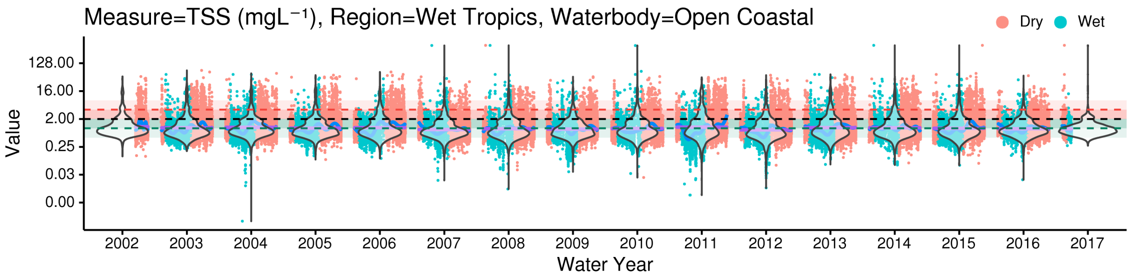
a) AIMS insitu



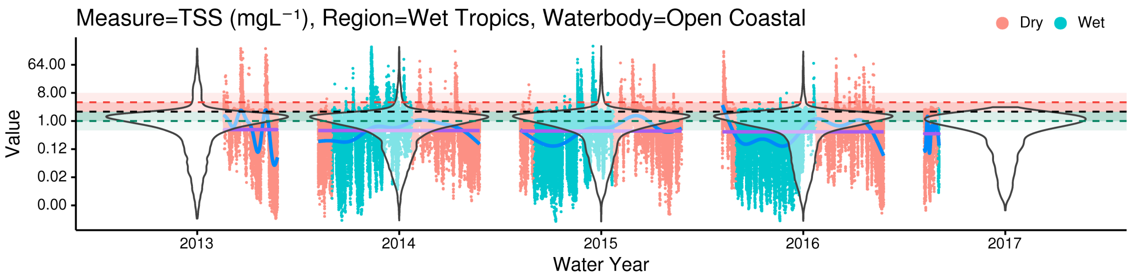
b) AIMS FLNTU



c) Satellite



d) eReefs



e) eReefs926

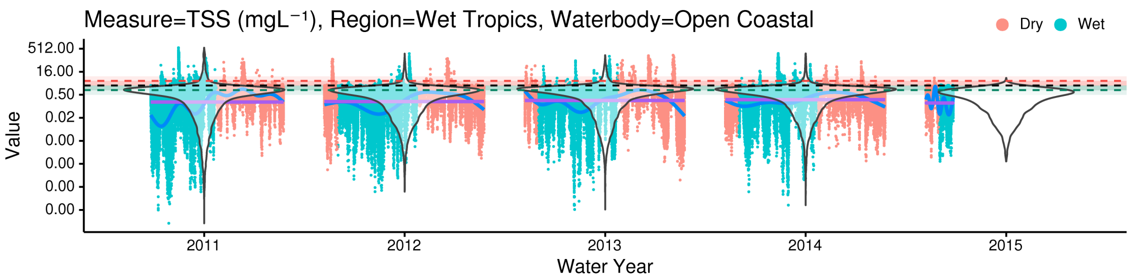
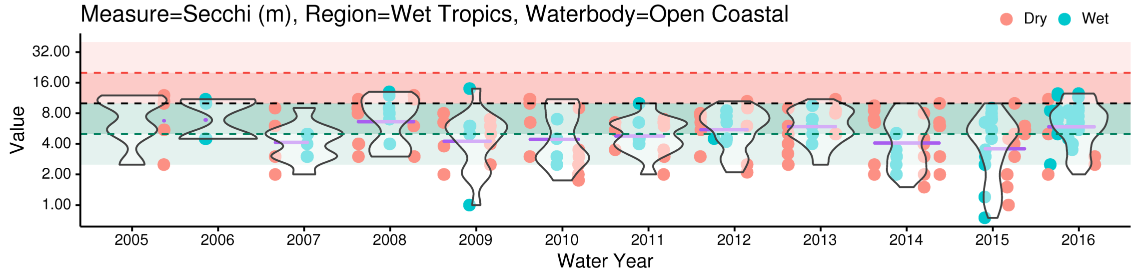
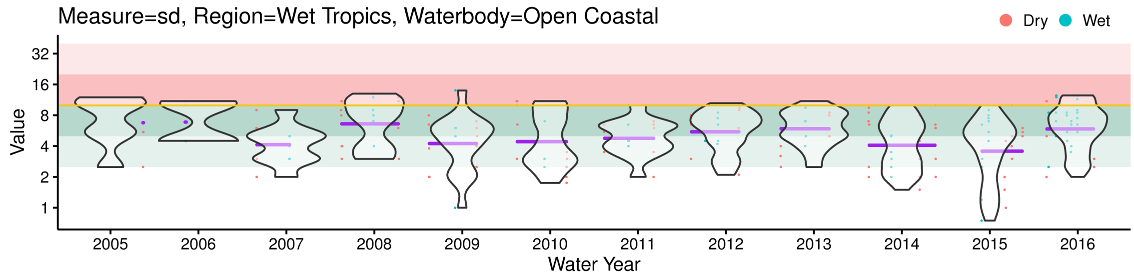


Figure 6: Observed (logarithmic axis with violin plot overlay) TSS data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU, c) Satellite, d) eReefs and e) eReefs926. Observations are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Blue smoother represents Generalized Additive Mixed Model within a water year and purple line represents average within the water year. Horizontal red, black and green dashed lines denote the twice threshold, threshold and half threshold values respectively. Red and green background shading indicates the range (10% shade: $\times 4, /4$, 30% shade: $\times 2, /2$) above and below threshold respectively.

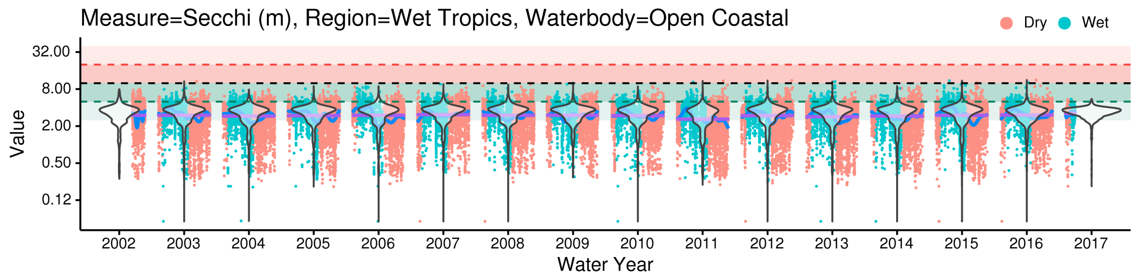
a) AIMS insitu



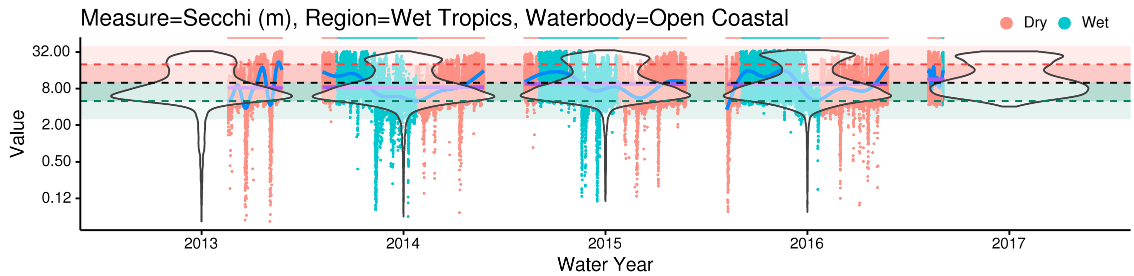
b) AIMS FLNTU



c) Satellite



d) eReefs



e) eReefs926

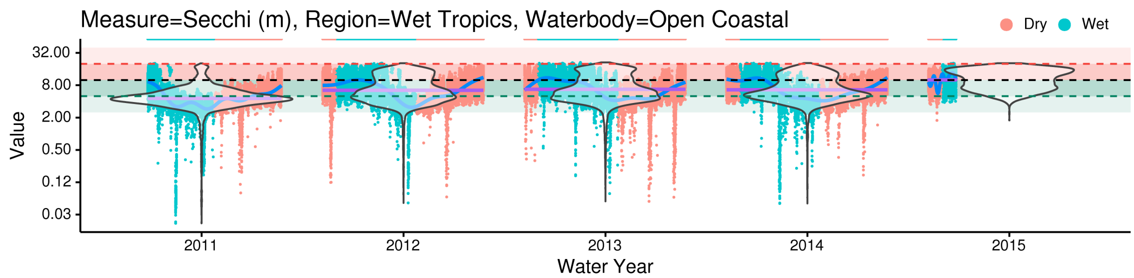


Figure 7: Observed (logarithmic axis with violin plot overlay) Secchi depth data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU, c) Satellite, d) eReefs and e) eReefs926. Observations are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Blue smoother represents Generalized Additive Mixed Model within a water year and purple line represents average within the water year. Horizontal red, black and green dashed lines denote the twice threshold, threshold and half threshold values respectively. Red and green background shading indicates the range (10% shade: $\times 4, /4$, 30% shade: $\times 2, /2$) above and below threshold respectively.

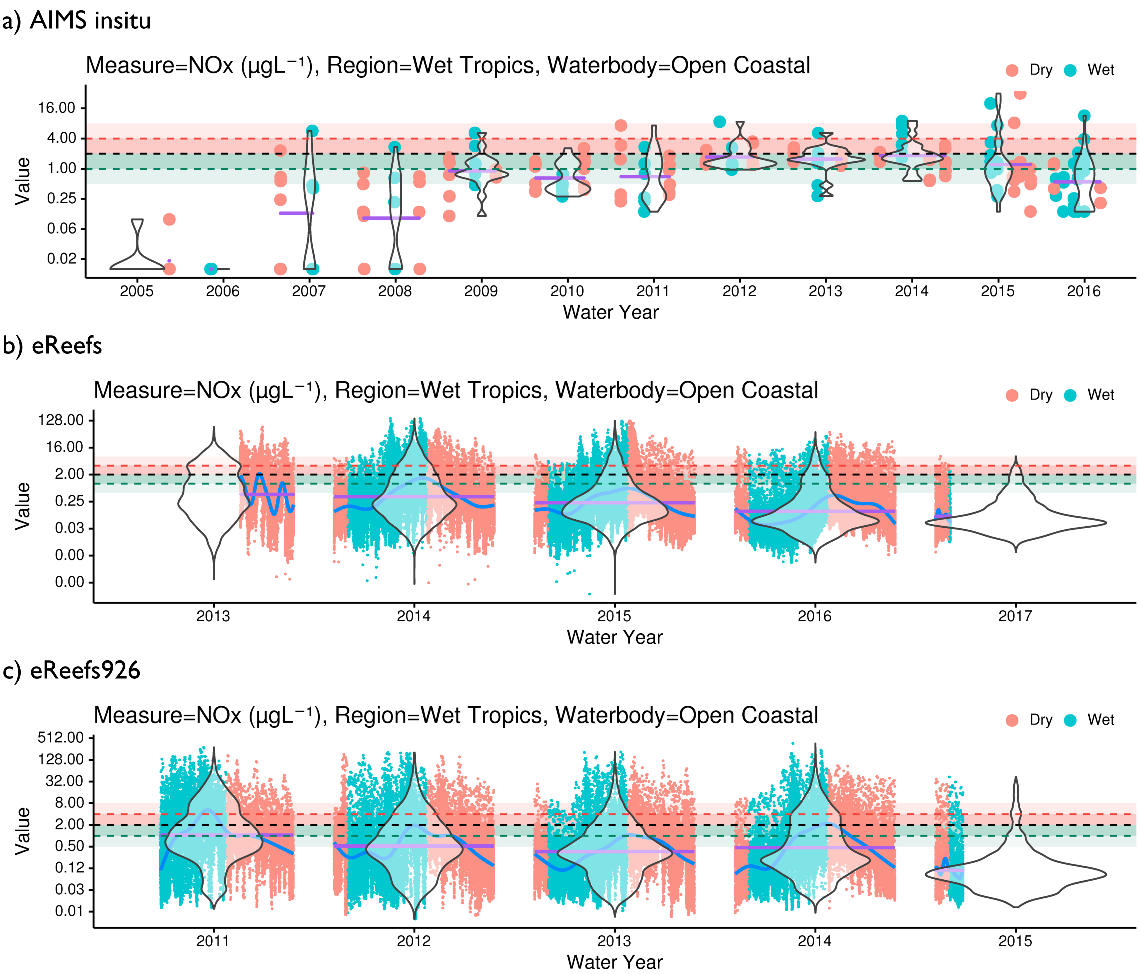


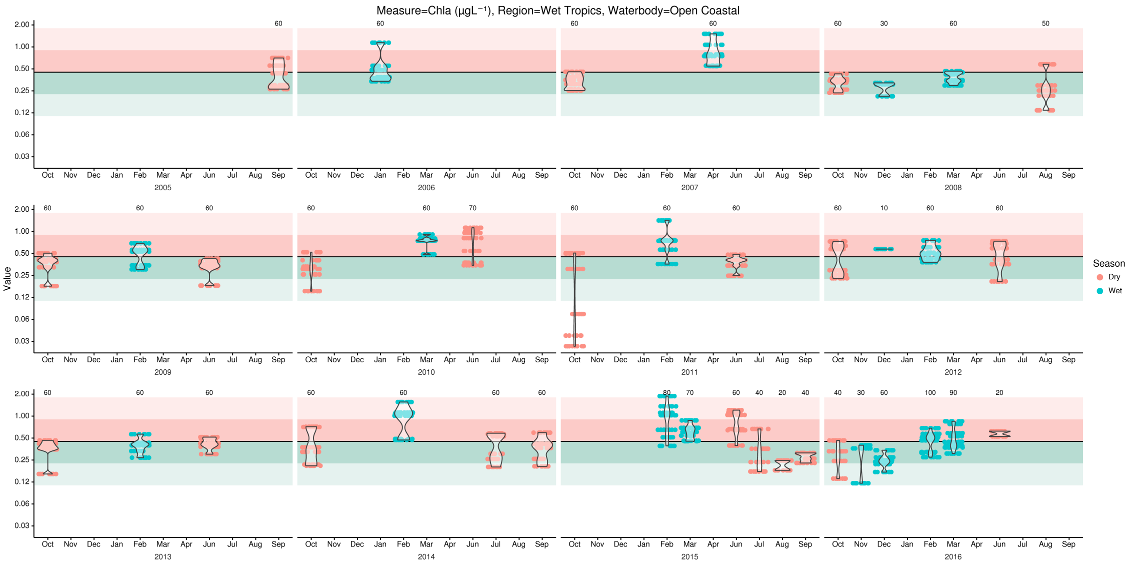
Figure 8: Observed (logarithmic axis with violin plot overlay) NOx data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) eReefs and c) eReefs926. Observations are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Blue smoother represents Generalized Additive Mixed Model within a water year and purple line represents average within the water year. Horizontal red, black and green dashed lines denote the twice threshold, threshold and half threshold values respectively. Red and green background shading indicates the range (10% shade: $\times 4, /4$, 30% shade: $\times 2, /2$) above and below threshold respectively.

4.3 Monthly data

Figures 9 – 14 provide finer temporal resolution by displaying the temporal distribution of Chlorophyll-a, TSS, Secchi depth and NOx observations for the each month within Wet Tropics Open Coastal Zone from AIMS insitu, AIMS FLNTU, Satellite, eReefs and eReefs926 sources. Additional combinations of Measure/Zone/Source can be found in Figures ??–??.

The monthly violin plots do not add any additional insights with respect to understanding the characteristics of the underlying data to help guide the selection of appropriate indexation formulation or perhaps even Measure/Source selection. Rather, they provide a less compacted view of the underlying data from which patterns highlighted in Section ?? might be more easily appreciated.

a) AIMS insitu



b) AIMS FLNTU

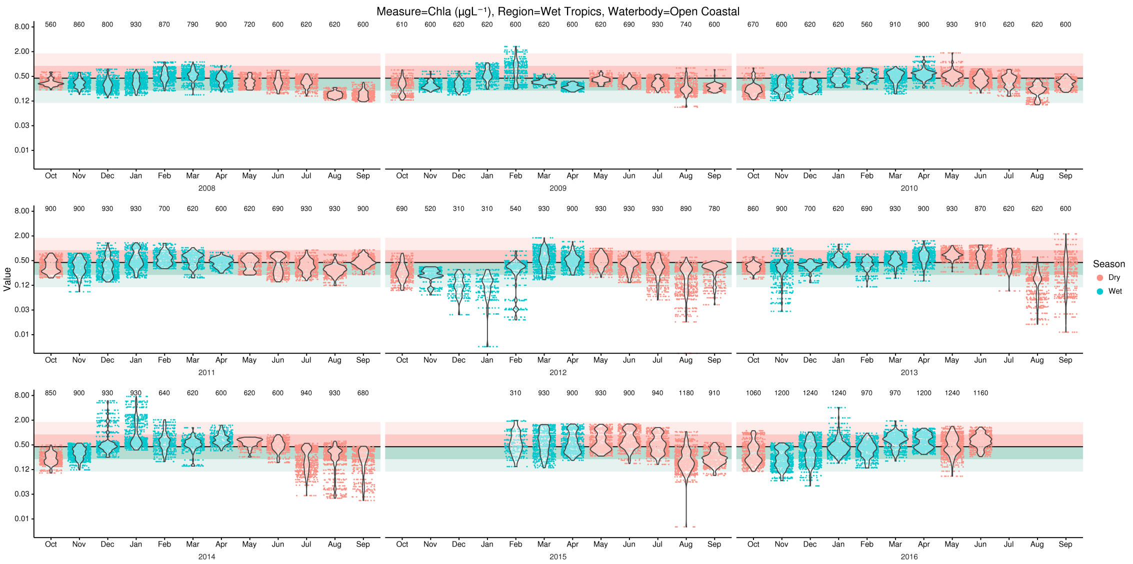
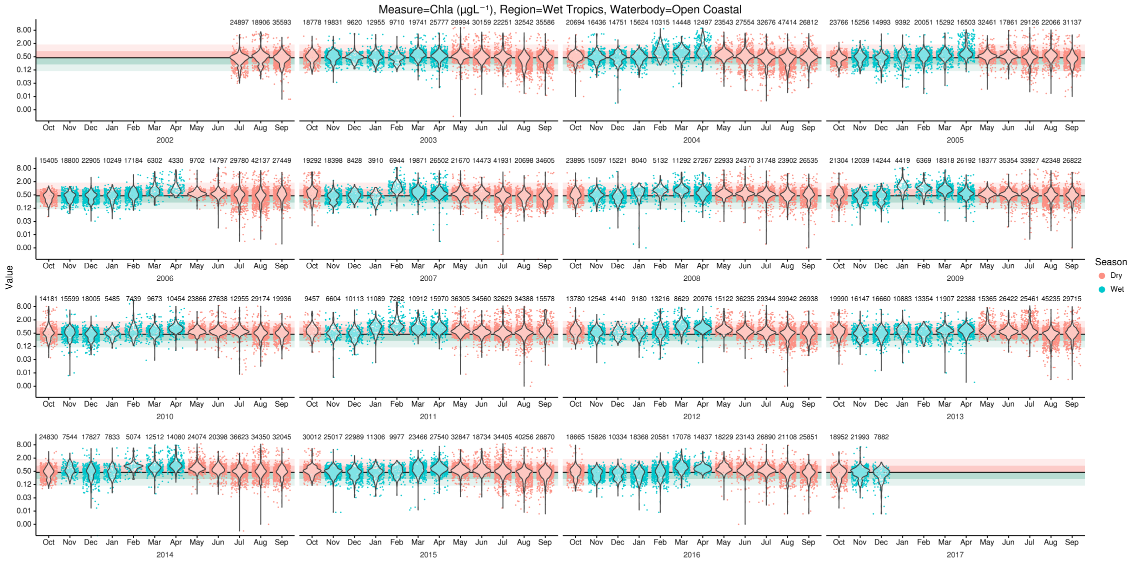


Figure 9: Observed (logarithmic axis with violin plot overlay) Chlorophyll-a data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: $\times 4, /4$, 30% shade: $\times 2, /2$) above and below threshold respectively.

a) Satellite



b) eReefs

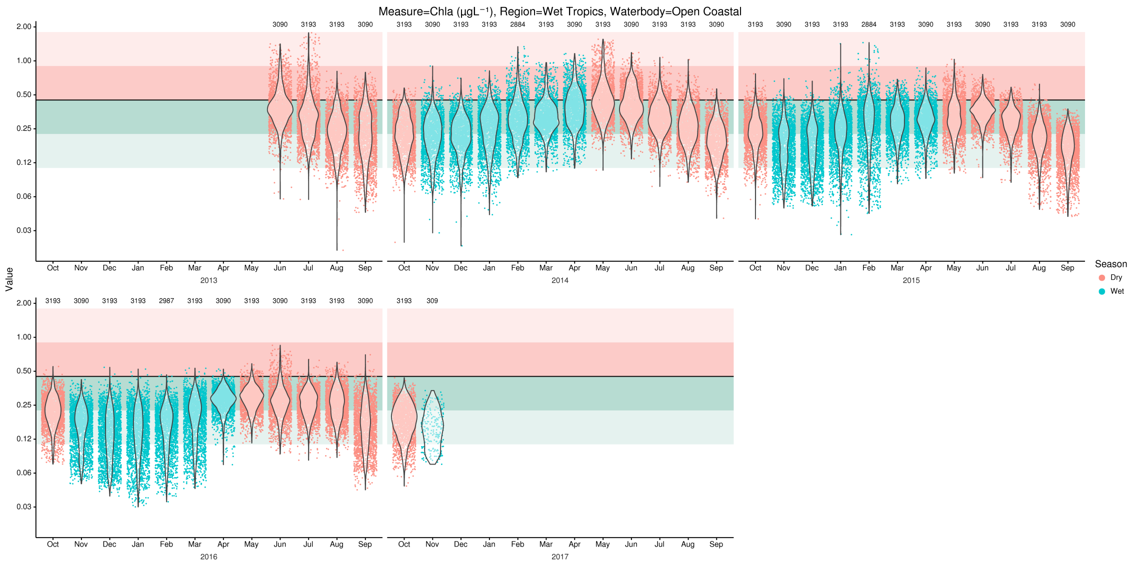
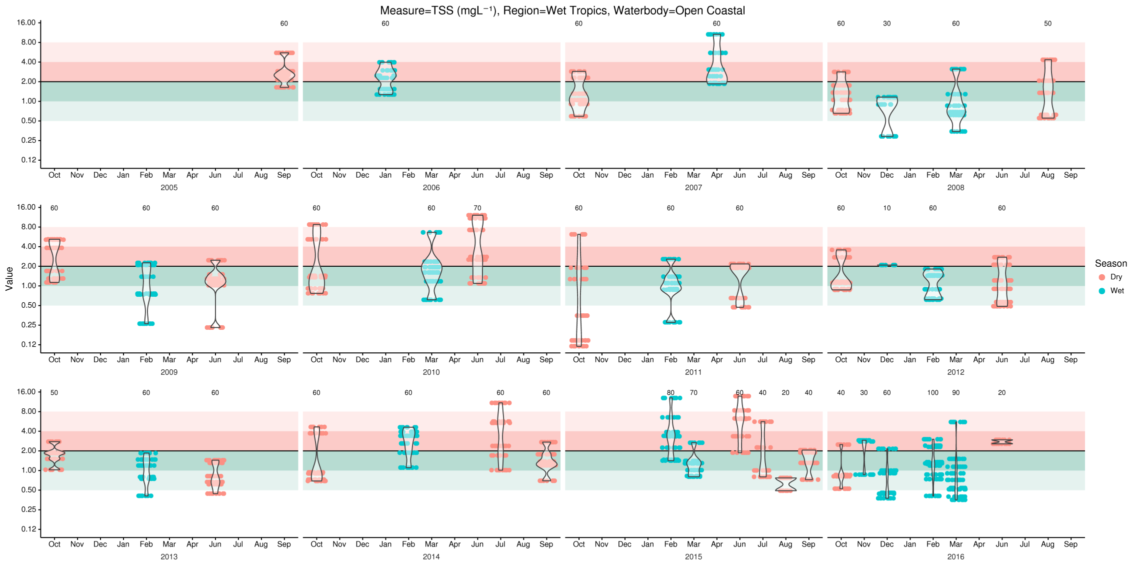


Figure 10: Observed (logarithmic axis with violin plot overlay) Chlorophyll-a data for the Wet Tropics Open Coastal Zone from a) Satellite, b) eReefs. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: $\times 4, /4$, 30% shade: $\times 2, /2$) above and below threshold respectively.

a) AIMS insitu



b) AIMS FLNTU

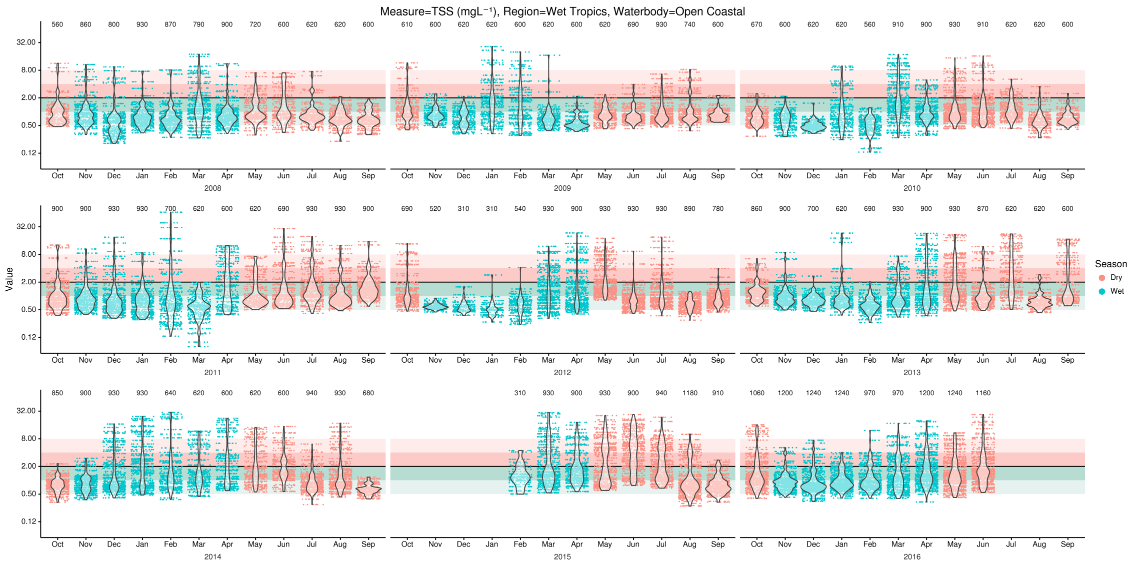
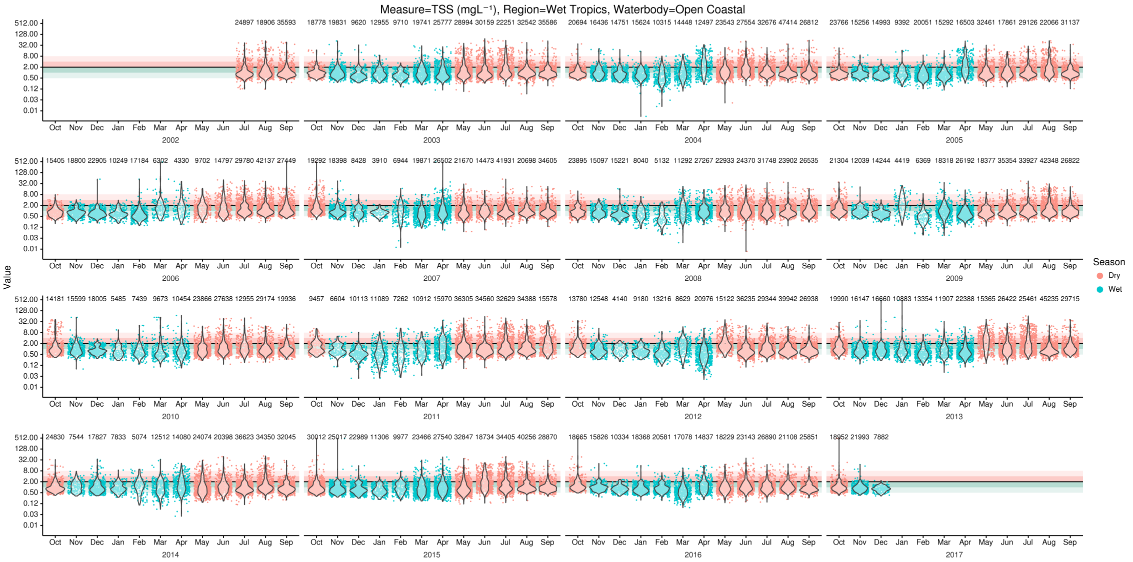


Figure 11: Observed (logarithmic axis with violin plot overlay) TSS data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: $\times 4, /4$, 30% shade: $\times 2, /2$) above and below threshold respectively.

a) Satellite



b) eReefs

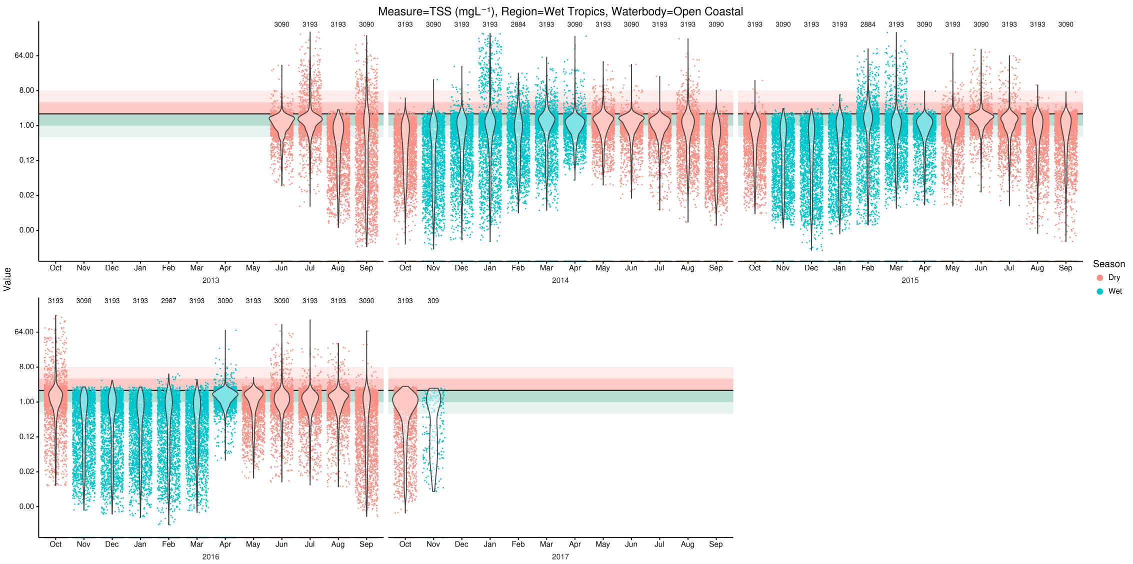
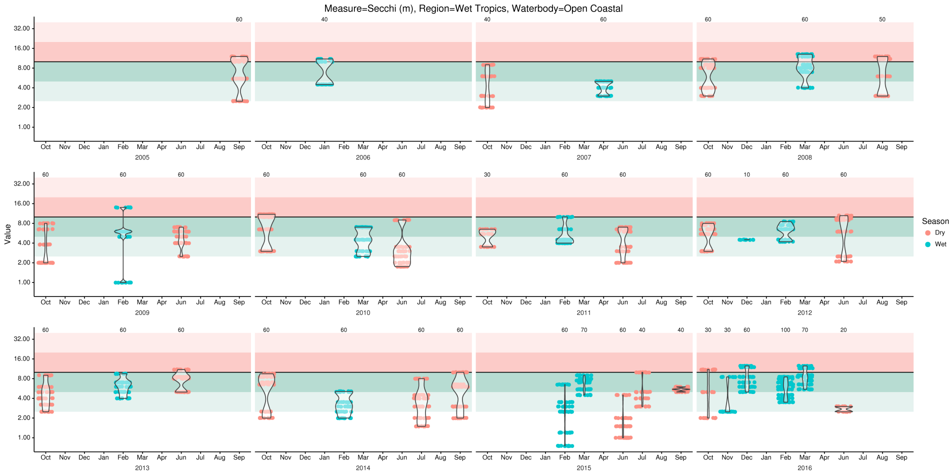
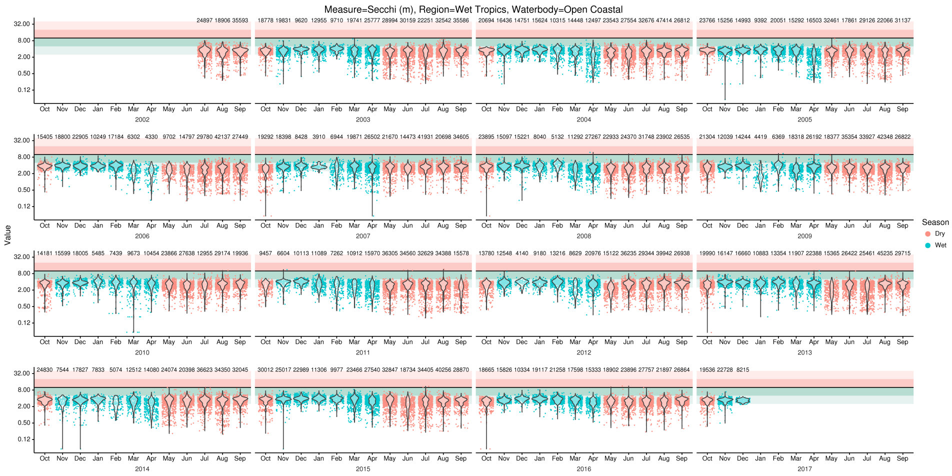


Figure 12: Observed (logarithmic axis with violin plot overlay) TSS data for the Wet Tropics Open Coastal Zone from a) Satellite, b) eReefs. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: $\times 4, /4$, 30% shade: $\times 2, /2$) above and below threshold respectively.

a) AIMS insitu



b) Satellite



c) eReefs

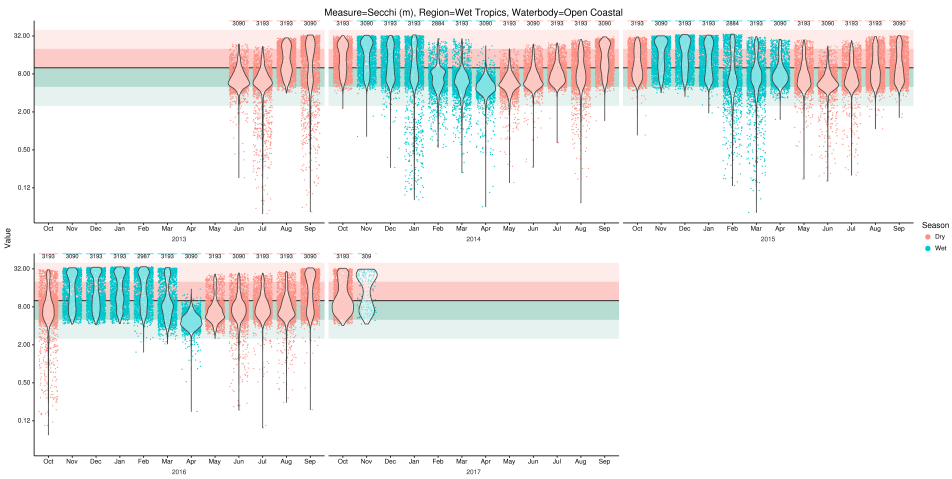
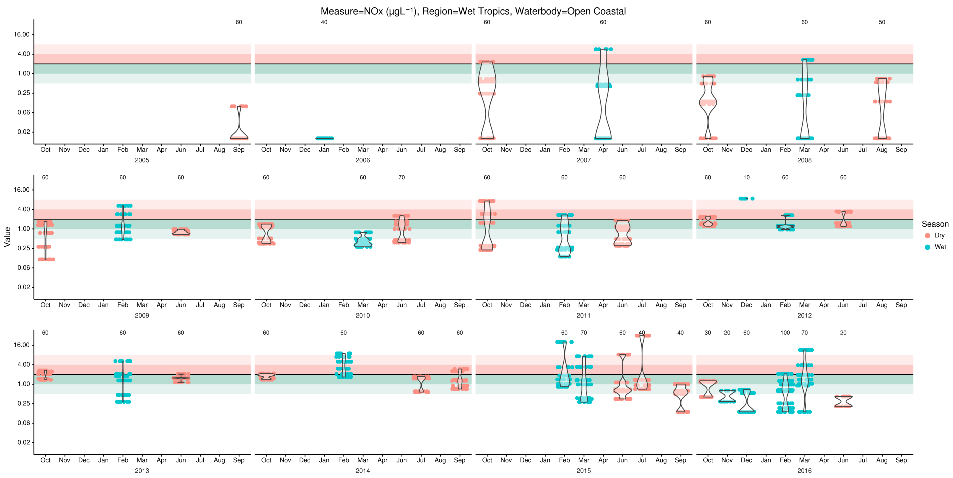
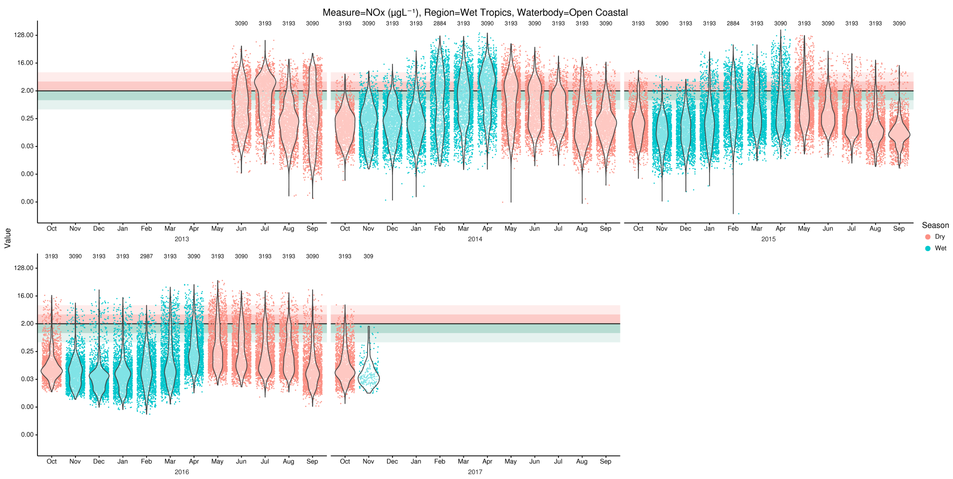


Figure 13: Observed (logarithmic axis with violin plot overlay) Secchi depth data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) Satellite and c) eReefs. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: $\times 4, /4$, 30% shade: $\times 2, /2$) above and below threshold respectively.

a) AIMS insitu



b) eReefs



c) eReefs926

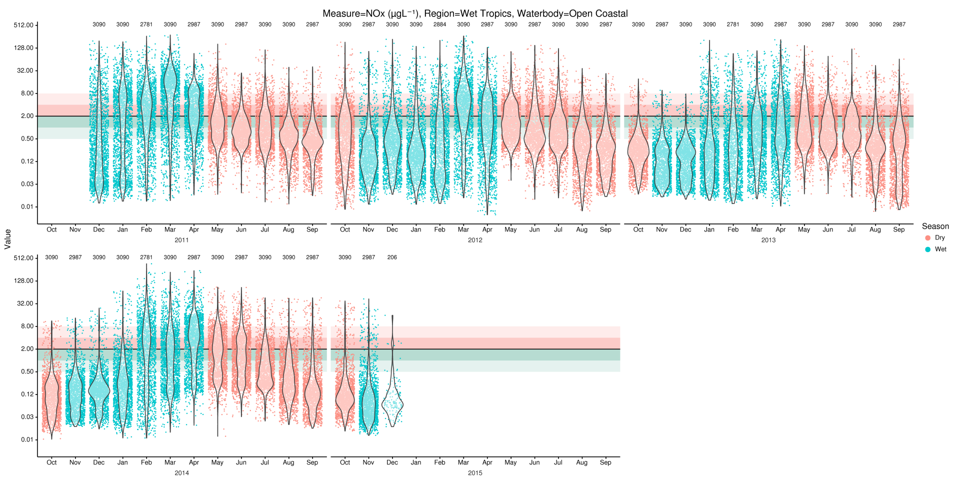


Figure 14: Observed (logarithmic axis with violin plot overlay) Secchi depth data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) eReefs c) eReefs926. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: $\times 4, / 4$, 30% shade: $\times 2, / 2$) above and below threshold respectively.

4.4 Spatial data

Figures 15 – 22 explore the spatio-temporal patterns in observed data from a finer spatial perspective (focussing on just the Wet Tropics Open Coastal and Dry Tropics Midshelf Zones). These figures highlight the disparity in resolution between the different data sources. The AIMS insitu data is spatially very sparse [1]. The Satellite data has the most extensive spatial resolution and notwithstanding the many gaps due to various optical interferences (such as cloud cover), also has the greatest temporal coverage.

For the selected Zones and span of water years, there is little evidence of a major latitudinal gradient in Satellite Chlorophyll-a with most of any change (if any) occurring across the shelf. Indeed, Satellite parameters are relatively constant over space and time for the Dry Tropics Midshelf Zone (see Figs. 19–21b). Moreover, the spatial patterns of Satellite derived Chlorophyll-a and TSS appear relatively invariant between years (see Figs.15–21b).

The eReefs and eReefs926 do show some variability in spatial and temporal Chlorophyll-a and Secchi depth (see Figs. 15c-d,17c-d,19c-d and 21c-d), yet relatively little for TSS and NOx (at least for Dry Tropics Midshelf).

[1] the AIMS FLNTU logger data is even more sparse and thus is not shown.

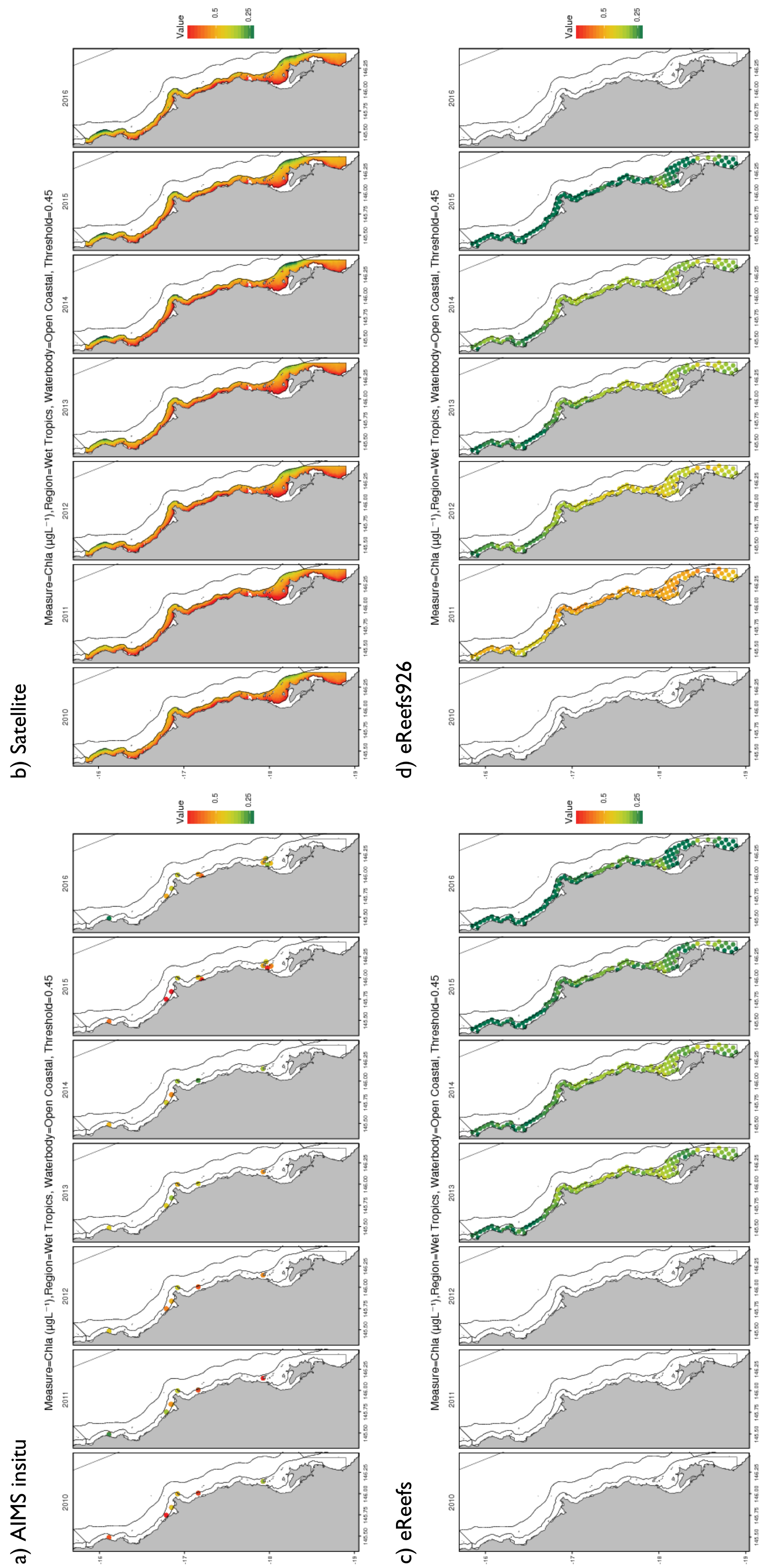


Figure 15: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 Chlorophyll-a (2009–2016) for the Wet Tropics Open Coastal Zone.

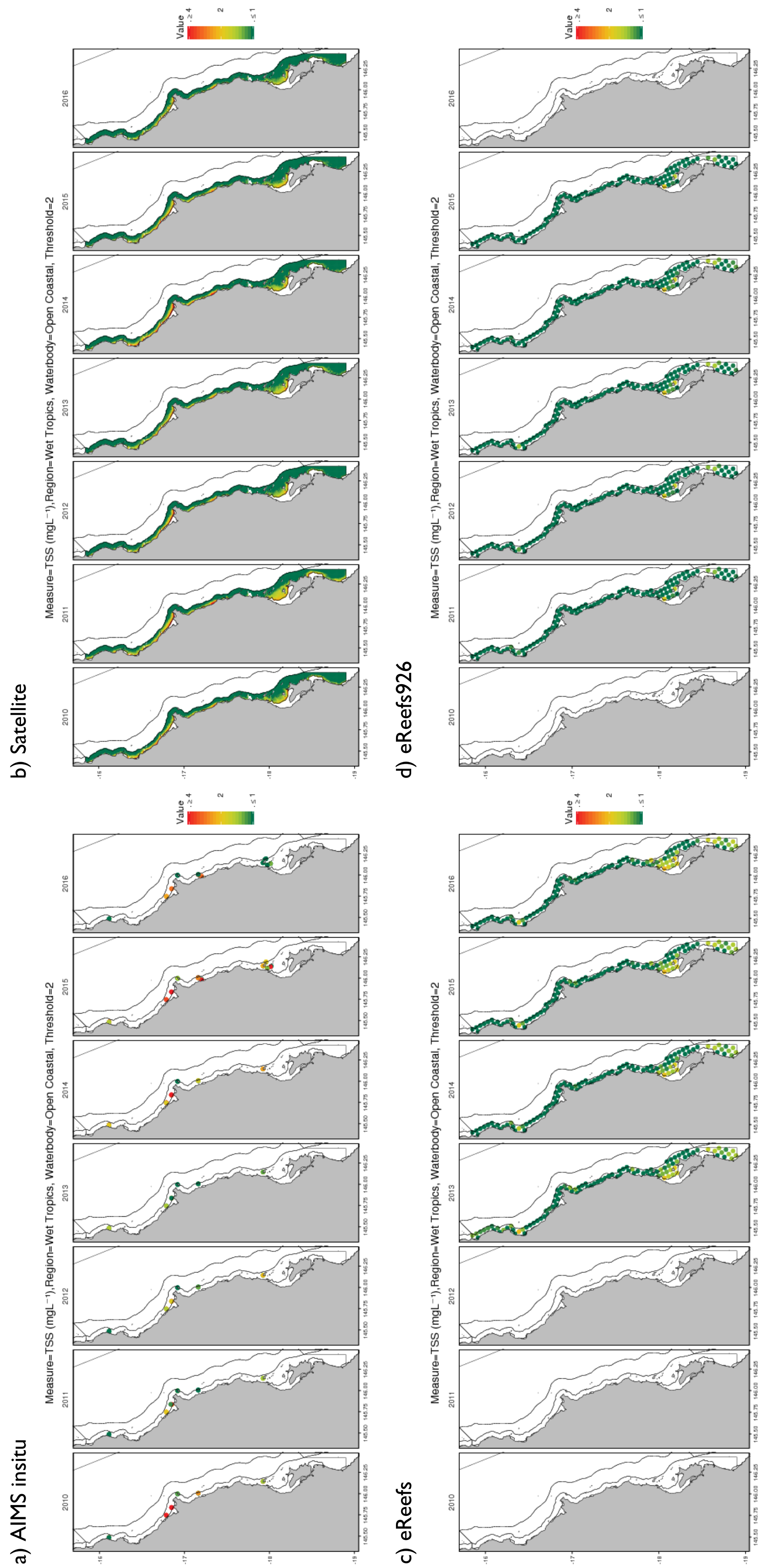


Figure 16: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 TSS (2009–2016) for the Wet Tropics Open Coastal Zone.

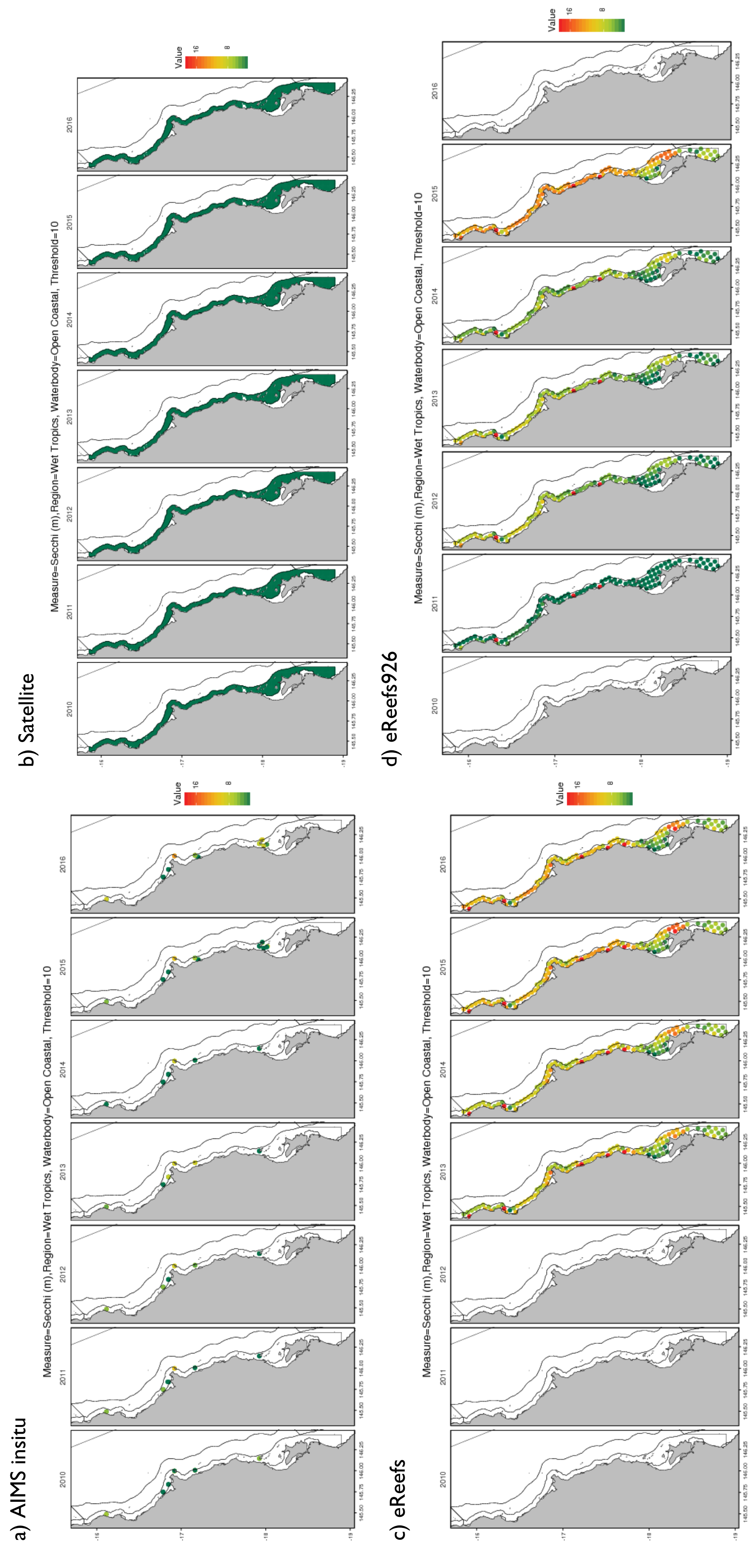


Figure 17: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 Secchi depth (2009–2016) for the Wet Tropics Open Coastal Zone.

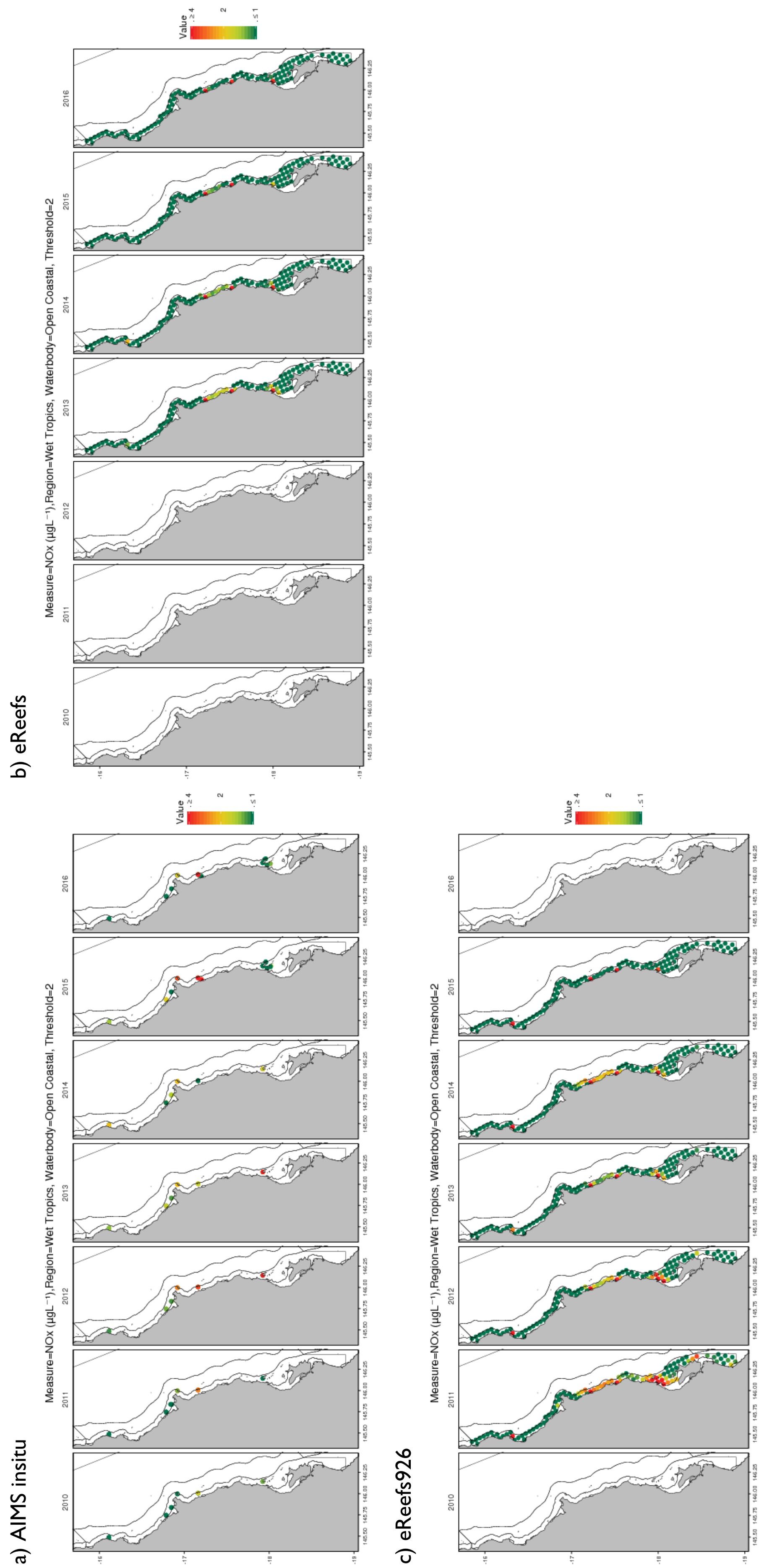


Figure 18: Spatial distribution of observed a) AIMS insitu, b) eReefs and c) eReefs926 NOx (2009–2016) for the Wet Tropics Open Coastal Zone.

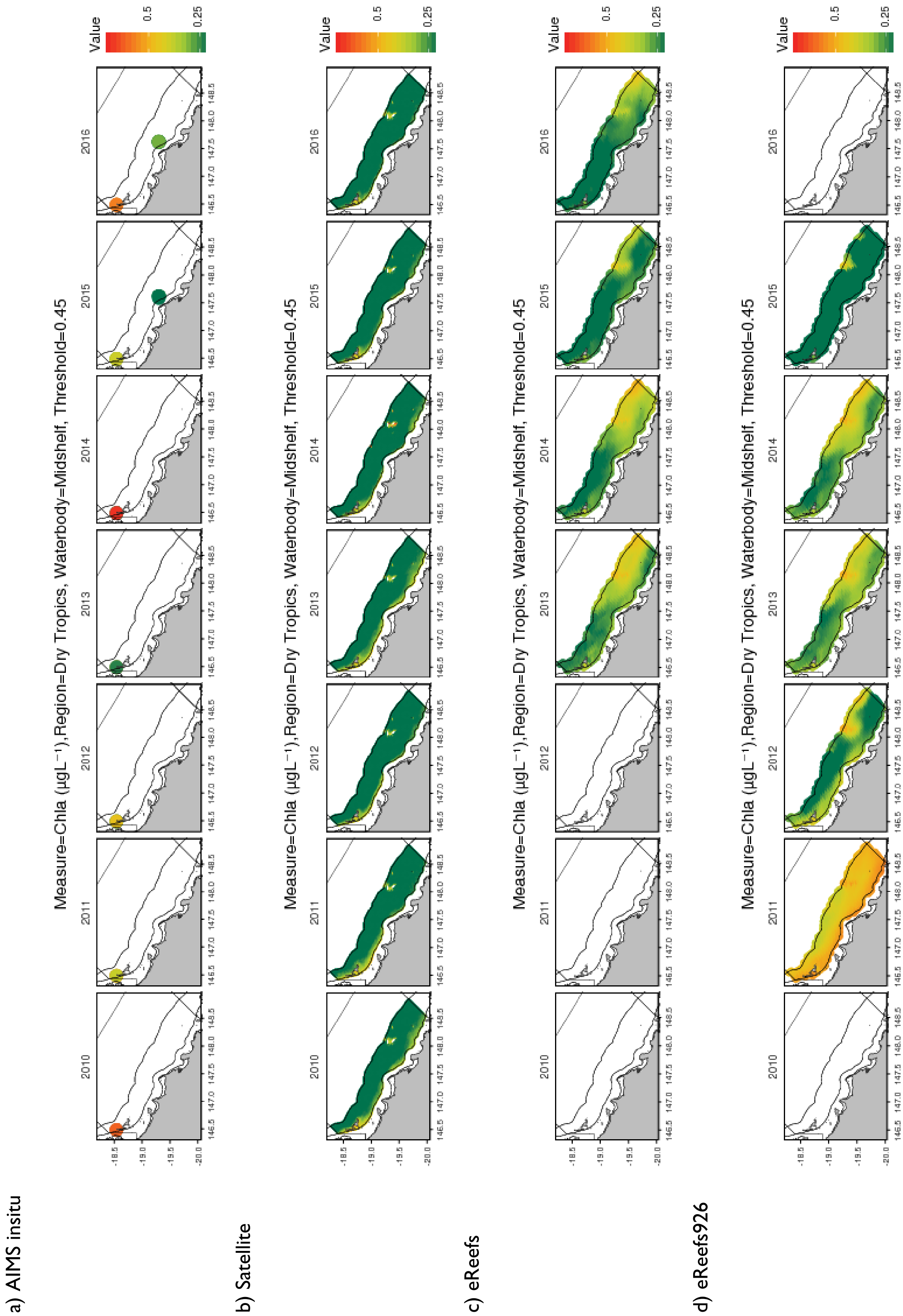


Figure 19: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 Chlorophyll-a (2009–2016) for the Dry Tropics Midshelf Zone.

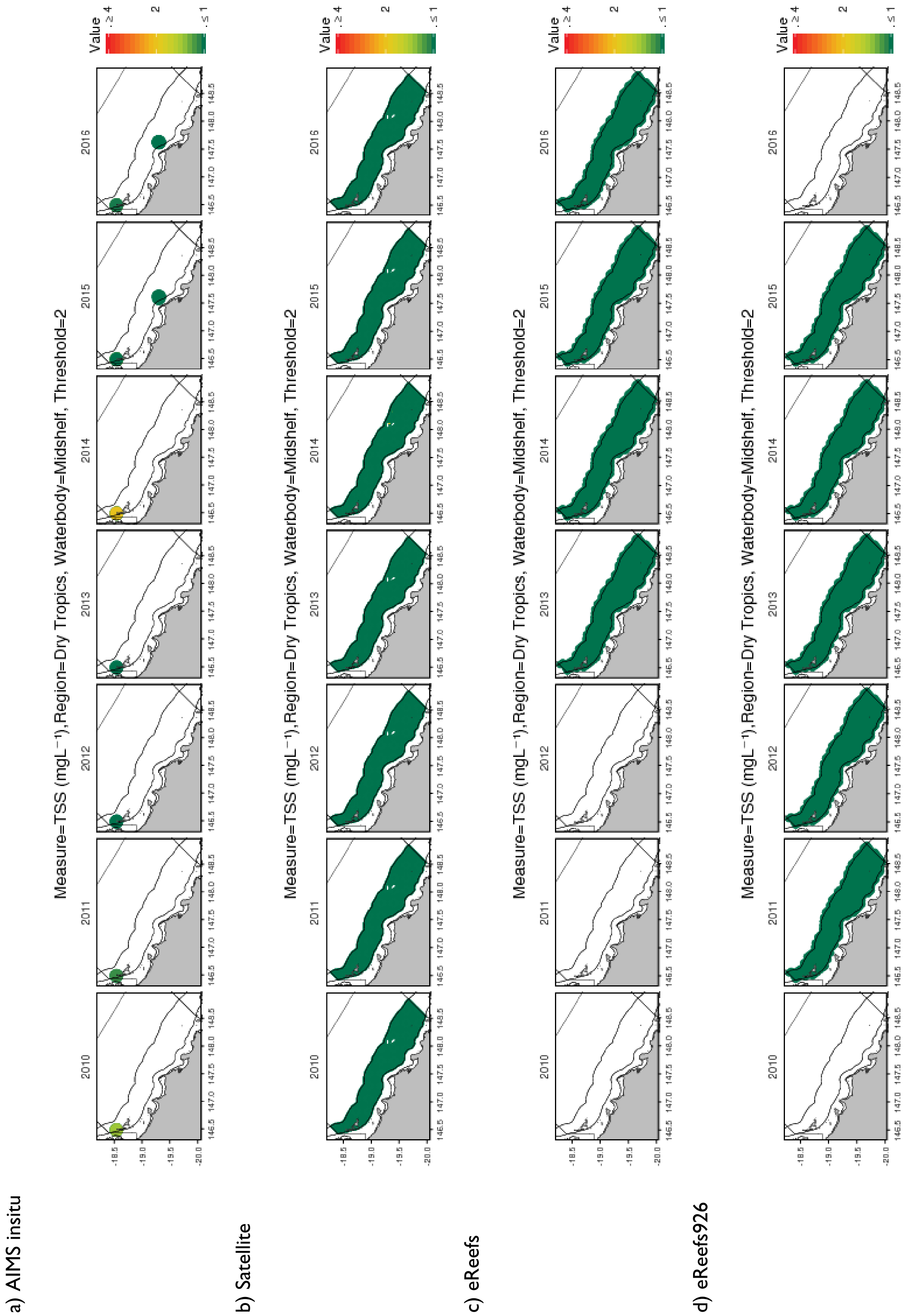


Figure 20: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 TSS (2009–2016) for the Dry Tropics Midshelf Zone.

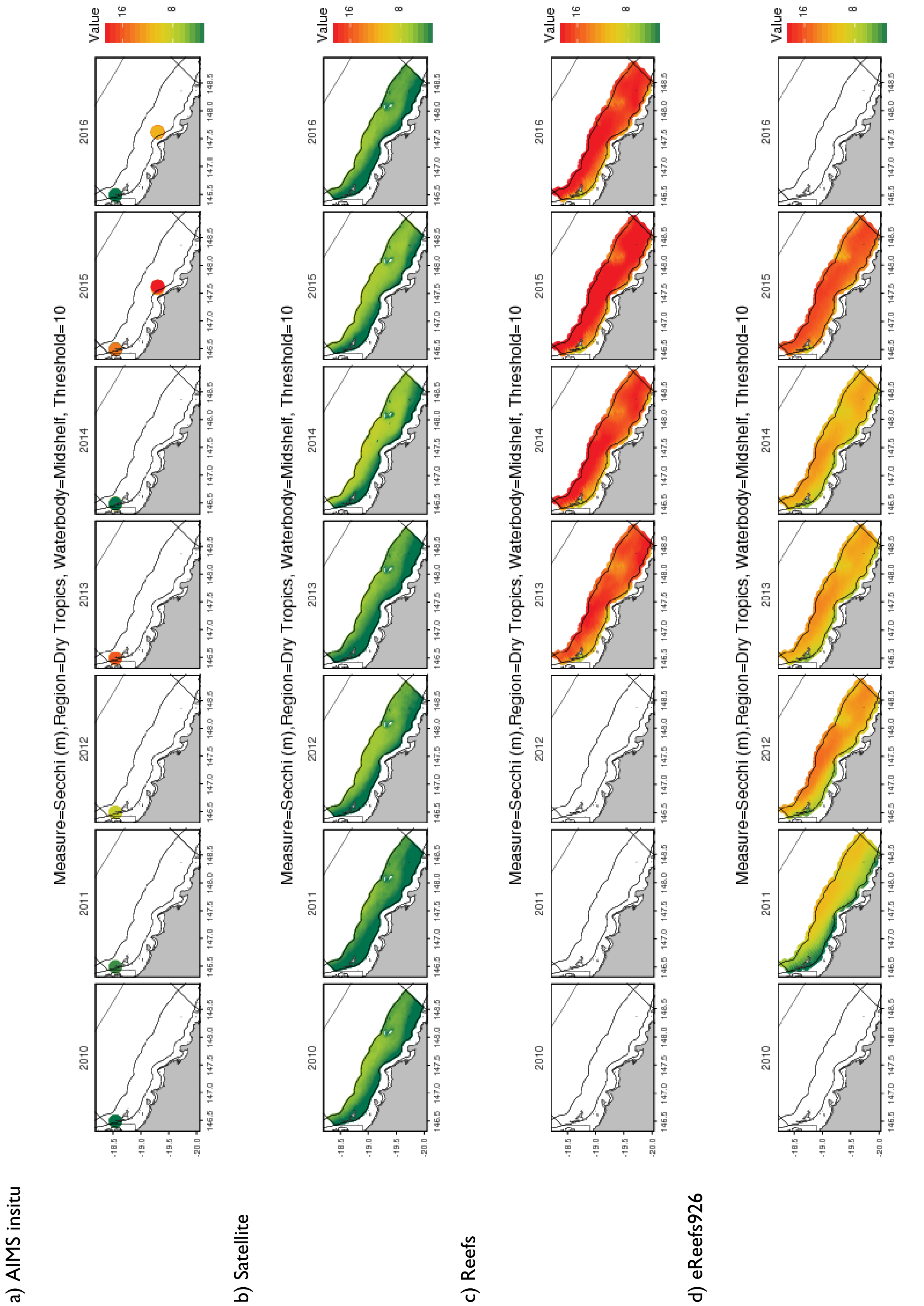
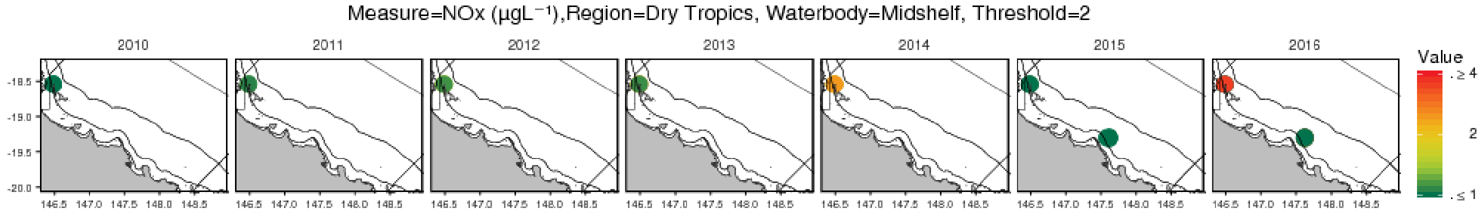
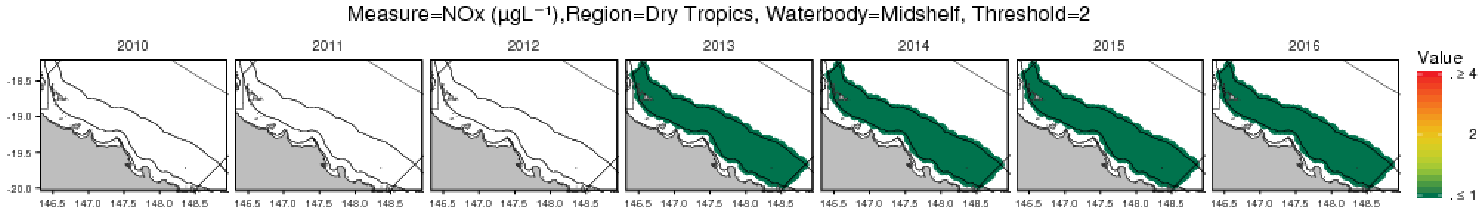


Figure 21: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 Secchi depth (2009–2016) for the Dry Tropics Midshelf Zone.

a) AIMS insitu



b) Reefs



c) eReefs926

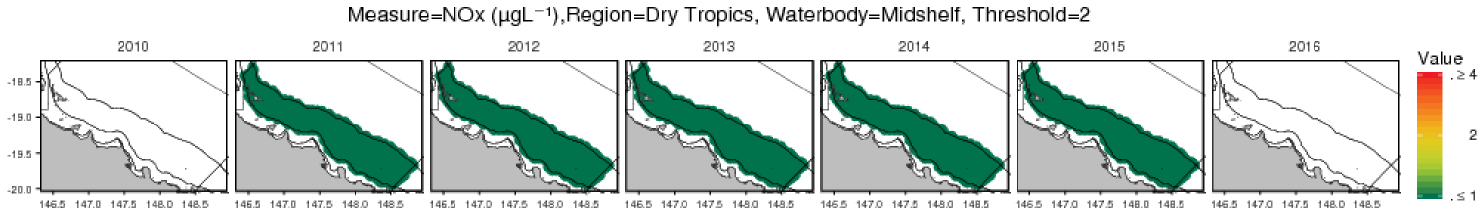


Figure 22: Spatial distribution of observed a) AIMS insitu, b) eReefs and c) eReefs926 NOx (2009–2016) for the Dry Tropics Midshelf Zone.

4.5 Comparison of data sources

Ensuring that the data underpinning the metric calculation is fit-for-purpose is a critical part of the process, especially if multiple data sources for a specific indicator are to be aggregated as part of this calculation. For example, successful aggregation of Chlorophyll-a as modelled by the eReefs BGC with Chlorophyll-a as extracted from satellite reflectance data (optical properties) will largely depend on the underlying compatibility of these two sources. Moreover, further combining with far more sparse and irregular sources (such as AIMS insitu Chlorophyll-a samples) relies on general patterns of spatial and temporal autocorrelation being present across the more dense data sources so as to facilitate a contagious projection of sparse data across the denser layers.

The project's recommendation is not to aggregate, for each indicator, independent streams of data. Instead the preferred approach is to assimilate satellite reflectance information into the eReefs BGC model and to rely on in situ measurements for verification of the model performance.

It is worthwhile noting that there is no single point of truth as the sparse in situ sampling does not account for the dynamic nature of the receiving environment, both temporally and spatially. It is however possible to compare different measurement methods at a high level.

These five different sources (Satellite, eReefs, eReefs926, AIMS Insitu and AIMS FLNTU loggers) were all collected at different spatio-temporal resolutions. Specifically:

- the Satellite data are collected on a 1km grid on a daily basis, however there are many gaps in the time series of each cell due to cloud cover and other issues that affect the reliability of observations.
- the eReefs data are collected on a 4km grid on a daily basis without any time series gaps between 2013 and 2016
- the eReefs926 data are collected on a 4km grid on a daily basis without any time series gaps between 2011 and 2014
- the AIMS Insitu samples are collected from specific sampling sites (28-32 throughout the GBR) and on an infrequent basis (approx. 3-4 times per year although more frequently in later years). Furthermore, with the exception of relatively recently, the majority of samples were collected in the dry season and thus these samples could be biased towards long term water quality trends rather than short-term pulses.
- the AIMS FLNTU logger data are deployed at a subset (16) of the AIMS Insitu sampling locations and record measurements every 10 minutes (although there are frequent gaps due to instrument failure).

The AIMS Insitu sampling locations are strategically positioned so as to generally represent transects away from major rivers discharging into the GBR. As such, they likely represent biased estimates of the water parameters of the surrounding water bodies. Nevertheless, the observed data are direct measurements of a range of parameters considered to be important measures of water quality and are therefore considered to be relatively accurate estimates of the true state - albiet for a potentially narrow spati-temporal window. By contrast, the Satellite data represent indirect proxies for some of these parameters (Chlorophyll-a, Total Suspended Solids and Secchi Depth) and similarly, the eReefs data are indirect modelled estimates simulated from a deterministic manifestation of a conceptual model. Hence, to gauge the accuracy of the Satellite and eReefs data (and thus inform qualitative confidence), time series and spatial patterns in the Satellite and eReefs observations were compared to the AIMS Insitu observations.

The disparate spatio-temporal resolutions of the data sources present substantial challenges for extracting comparable data. For example, the proximity of AIMS Insitu samples to reefs and the spatial resolution (1km or 4km grid) frequently results in an inability to obtain matching spatial location for all three sources. Furthermore, gaps in the Satellite time series frequently prevent matching Satellite data to the same day as AIMS Insitu sampling.

The degree to which the discrete AIMS Insitu samples reflect space and time around the actual sampling sites/times is largely unknown. That is, we don't know how broadly representative the direct observations are. Consequently, it is difficult to estimate how broadly to filter the Satellite and eReefs data in space and time around the AIMS Insitu sampling events in order to generate comparable data. The 'best' breadth is likely to be a compromise between data availability (time limited for Satellite and space limited for eReefs)

Figures 23 and 24 illustrate the spatial distribution of Satellite and eReefs grid cell centroid locations relative to the AIMS Insitu sampling locations. The different color spokes denote distance categories (red: <1km, purple: <5km) from the AIMS Insitu data.

The approach we took was to extract all observations within a specific series of spatio-temporal windows or neighbourhoods from which we could calculate a range of association and correspondence (such as RMSE and R^2) metrics (see Tables 8, ??, ?? & ??). Tables ??, ?? & ?? document the top 5 ranked (according to RMSE, MAE and MAPE respectively) spatio-temporal lag associations between Satellite/eReefs data and AIMS Insitu data.

Table 8: Association and correspondence metrics between Satellite/eReefs observations ($\hat{\theta}_i$) and AIMS Niskin observations (θ_i). Similar calculations can be performed on model residuals.

Metric	Description	Formulation
RMSE	Root Mean Square Error - is a measure of accuracy	$RMSE(\hat{\theta}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}$
MAE	Mean Absolute Error - is a measure of accuracy	$MAE(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\lvert(\hat{\theta}_i - \theta_i)\rvert$
MAPE	Mean Absolute Percentage Error - is a measure of accuracy expressed as a percentage of AIMS insitu samples	$MAPE(\hat{\theta}) = \frac{100}{n}\sum_{i=1}^{n}\frac{\hat{\theta}_i - \theta_i}{\theta_i}$

Whilst it is well established that water quality parameters can be highly varied over time and space, even approximate degrees of spatio-temporal autocorrelation for these parameters remains largely unknown. Nevertheless, We might expect that observations from different sources collected at similar locations and at similar times should be more similar to one another than they are to more distal observations. Furthermore, whilst the absolute values derived from different sources might not be exactly the same, we should expect a reasonable degree of correlation between the sources. Given these two positions (that observations should be autocorrelated and that different sources should be correlated), we should expect that the degree of correlation between the different sources for a given measure should be strongest for observation pairs closer together in space and time.

Tables 9 – 11 tabulate the association and correspondence metrics between the AIMS insitu samples and either the Satellite or eReefs data for each Measure. Irrespective of the association metric (RMSE, MAE or MAPE), closest associations with AIMS insitu observations tend to occur at shorter spatial distances for eReefs data than Satellite data, yet the opposite is apparent for temporal lags. We might have expected that associations would be strongest proximal (in both time and space) to the AIMS insitu samples and associations to weaken in some sort of multidimentional decaying pattern with increasing separation. Such a pattern would permit relatively straight forward integration of the AIMS insitu observational data into the Satellite or eReefs layers[2] However this is not the case and thus it is very difficult to formulate an integration routine that does more than just update a very limited number of points in space and time.

The other rationale for exploring the spatio-temporal associations between AIMS insitu data and Satellite/eReefs data is to be able to determine the an optimal temporal lag and spatial distance for making comparisons of trends. Given that AIMS insitu data are in some respects considered the more accurate (albeit limited in the degree to which they more broadly represent space and time around the samples), a comparison of the general temporal trends of each source should give some idea of the relative accuracy of the sources of indirect measurements (Satellite and eReefs). Figures. 25 – 28 illustrate the temporal patterns of Chlorophyll-a, TSS, Secchi depth and NOx for each source (AIMS insitu, AIMS FLNTU, Satellite, eReefs and eReefs926) for each of the AIMS insitu sampling locations. The background fills of the site titles are colored according to water body (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

All sources of data are typically most variable at Enclosed Coastal sites and substantially less variable at Midshelf sites. Moreover, the alignment of trends also appears to be substantially better at Midshore sites. Enclosed Coastal and Open Coastal sites are closer to the coasts and in particular, closer to major sources of discharge (as intended by the AIMS Water Quality MMP) whereby water conditions are subject to more extreme fluctuations that result in conditions varying rapidly in time and space. Moreover, these sites are likely to be in shallower water or water whose depth is relatively heterogeneous. As a result, data pooled within a 5km radius might represent a substantially different body of water than that represented by the AIMS insitu point sources. By contrast, the conditions represented within a 5km radius at Midshelf sites are likely to be more homogeneous and thereby resulting in a fairer comparison.

[2]Having a robust and consistent pattern of spatial and temporal autocorrelation would allow us to model the expected value of AIMS insitu data at unobserved locations.

Notwithstanding the disparity in fairness between different water bodies as a result of how well the various sources represent spatial and temporal envelopes, it is unlikely that either the eReefs models or Satellite data are going to provide accurate estimates for Enclosed Coastal water bodies. However, the accuracy for Midshelf and Offshore are likely to be sufficient.

Table 9: Top five ranked AIMS Niskin vs Satellite/eReefs observation association metrics (RMSE: root mean square error, MAE: mean absolute error, MAPE: mean percent error, Value: regression slope, residual.RMSE: residual root mean square error, residual.MAE: residual mean absolute error, R2.marginal: R^2 marginalized over sites, R2.conditional: R^2 conditional on sites) per Measure per source (Satellite, eReefs) for spatial/temporal lags. Rows ranked and filtered based on RMSE. Dist and Lag represent spatial (km) and temporal (days) lags.

Source	Measure	Dist	Lag	RMSE	MAE	MAPE	Value	Std.Error	DF	t.value	p.value	residual.RMSE	residual.MAE	R2.marginal	R2.conditional
Satellite	chl	8.00	6.00	0.33	0.22	0.69	0.42	0.04	566.00	11.43	0.00	0.22	0.14	0.10	0.66
Satellite	chl	9.00	6.00	0.33	0.22	0.69	0.42	0.04	566.00	11.37	0.00	0.22	0.14	0.09	0.67
Satellite	chl	6.00	6.00	0.33	0.22	0.69	0.43	0.04	566.00	11.54	0.00	0.22	0.14	0.10	0.65
Satellite	chl	10.00	6.00	0.33	0.22	0.69	0.42	0.04	566.00	11.30	0.00	0.22	0.13	0.09	0.67
Satellite	chl	11.00	6.00	0.33	0.22	0.69	0.42	0.04	566.00	11.27	0.00	0.22	0.13	0.09	0.67
eReefs	chl	1.00	5.00	0.34	0.24	0.44	0.13	0.03	96.00	3.67	0.00	0.10	0.08	0.08	0.48
eReefs	chl	1.00	4.00	0.34	0.24	0.44	0.14	0.04	96.00	3.85	0.00	0.10	0.08	0.09	0.48
eReefs	chl	1.00	6.00	0.34	0.24	0.45	0.12	0.03	96.00	3.63	0.00	0.09	0.08	0.08	0.49
eReefs	chl	1.00	3.00	0.34	0.24	0.45	0.16	0.04	96.00	3.76	0.00	0.12	0.09	0.09	0.42
eReefs	chl	1.00	7.00	0.34	0.24	0.45	0.11	0.03	96.00	3.46	0.00	0.09	0.07	0.07	0.50
Satellite	nap	4.00	1.00	1.65	0.90	1.02	0.48	0.03	432.00	16.60	0.00	1.15	0.54	0.40	0.45
Satellite	nap	1.00	1.00	1.66	0.87	1.08	0.54	0.04	358.00	14.58	0.00	1.30	0.57	0.38	0.45
Satellite	nap	4.00	0.00	1.67	0.87	1.21	0.51	0.04	225.00	13.99	0.00	1.17	0.52	0.45	0.49
Satellite	nap	3.00	1.00	1.70	0.91	0.97	0.47	0.03	427.00	15.41	0.00	1.19	0.55	0.37	0.43
Satellite	nap	3.00	0.00	1.73	0.90	1.11	0.54	0.04	214.00	13.28	0.00	1.23	0.57	0.43	0.53
eReefs	nap	5.00	3.00	2.07	1.18	0.73	0.12	0.02	239.00	6.20	0.00	0.57	0.38	0.13	0.16
eReefs	nap	5.00	4.00	2.07	1.17	0.73	0.11	0.02	239.00	5.51	0.00	0.56	0.39	0.11	0.16
eReefs	nap	4.00	3.00	2.08	1.17	0.70	0.11	0.02	239.00	5.78	0.00	0.53	0.37	0.12	0.18
eReefs	nap	4.00	4.00	2.08	1.16	0.70	0.09	0.02	239.00	5.03	0.00	0.54	0.39	0.09	0.16
eReefs	nap	5.00	2.00	2.08	1.18	0.74	0.12	0.02	239.00	6.00	0.00	0.57	0.39	0.13	0.16
Satellite	sd	5.00	2.00	4.47	3.38	0.44	0.11	0.01	463.00	11.77	0.00	0.55	0.42	0.24	0.54
Satellite	sd	4.00	2.00	4.48	3.38	0.44	0.11	0.01	462.00	11.71	0.00	0.56	0.42	0.24	0.52
Satellite	sd	3.00	2.00	4.48	3.39	0.44	0.12	0.01	455.00	11.73	0.00	0.57	0.42	0.25	0.51
Satellite	sd	11.00	2.00	4.48	3.37	0.44	0.11	0.01	470.00	11.65	0.00	0.53	0.41	0.20	0.61
Satellite	sd	12.00	2.00	4.48	3.37	0.44	0.11	0.01	470.00	11.65	0.00	0.53	0.41	0.20	0.61
eReefs	sd	4.00	1.00	13.13	11.31	2.37	1.23	0.12	196.00	10.39	0.00	6.47	4.92	0.35	0.37
eReefs	sd	4.00	2.00	13.29	11.68	2.49	1.14	0.11	196.00	9.89	0.00	6.10	4.75	0.34	0.37
eReefs	sd	5.00	1.00	13.46	11.62	2.36	1.29	0.12	185.00	10.81	0.00	6.61	5.12	0.38	0.39
eReefs	sd	6.00	1.00	13.53	11.69	2.37	1.30	0.13	185.00	10.40	0.00	6.43	4.96	0.38	0.41
eReefs	sd	5.00	2.00	13.66	12.02	2.48	1.18	0.12	185.00	10.20	0.00	6.30	5.02	0.36	0.37

Table 10: Top five ranked AIMS Niskin vs Satellite/eReefs observation association metrics (RMSE: root mean square error, MAE: mean absolute error, MAPE: mean percent error, Value: regression slope, residual.RMSE: residual root mean square error, residual.MAE: residual mean absolute error, R2.marginal: R^2 marginalized over sites, R2.conditional: R^2 conditional on sites) per Measure per source (Satellite, eReefs) for spatial/temporal lags. Rows ranked and filtered based on MAE. Dist and Lag represent spatial (km) and temporal (days) lags.

Source	Measure	Dist	Lag	RMSE	MAE	MAPE	Value	Std.Error	DF	t.value	p.value	residual.RMSE	residual.MAE	R2.marginal	R2.conditional
Satellite	chl	10.00	0.00	0.38	0.21	0.64	0.82	0.08	253.00	9.99	0.00	0.33	0.17	0.27	0.37
Satellite	chl	11.00	0.00	0.38	0.21	0.65	0.81	0.08	254.00	9.89	0.00	0.33	0.17	0.26	0.38
Satellite	chl	12.00	0.00	0.38	0.21	0.65	0.81	0.08	254.00	9.89	0.00	0.33	0.17	0.26	0.38
Satellite	chl	4.00	0.00	0.38	0.21	0.65	0.91	0.08	226.00	10.82	0.00	0.33	0.17	0.32	0.44
Satellite	chl	9.00	0.00	0.39	0.21	0.64	0.84	0.09	250.00	9.86	0.00	0.35	0.17	0.27	0.36
eReefs	chl	3.00	5.00	0.34	0.23	0.43	0.14	0.02	221.00	6.09	0.00	0.09	0.08	0.11	0.46
eReefs	chl	3.00	6.00	0.34	0.23	0.43	0.13	0.02	221.00	6.09	0.00	0.09	0.07	0.11	0.46
eReefs	chl	3.00	4.00	0.34	0.23	0.43	0.14	0.02	221.00	6.09	0.00	0.10	0.08	0.11	0.45
eReefs	chl	3.00	7.00	0.35	0.23	0.43	0.12	0.02	221.00	5.88	0.00	0.09	0.07	0.10	0.46
eReefs	chl	4.00	5.00	0.34	0.23	0.43	0.13	0.02	239.00	5.98	0.00	0.09	0.07	0.10	0.46
Satellite	nap	4.00	0.00	1.67	0.87	1.21	0.51	0.04	225.00	13.99	0.00	1.17	0.52	0.45	0.49
Satellite	nap	1.00	1.00	1.66	0.87	1.08	0.54	0.04	358.00	14.58	0.00	1.30	0.57	0.38	0.45
Satellite	nap	4.00	1.00	1.65	0.90	1.02	0.48	0.03	432.00	16.60	0.00	1.15	0.54	0.40	0.45
Satellite	nap	3.00	0.00	1.73	0.90	1.11	0.54	0.04	214.00	13.28	0.00	1.23	0.57	0.43	0.53
Satellite	nap	3.00	1.00	1.70	0.91	0.97	0.47	0.03	427.00	15.41	0.00	1.19	0.55	0.37	0.43
eReefs	nap	4.00	4.00	2.08	1.16	0.70	0.09	0.02	239.00	5.03	0.00	0.54	0.39	0.09	0.16
eReefs	nap	4.00	3.00	2.08	1.17	0.70	0.11	0.02	239.00	5.78	0.00	0.53	0.37	0.12	0.18
eReefs	nap	4.00	2.00	2.09	1.17	0.72	0.11	0.02	239.00	5.52	0.00	0.55	0.38	0.11	0.18
eReefs	nap	5.00	4.00	2.07	1.17	0.73	0.11	0.02	239.00	5.51	0.00	0.56	0.39	0.11	0.16
eReefs	nap	5.00	3.00	2.07	1.18	0.73	0.12	0.02	239.00	6.20	0.00	0.57	0.38	0.13	0.16
Satellite	sd	11.00	2.00	4.48	3.37	0.44	0.11	0.01	470.00	11.65	0.00	0.53	0.41	0.20	0.61
Satellite	sd	12.00	2.00	4.48	3.37	0.44	0.11	0.01	470.00	11.65	0.00	0.53	0.41	0.20	0.61
Satellite	sd	10.00	2.00	4.48	3.37	0.44	0.11	0.01	470.00	11.67	0.00	0.53	0.41	0.20	0.61
Satellite	sd	4.00	2.00	4.48	3.38	0.44	0.11	0.01	462.00	11.71	0.00	0.56	0.42	0.24	0.52
Satellite	sd	9.00	2.00	4.49	3.38	0.44	0.11	0.01	468.00	11.89	0.00	0.53	0.41	0.22	0.60
eReefs	sd	4.00	1.00	13.13	11.31	2.37	1.23	0.12	196.00	10.39	0.00	6.47	4.92	0.35	0.37
eReefs	sd	1.00	1.00	14.04	11.52	2.73	1.10	0.29	85.00	3.86	0.00	7.61	5.43	0.15	0.22
eReefs	sd	1.00	2.00	13.71	11.58	2.79	1.12	0.26	85.00	4.31	0.00	6.87	5.36	0.18	0.26
eReefs	sd	5.00	1.00	13.46	11.62	2.36	1.29	0.12	185.00	10.81	0.00	6.61	5.12	0.38	0.39
eReefs	sd	4.00	2.00	13.29	11.68	2.49	1.14	0.11	196.00	9.89	0.00	6.10	4.75	0.34	0.37

Table 11: Top five ranked AIMS Niskin vs Satellite/eReefs observation association metrics (RMSE: root mean square error, MAE: mean absolute error, MAPE: mean percent error, Value: regression slope, residual.RMSE: residual root mean square error, residual.MAE: residual mean absolute error, R2.marginal: R^2 marginalized over sites, R2.conditional: R^2 conditional on sites) per Measure per source (Satellite, eReefs) for spatial/temporal lags. Rows ranked and filtered based on MAPE. Dist and Lag represent spatial (km) and temporal (days) lags.

Source	Measure	Dist	Lag	RMSE	MAE	MAPE	Value	Std.Error	DF	t.value	p.value	residual.RMSE	residual.MAE	R2.marginal	R2.conditional
Satellite	chl	4.00	2.00	0.37	0.21	0.62	0.64	0.05	508.00	12.12	0.00	0.30	0.15	0.18	0.48
Satellite	chl	3.00	2.00	0.37	0.21	0.63	0.67	0.05	501.00	12.20	0.00	0.30	0.15	0.19	0.46
Satellite	chl	2.00	2.00	0.35	0.21	0.63	0.63	0.05	492.00	12.64	0.00	0.27	0.15	0.19	0.54
Satellite	chl	8.00	0.00	0.41	0.21	0.64	0.87	0.09	248.00	9.86	0.00	0.36	0.17	0.27	0.34
Satellite	chl	10.00	0.00	0.38	0.21	0.64	0.82	0.08	253.00	9.99	0.00	0.33	0.17	0.27	0.37
eReefs	chl	3.00	6.00	0.34	0.23	0.43	0.13	0.02	221.00	6.09	0.00	0.09	0.07	0.11	0.46
eReefs	chl	4.00	6.00	0.34	0.23	0.43	0.12	0.02	239.00	6.03	0.00	0.09	0.07	0.10	0.47
eReefs	chl	3.00	5.00	0.34	0.23	0.43	0.14	0.02	221.00	6.09	0.00	0.09	0.08	0.11	0.46
eReefs	chl	2.00	6.00	0.35	0.23	0.43	0.13	0.02	195.00	5.72	0.00	0.09	0.07	0.11	0.45
eReefs	chl	3.00	7.00	0.35	0.23	0.43	0.12	0.02	221.00	5.88	0.00	0.09	0.07	0.10	0.46
Satellite	nap	3.00	2.00	1.76	0.95	0.90	0.35	0.02	500.00	15.62	0.00	0.94	0.50	0.31	0.50
Satellite	nap	2.00	2.00	1.81	0.96	0.91	0.35	0.02	491.00	14.78	0.00	0.97	0.50	0.27	0.52
Satellite	nap	7.00	2.00	1.88	1.00	0.93	0.34	0.02	514.00	13.50	0.00	1.04	0.54	0.22	0.52
Satellite	nap	8.00	2.00	1.88	1.01	0.93	0.33	0.02	514.00	13.35	0.00	1.03	0.54	0.21	0.54
Satellite	nap	9.00	2.00	1.88	1.01	0.93	0.33	0.02	514.00	13.43	0.00	1.01	0.53	0.20	0.56
eReefs	nap	1.00	4.00	2.34	1.36	0.68	0.10	0.03	96.00	3.12	0.00	0.76	0.50	0.08	0.08
eReefs	nap	1.00	3.00	2.37	1.37	0.68	0.12	0.04	96.00	3.11	0.00	0.87	0.51	0.08	0.08
eReefs	nap	11.00	4.00	2.57	1.28	0.69	0.07	0.02	246.00	4.48	0.00	0.55	0.39	0.07	0.17
eReefs	nap	12.00	4.00	2.57	1.28	0.69	0.07	0.02	246.00	4.49	0.00	0.55	0.39	0.07	0.17
eReefs	nap	10.00	4.00	2.57	1.28	0.69	0.07	0.02	246.00	4.45	0.00	0.56	0.39	0.07	0.16
Satellite	sd	6.00	0.00	4.64	3.50	0.43	0.16	0.02	217.00	10.16	0.00	0.74	0.54	0.34	0.42
Satellite	sd	4.00	0.00	4.73	3.59	0.43	0.16	0.01	207.00	11.42	0.00	0.70	0.54	0.40	0.45
Satellite	sd	7.00	0.00	4.63	3.51	0.43	0.15	0.02	224.00	10.00	0.00	0.73	0.55	0.33	0.41
Satellite	sd	10.00	0.00	4.62	3.50	0.43	0.15	0.02	231.00	9.27	0.00	0.75	0.57	0.29	0.38
Satellite	sd	5.00	0.00	4.70	3.56	0.43	0.16	0.01	211.00	11.05	0.00	0.70	0.53	0.38	0.44
eReefs	sd	5.00	1.00	13.46	11.62	2.36	1.29	0.12	185.00	10.81	0.00	6.61	5.12	0.38	0.39
eReefs	sd	4.00	1.00	13.13	11.31	2.37	1.23	0.12	196.00	10.39	0.00	6.47	4.92	0.35	0.37
eReefs	sd	6.00	1.00	13.53	11.69	2.37	1.30	0.13	185.00	10.40	0.00	6.43	4.96	0.38	0.41
eReefs	sd	8.00	1.00	13.91	12.00	2.38	1.38	0.13	185.00	10.31	0.00	6.39	4.97	0.40	0.45
eReefs	sd	7.00	1.00	13.75	11.88	2.39	1.33	0.13	185.00	10.30	0.00	6.45	4.98	0.38	0.42

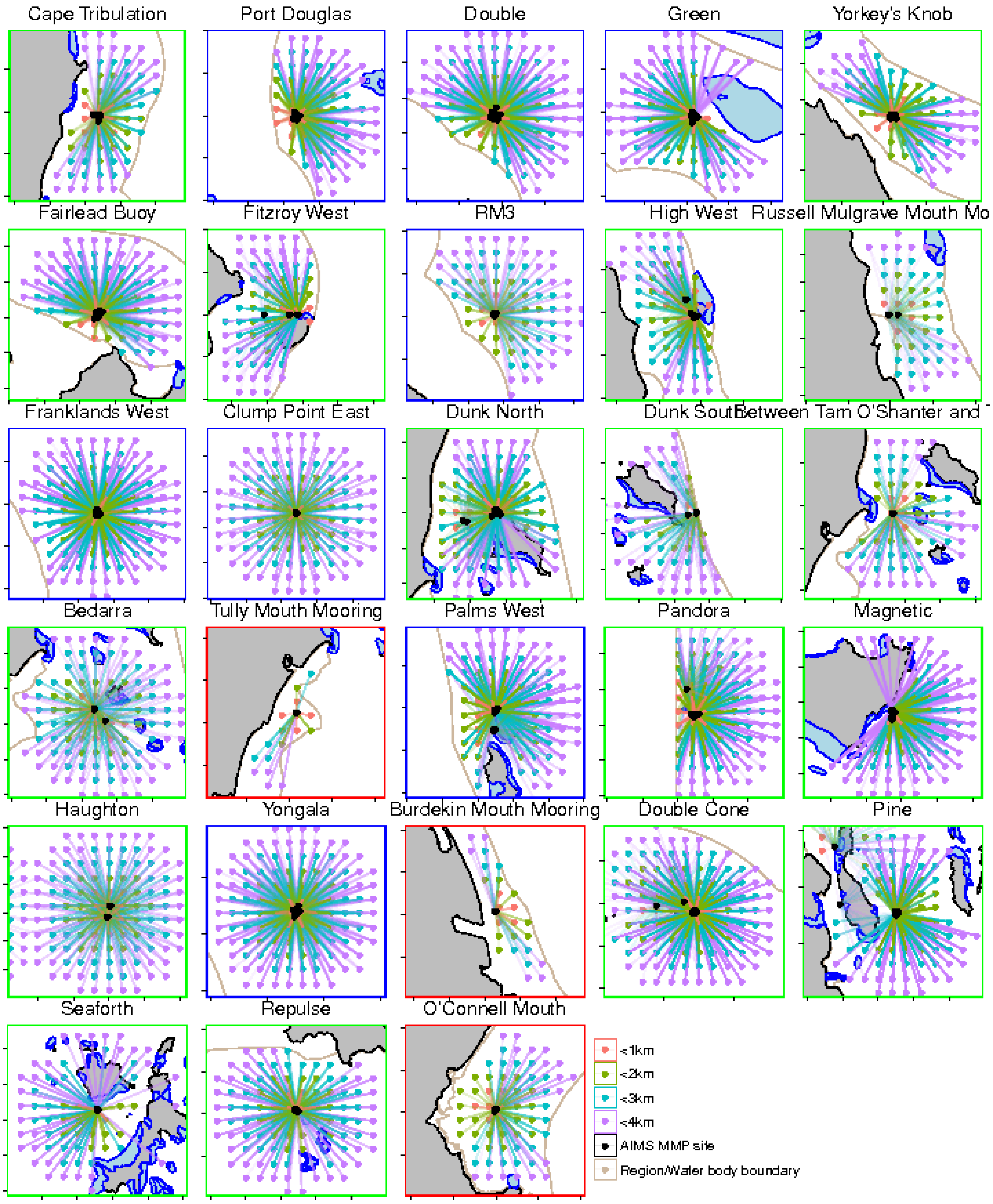


Figure 23: Location of Satellite cells within 5km of AIMS niskin samples. Panel borders represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

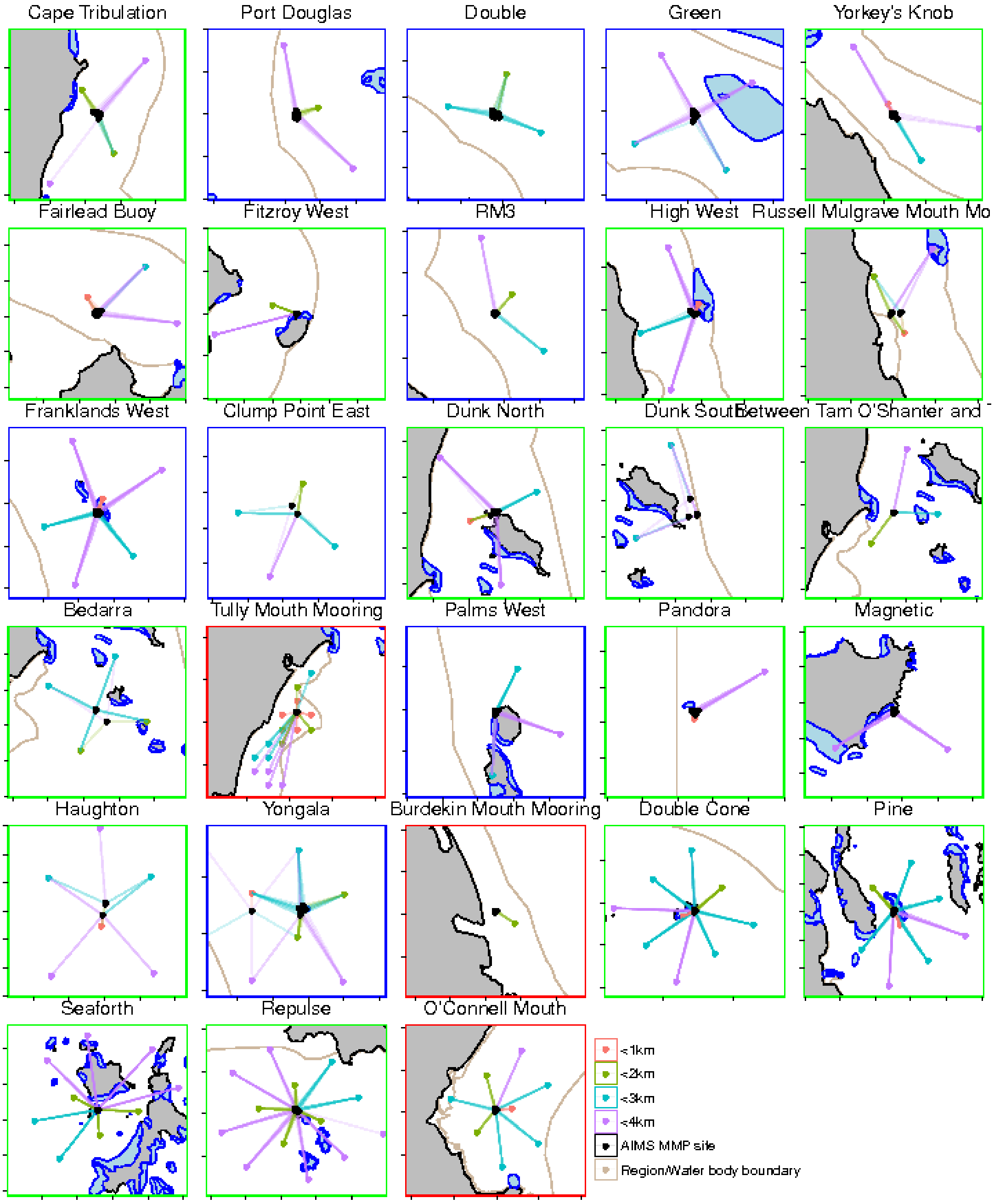


Figure 24: Location of eReefs cells within 5km of AIMS niskin samples. Panel borders represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

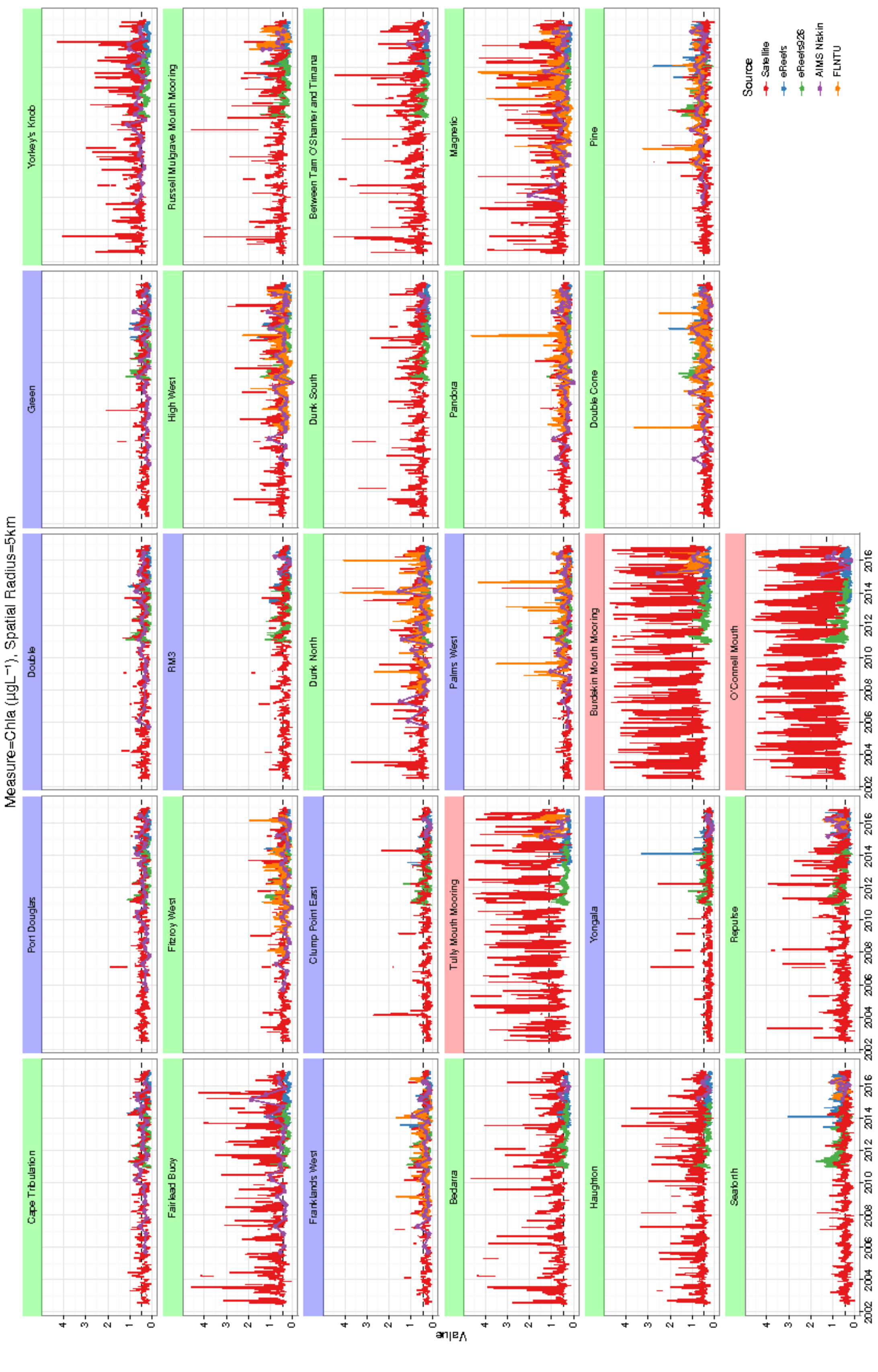


Figure 25: Temporal patterns in Chlorophyll-a within 5km of each AIMS MMP sampling site for eReefs, Satellite and AIMS insitu and FLNTU logger sources. Horizontal dashed line represents the guideline value. Title backgrounds represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

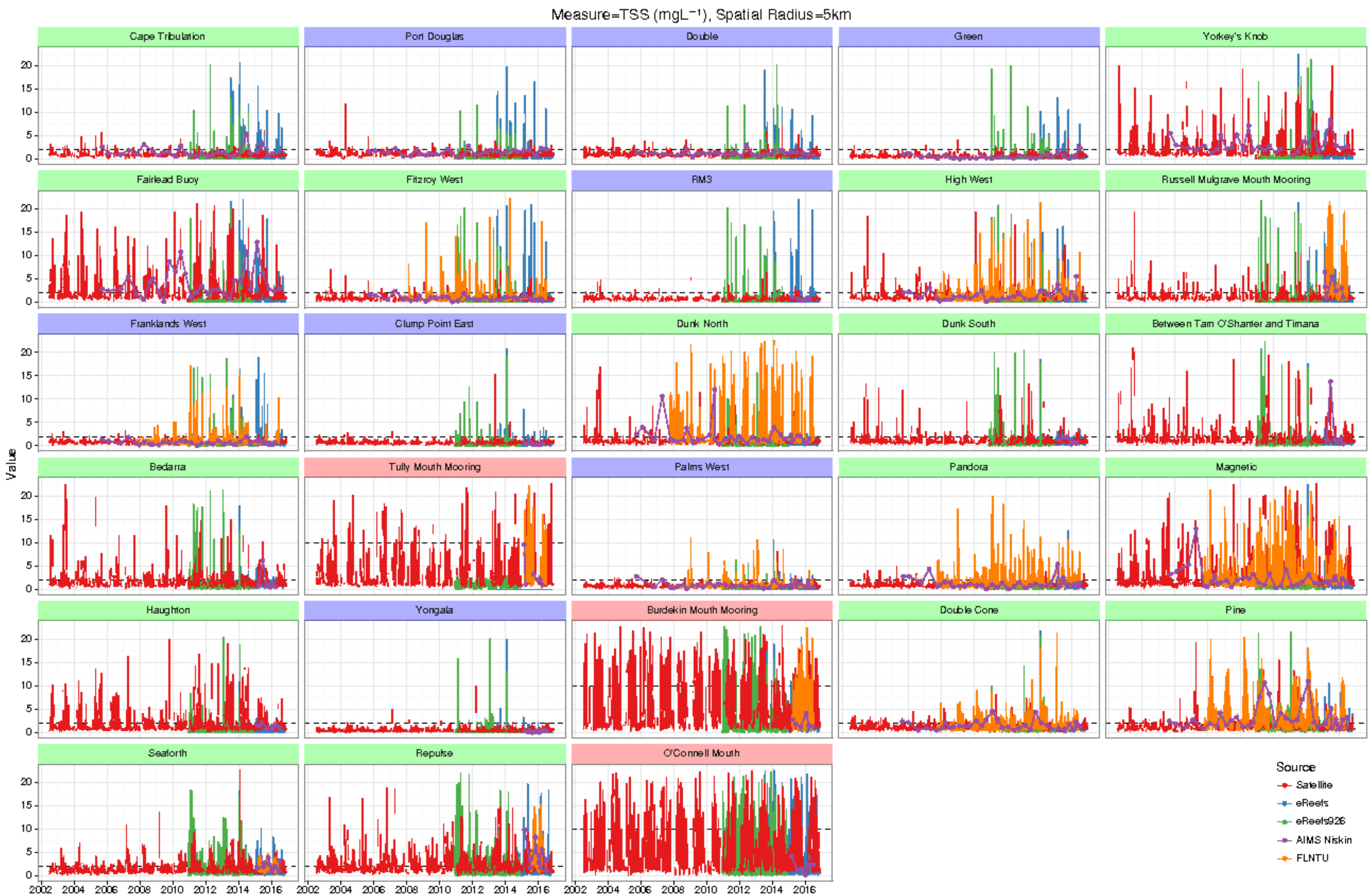


Figure 26: Temporal patterns in TSS within 5km of each AIMS MMP sampling site for eReefs, Satellite and AIMS insitu and FLNTU logger sources. Horizontal dashed line represents the guideline value. Title backgrounds represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

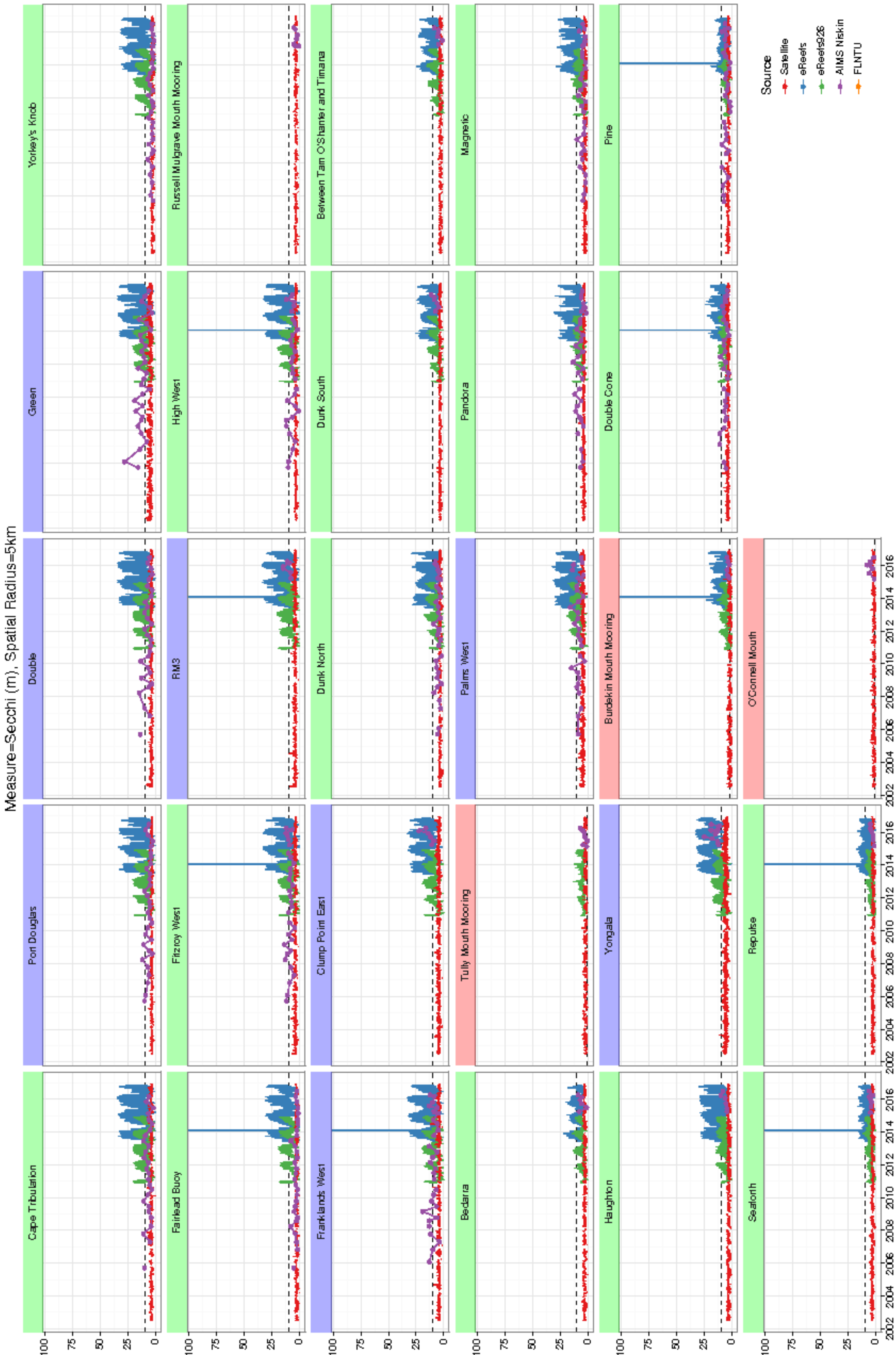


Figure 27: Temporal patterns in Secchi Depth within 5km of each AIMS MMP sampling site for eReefs, Satellite and AIMS insitu and FLNTU logger sources. Horizontal dashed line represents the guideline value. Title backgrounds represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

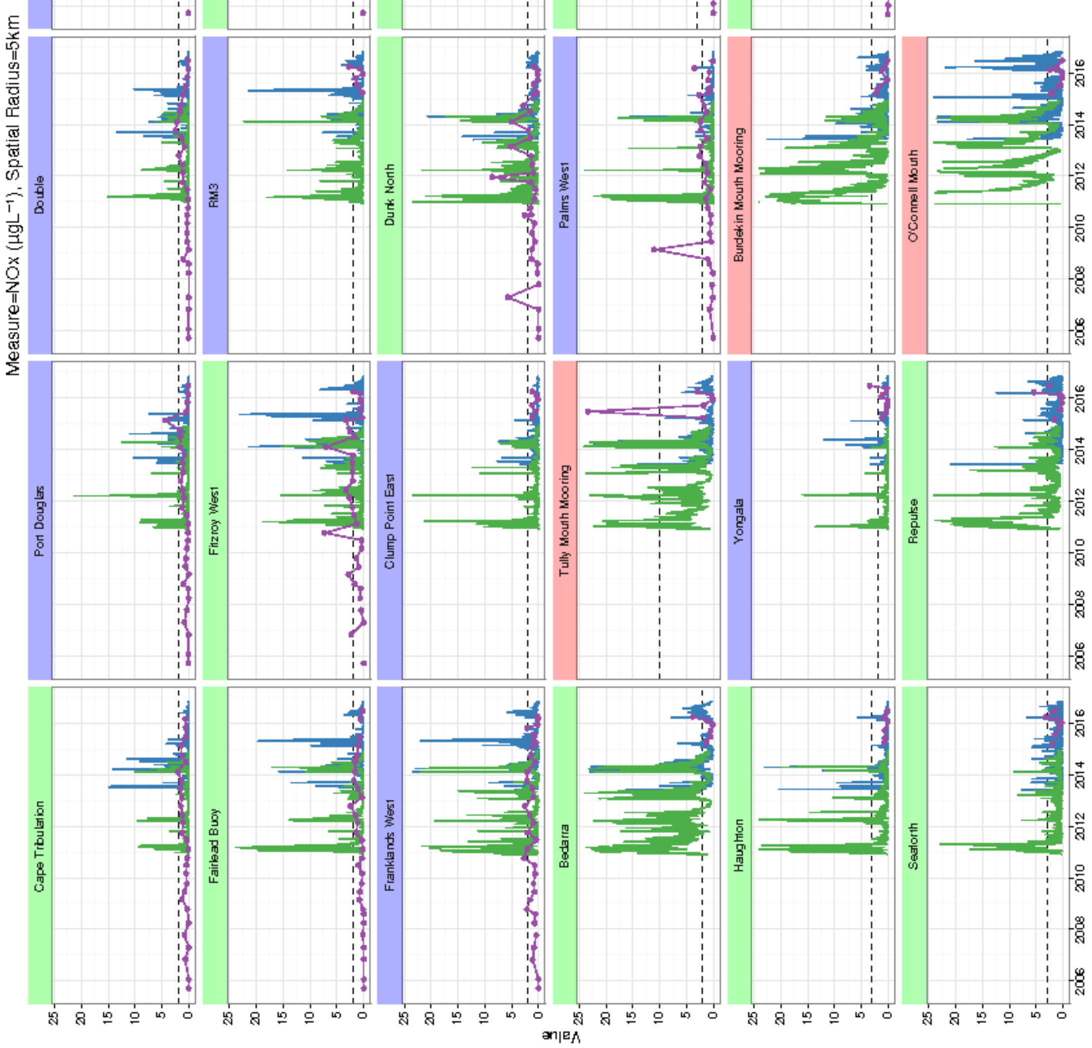


Figure 28: Temporal patterns in NOx within 5km of each AIMS MMP sampling site for eReefs, Satellite and AIMS insitu and FLNTU logger sources. Horizontal dashed line represents the guideline value. Title backgrounds represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

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).

$$
\begin{aligned}
F_1 &= 100.\left(\frac{Number\ of\ failed\ indicators}{Total\ number\ of\ indicators}\right)\\
F_2 &= 100.\left(\frac{Number\ of\ failed\ comparisons}{Total\ number\ of\ comparisons}\right)\\
F_3 &= \frac{100.E}{1+E}; \qquad E = \frac{\sum_{i=1}^{n} e_i}{n}; \qquad e_i = z_i.\left[\left(\frac{x_i}{benchmark_i}\right)^{\lambda_i} - 1\right]\\
z_i &= \begin{cases} 1 & \text{if ith comparison fails} \\ 0 & \text{otherwise} \end{cases}; \qquad \lambda_1 = \begin{cases} 1 & \text{If} < benchmark_i = \text{fail} \\ -1 & \text{If} > benchmark_i = \text{fail} \end{cases}\\
CCMEWQI &= 100 - \left(\frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732}\right)
\end{aligned}
$$

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.

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.

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 128) 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[3] There are two broad approaches to scaling (see Table 12):

[3] 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.

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. 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.

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.

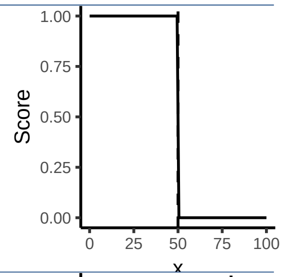
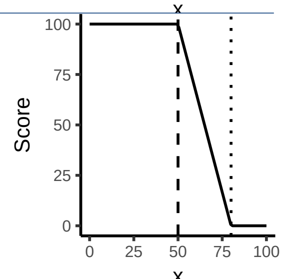
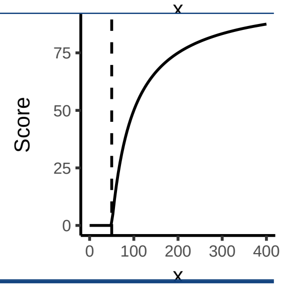
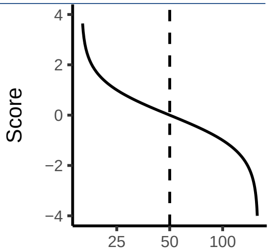
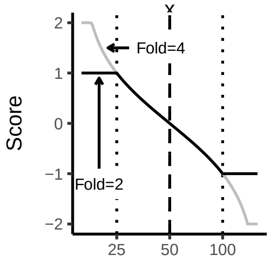
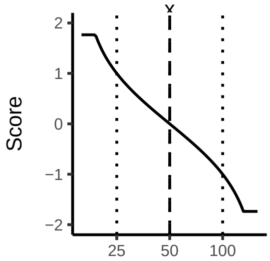
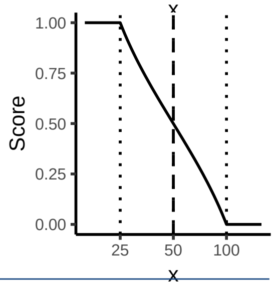
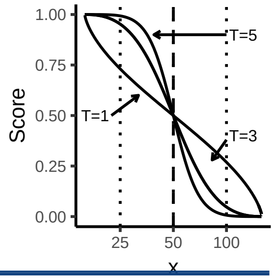
Method	Formulation	Response curve
Binary compliance	$Score_i = \begin{cases} 1 & \text{if } x_i \leq B_i \\ 0 & \text{if } x_i \text{ else} \end{cases}$	
Benchmark and WCS	$Score_i = \begin{cases} 100 & \text{if } x_i \leq B_i \\ 0 & \text{if } x_i \geq WCS_i \\ \left[1.0 - \left\lvert \frac{x_i - B_i}{WCS_i - B_i} \right\rvert\right].100 & \text{else} \end{cases}$	
Amplitude	$Score_i = \begin{cases} (\frac{x_i}{B_i})^{-1} & \text{if} > B_i = \text{fail} \\ (\frac{x_i}{B_i})^{1} & \text{if} < B_i = \text{fail} \end{cases}$ $Score_i = \frac{100 \times Score_i}{1 + Score_i}$	

Table 12: Report Card indexing methods, continued

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

5.2 Summary of adopted methodologies

5.3 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 29, 32 and 33, 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 29: 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

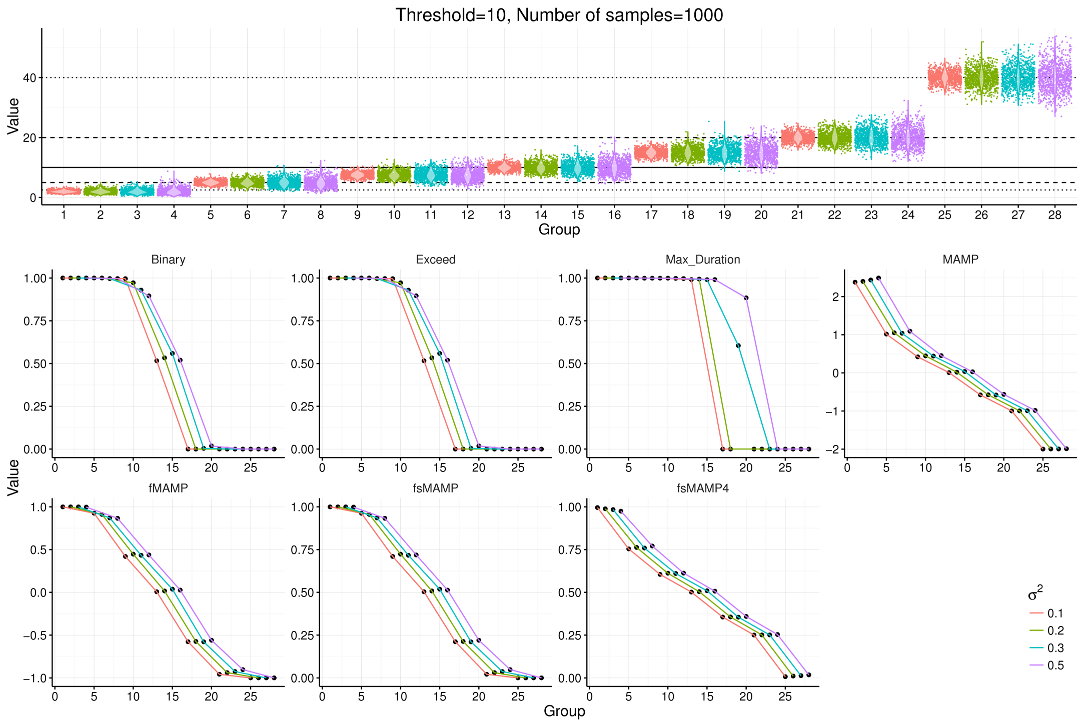


Figure 30: Simulated data and associated indices for threshold of 10 and very large sample sizes (R=1000).

indices to various combinations allowed the identification of the most appropriate and robust index calculation method.

When the number of samples and the relative sample variability is very large (e.g. fig. 29), 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. 29, 30 and 31), 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.33).

Over all 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.

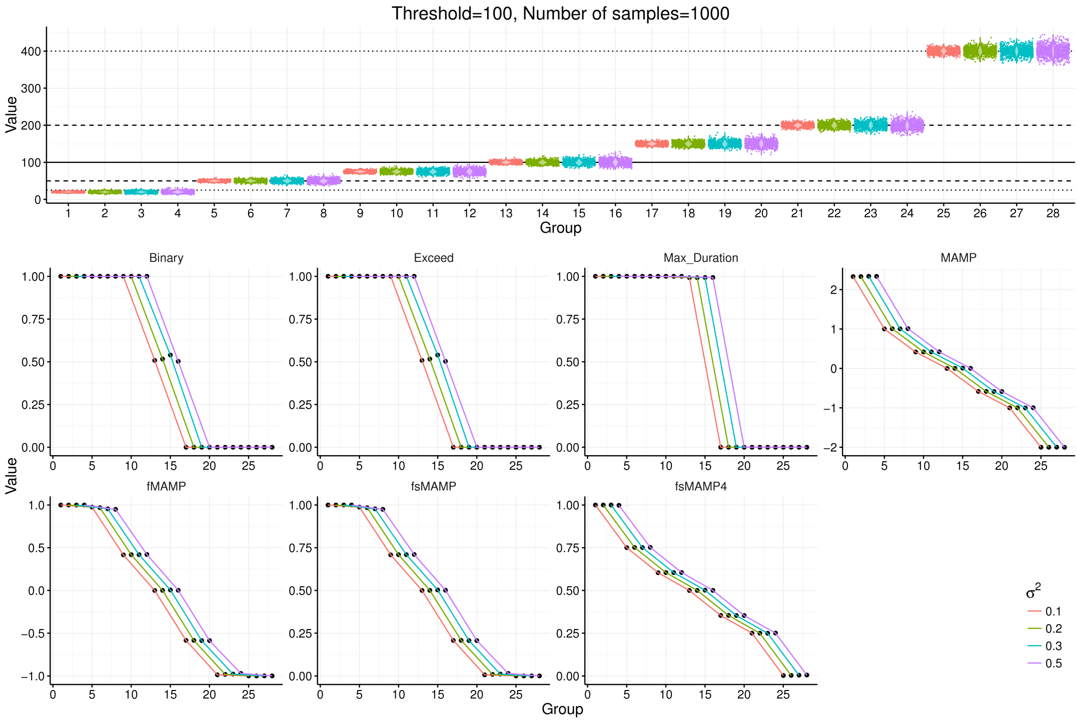


Figure 31: Simulated data and associated indices for threshold of 100 and very large sample sizes (R=1000).

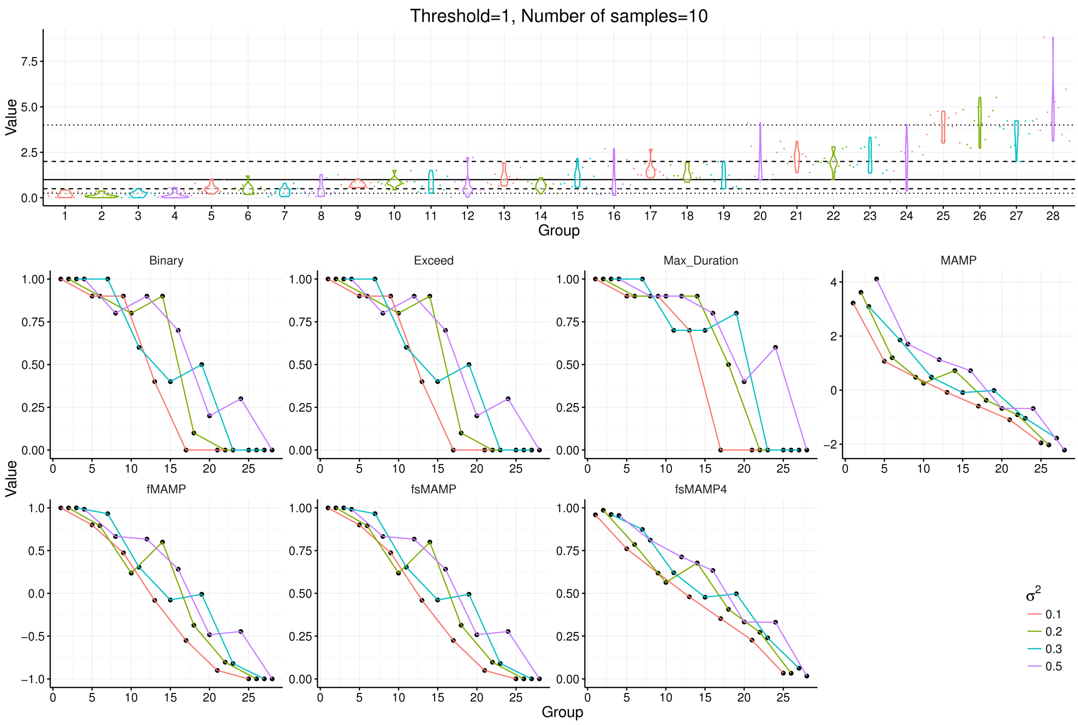


Figure 32: Simulated data and associated indices for threshold of 1 and large sample sizes (R=100).

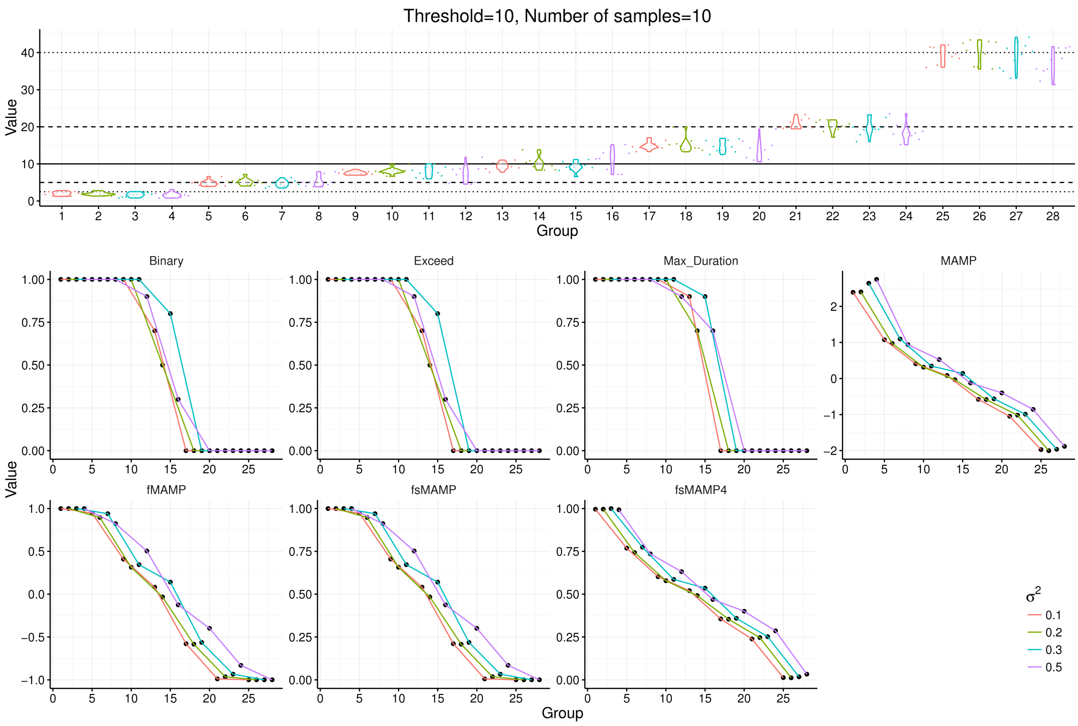


Figure 33: Simulated data and associated indices for threshold of 10 and small sample sizes (R=10).

- 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

5.4 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 expectations, there is very little (if any) variability and thus the realised utility of the parameter is low (or else the threshold is inappropriate).

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 in which the 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

5.4.1 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 5 (Chlorophyll-a from Wet Tropics, Open Coastal). Figures 34 – 38 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[4].

Inter and Intra annual variation is greatest in the Binary index method for each data Source[5]. 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.

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

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

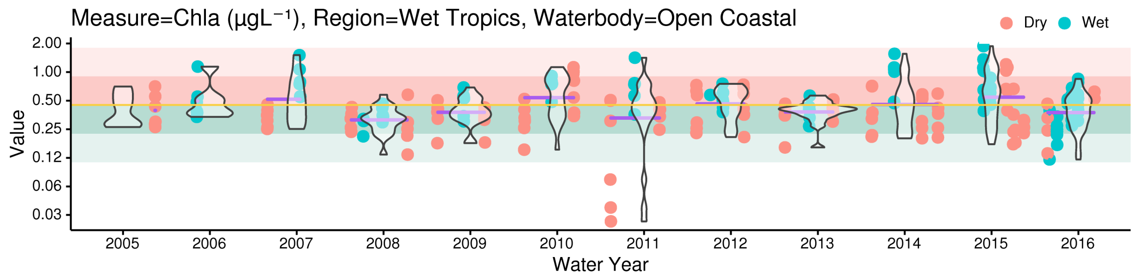
[5]this pattern also persists across all Zones (Region/Water body) and Measures - although other Measures and Zones not provided here to reduce space.

threshold) appears to in between these two extremes and thus could be considered a reasonable comprimise between over and under sensitivity.

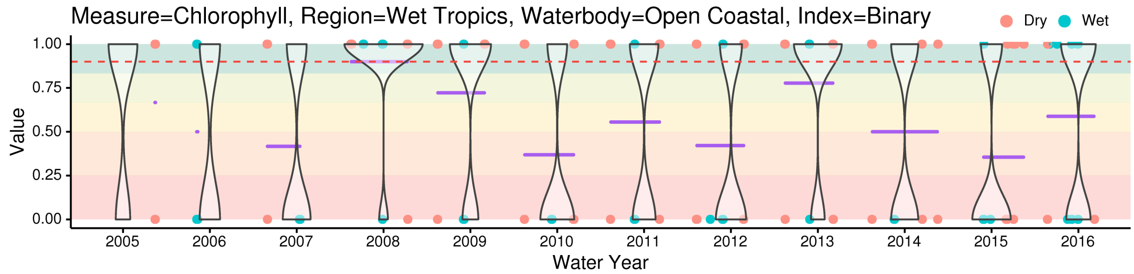
Spatial representations for Wet Tropics Open Coastal Chlorophyll-a (figs. 39 – 42) and Dry Tropics Midshelf Chlorophyll-a (figs. 43 – 47) offer similar assessments - that fsMAMP provides a reasonable compromise between the potentially under and over sensitive fsMAMP4 and Binary formulations.

Time series of annually aggregated observations and associated annually aggregated indices (figs. 48 – 52) 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.

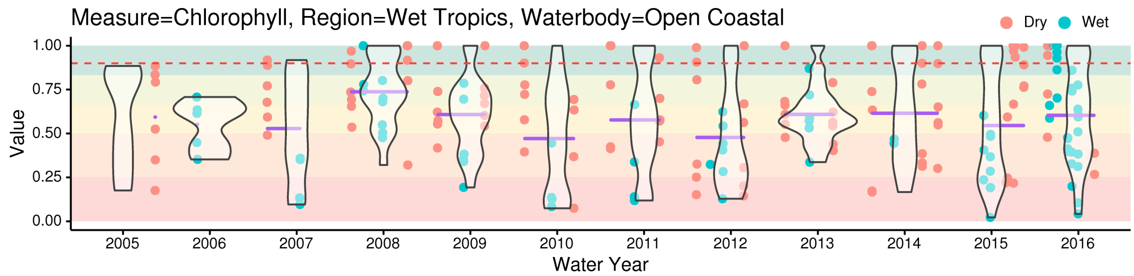
a) AIMS insitu site means



b) AIMS insitu site mean Binary



c) AIMS insitu site mean fsMAMP



d) AIMS insitu site mean fsMAMP4

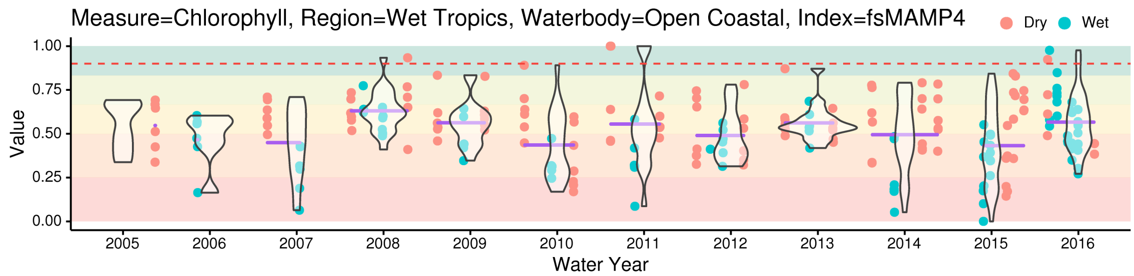
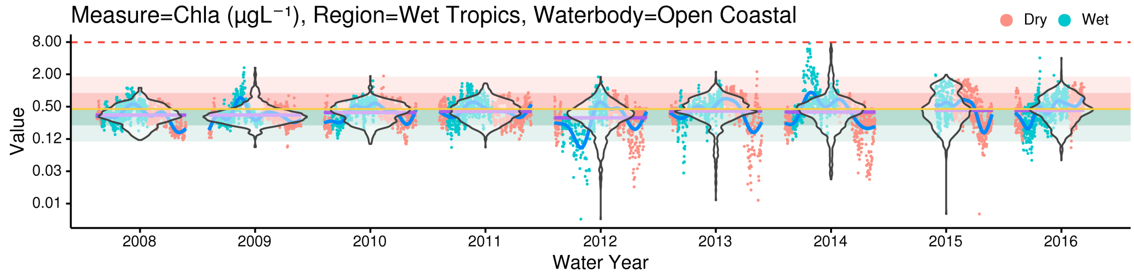
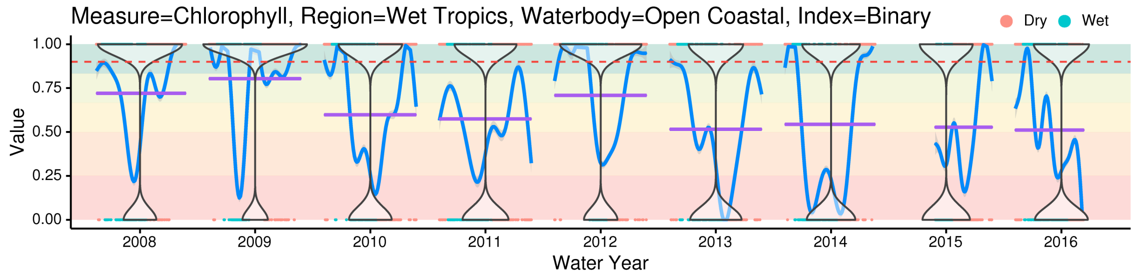


Figure 34: Temporal distribution of AIMS insitu Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) 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.

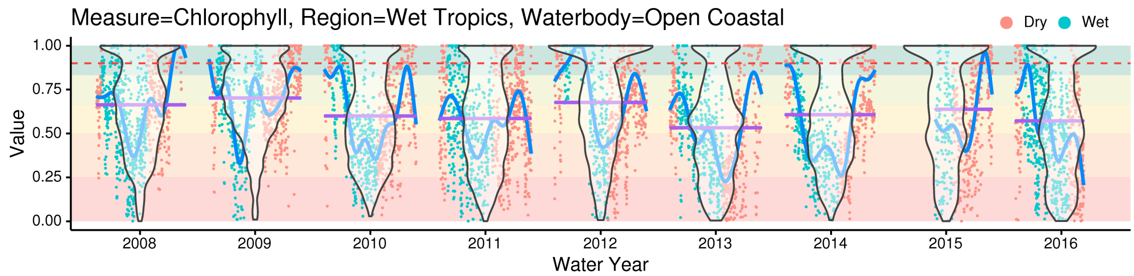
a) AIMS FLNTU raw site means



b) AIMS FLNTU site mean Binary



c) AIMS FLNTU site mean fsMAMP



d) AIMS FLNTU site mean fsMAMP4

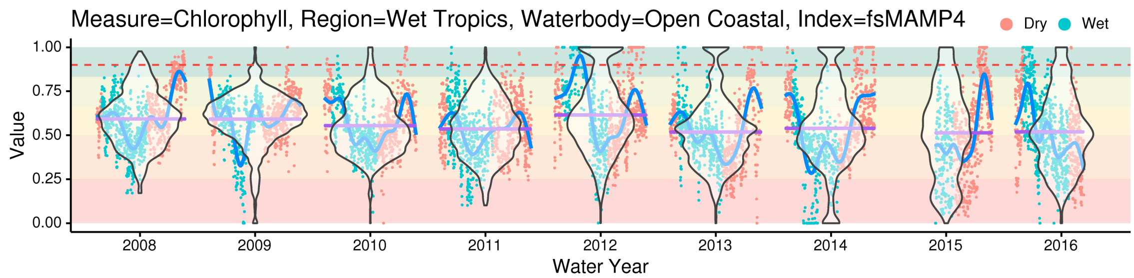
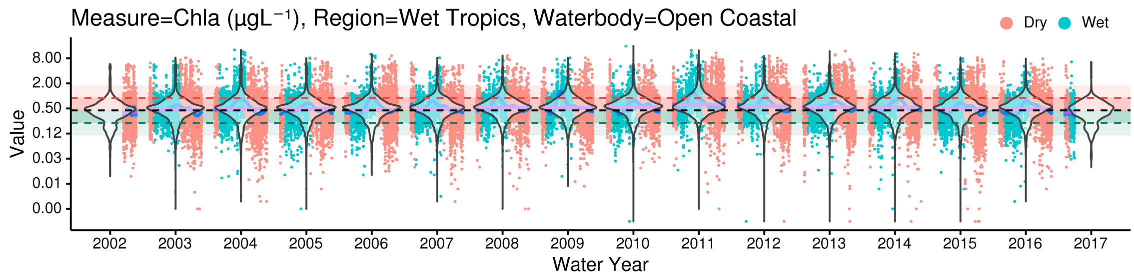
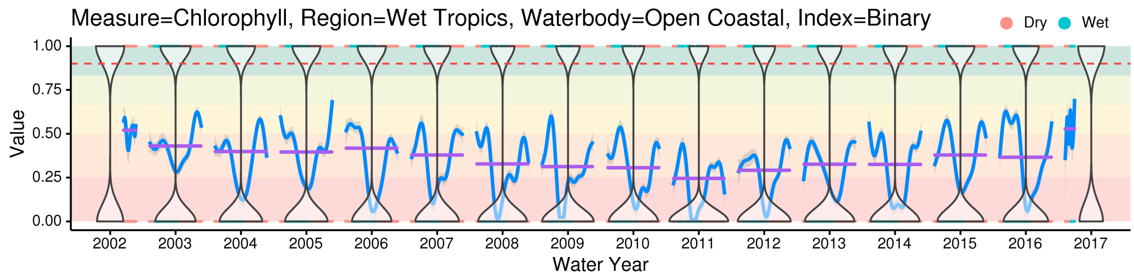


Figure 35: Temporal distribution of AIMS FLNTU Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) 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.

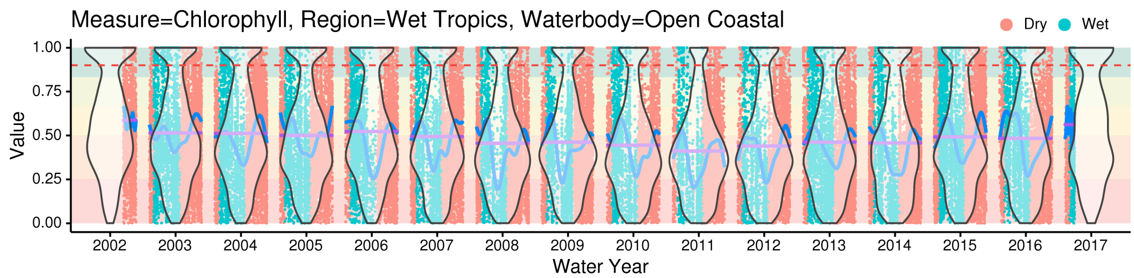
a) Satellite raw site means



b) Satellite site mean Binary



c) Satellite site mean fsMAMP



d) Satellite site mean fsMAMP4

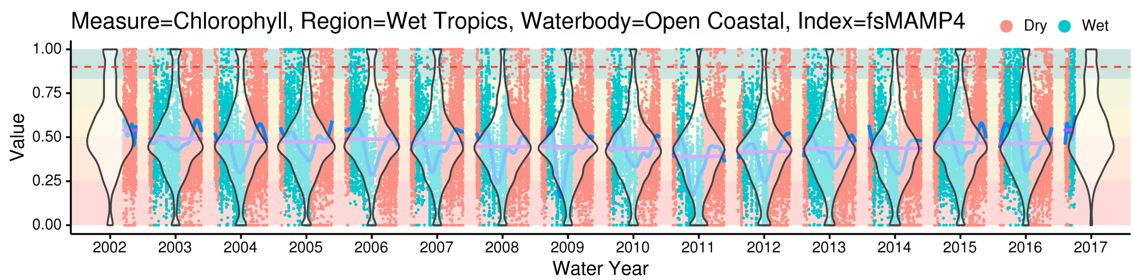
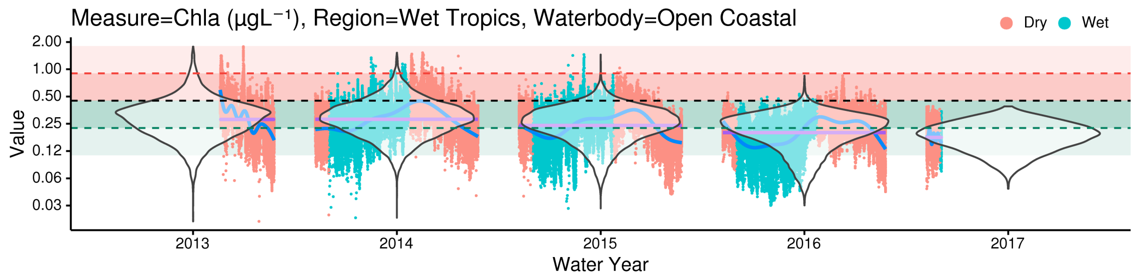
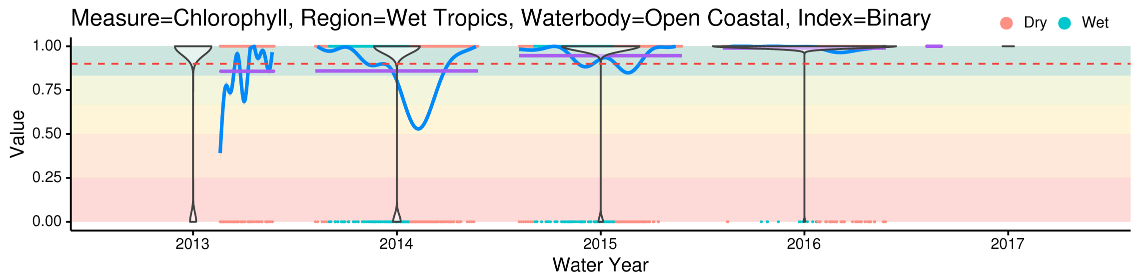


Figure 36: Temporal distribution of Satellite Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) 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.

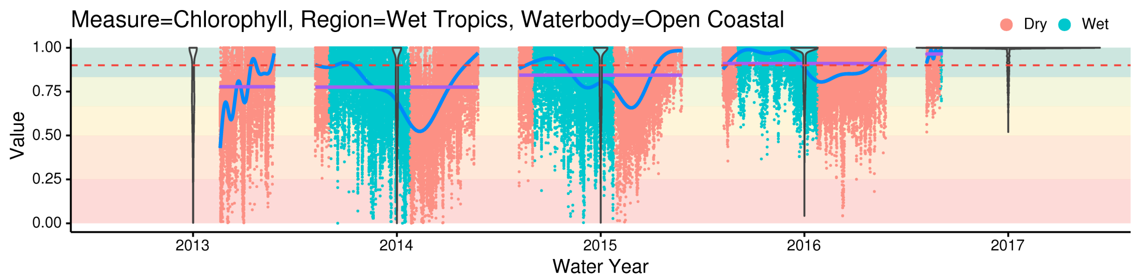
a) eReefs raw site means



b) eReefs site mean Binary



c) eReefs site mean fsMAMP



d) eReefs site mean fsMAMP4

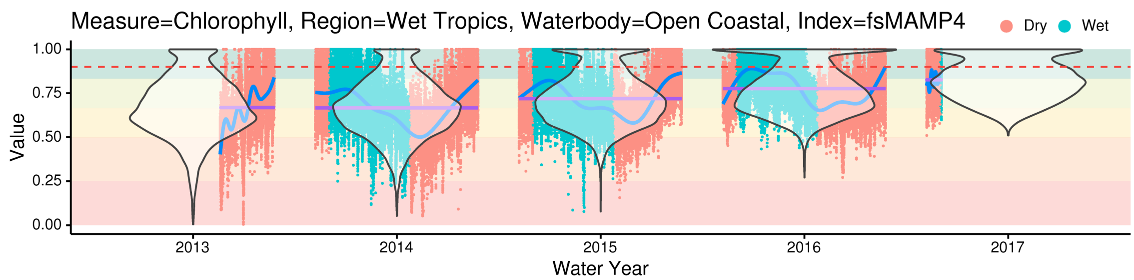
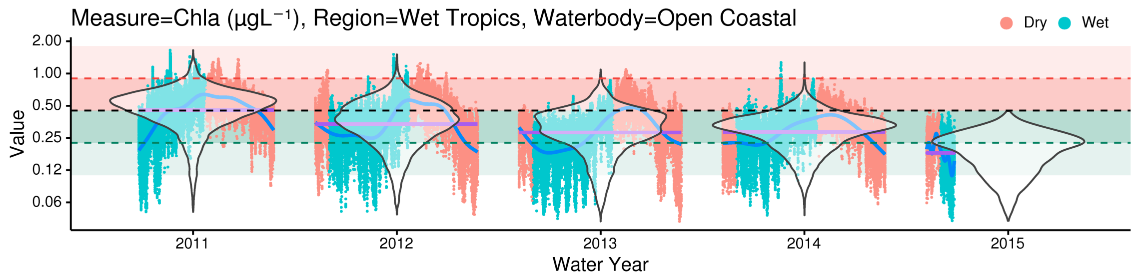
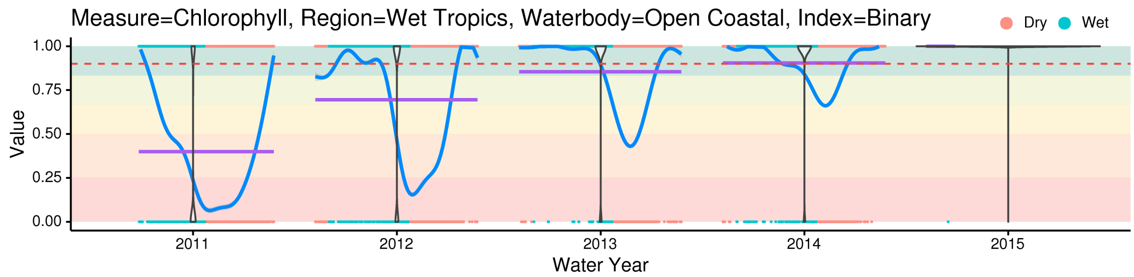


Figure 37: Temporal distribution of eReefs Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) respectively 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.

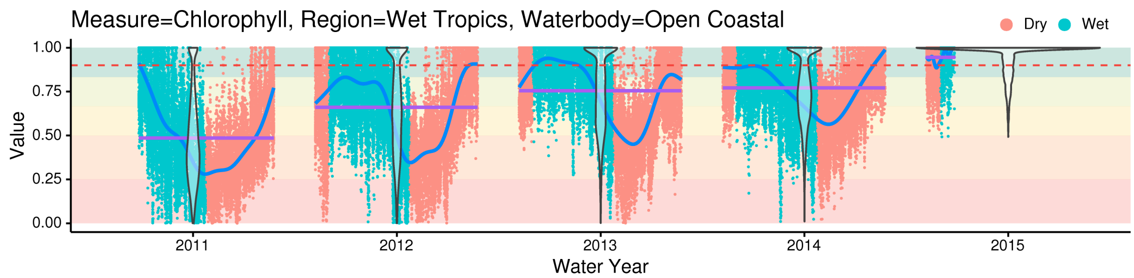
a) eReefs926 raw site means



b) eReefs926 site mean Binary



c) eReefs926 site mean fsMAMP



d) eReefs926 site mean fsMAMP4

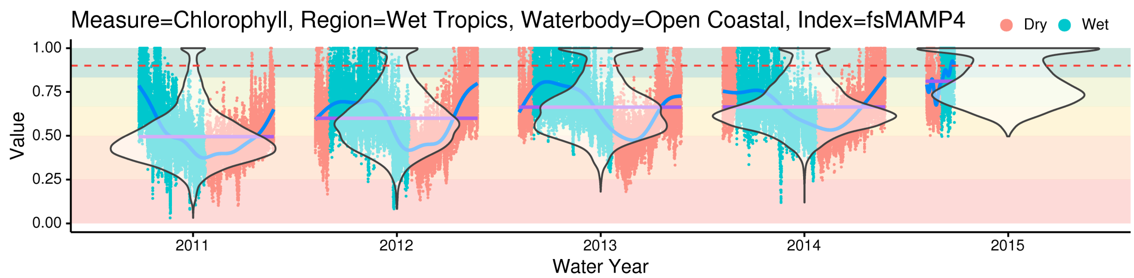


Figure 38: Temporal distribution of eReefs926 Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Red and Blue symbols represent samples collected in Dry and Wet seasons respectively. Green and red shaded banding on a) 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.

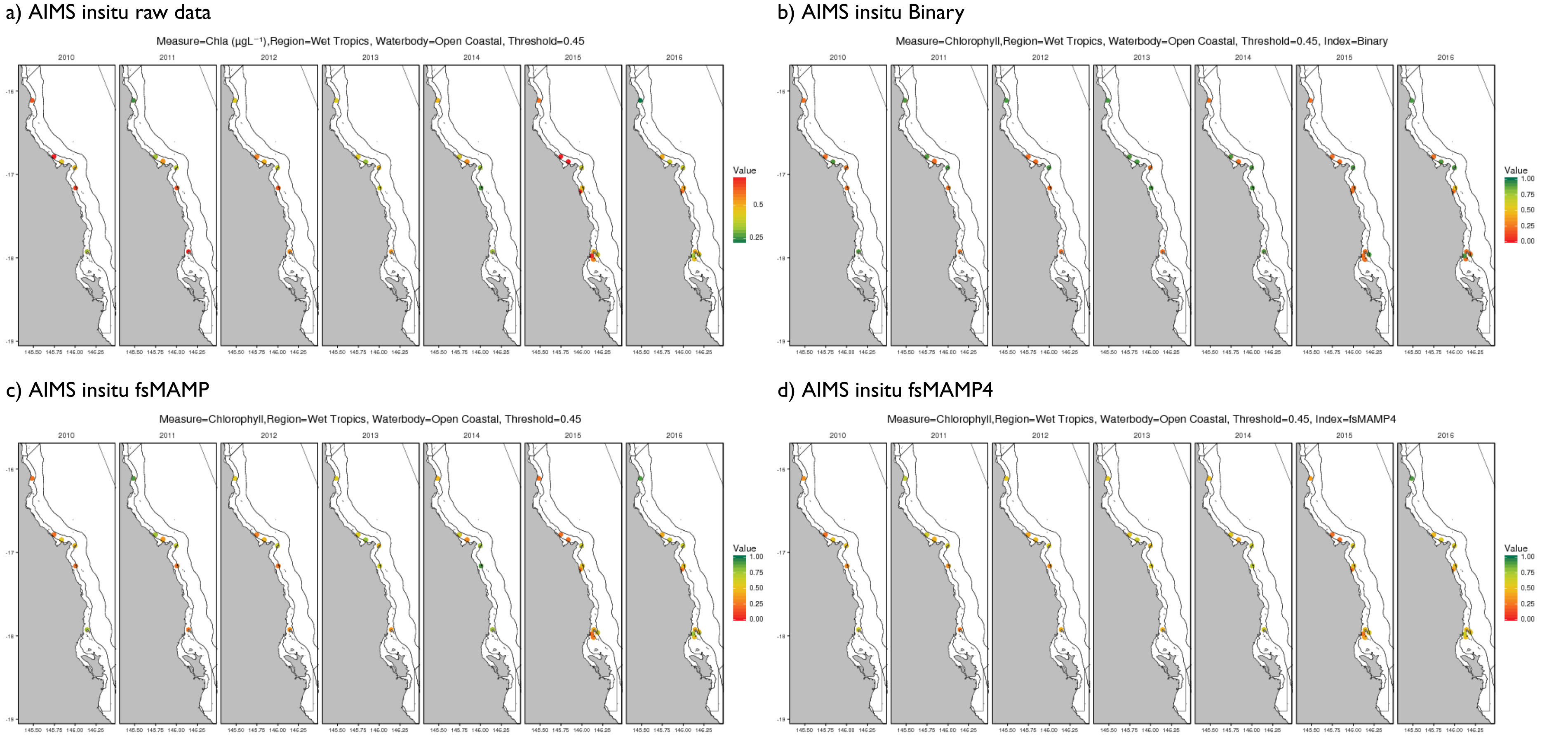


Figure 39: Spatial distribution of AIMS in situ Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

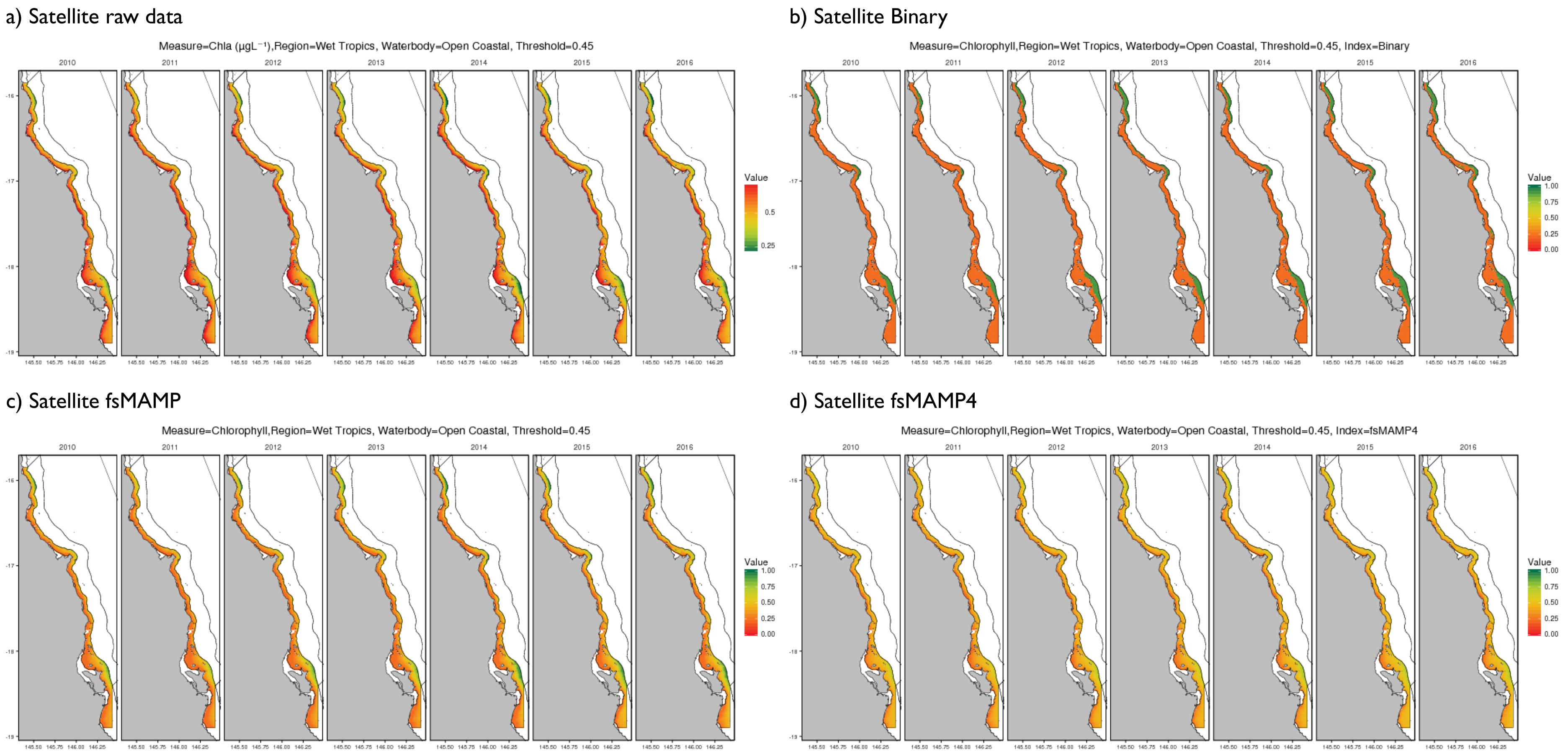


Figure 40: Spatial distribution of Satellite Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

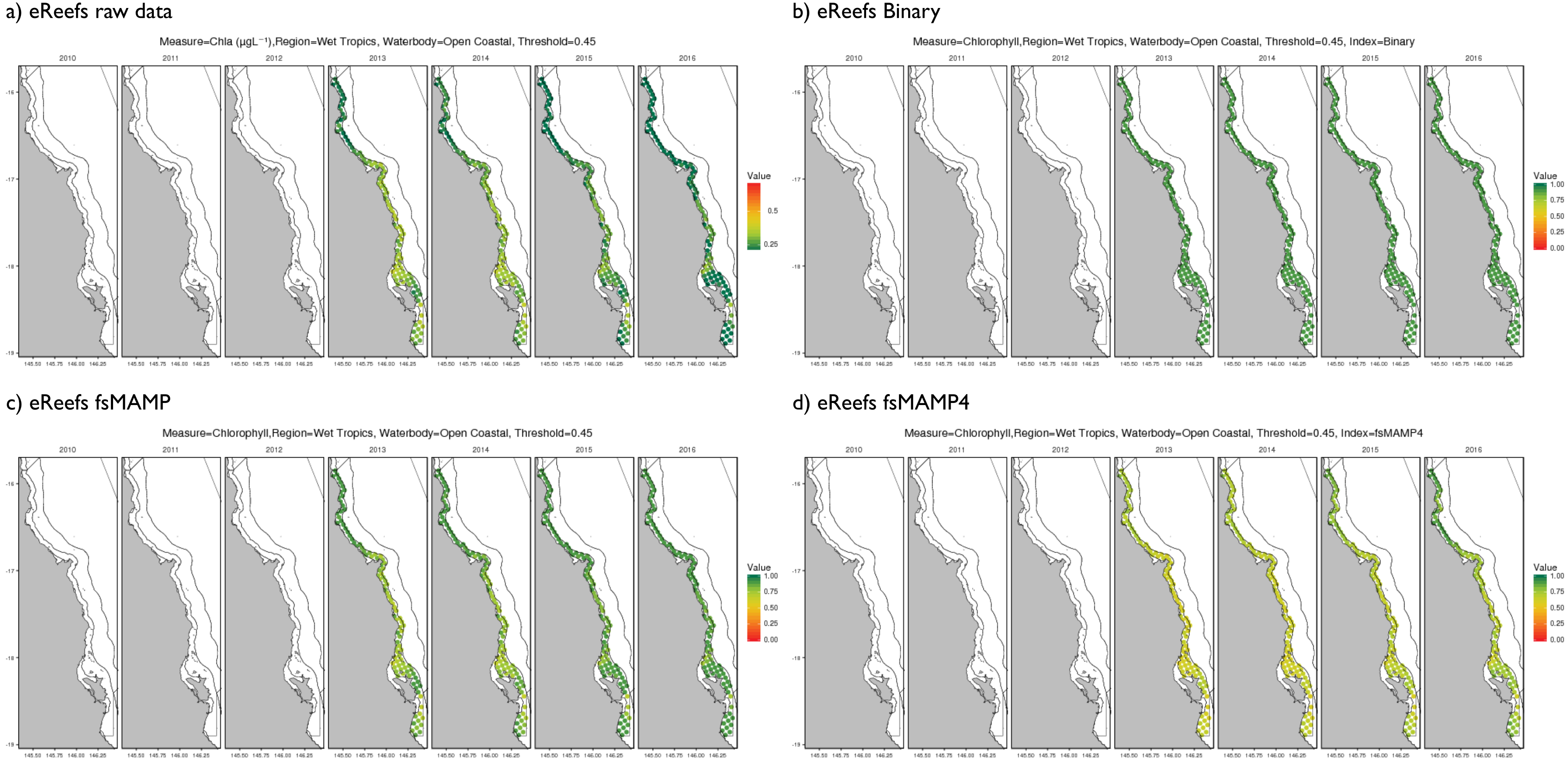


Figure 41: Spatial distribution of eReefs Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

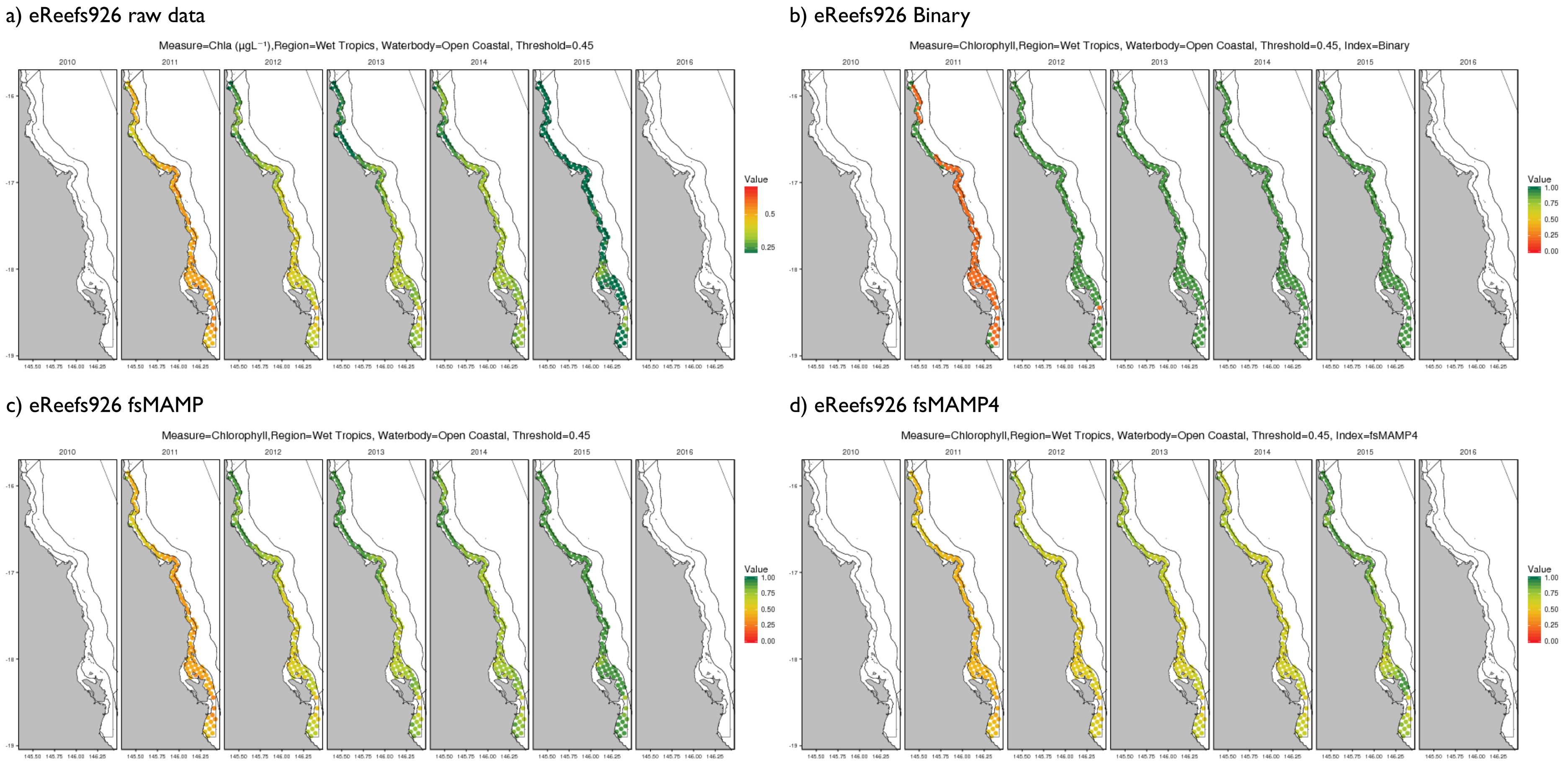
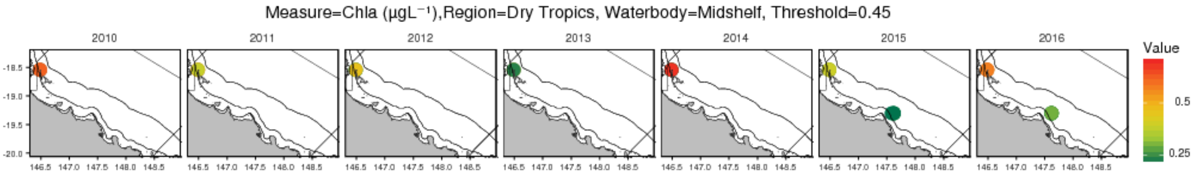
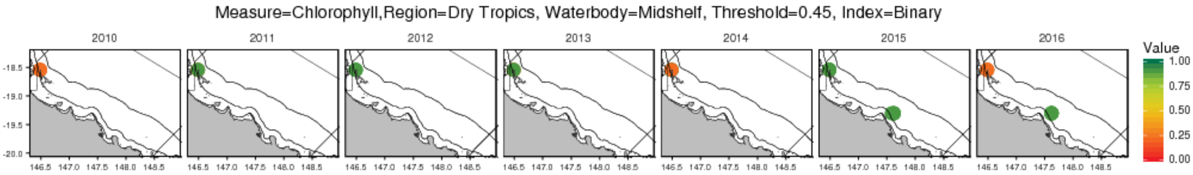


Figure 42: Spatial distribution of eReefs926 Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Wet Tropics Open Coastal zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

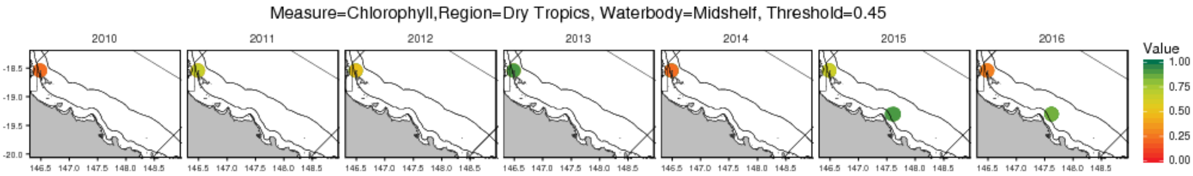
a) AIMS insitu raw data



b) AIMS insitu Binary



c) AIMS insitu fsMAMP



d) AIMS insitu fsMAMP4

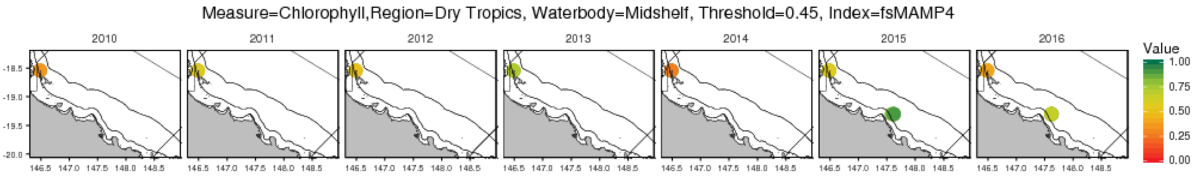
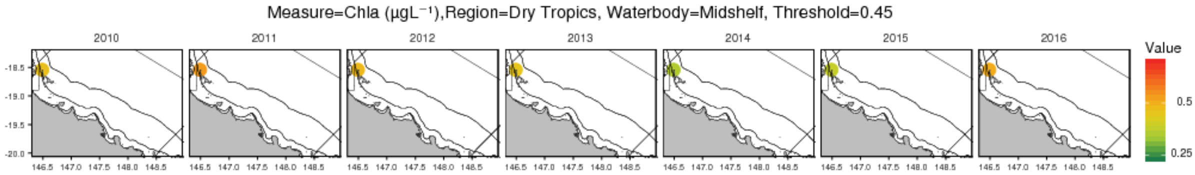
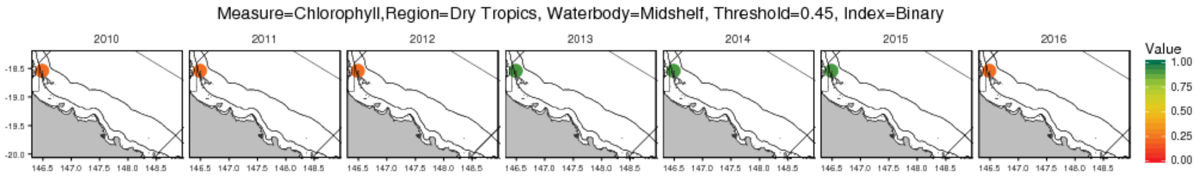


Figure 43: Spatial distribution of AIMS insitu Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP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

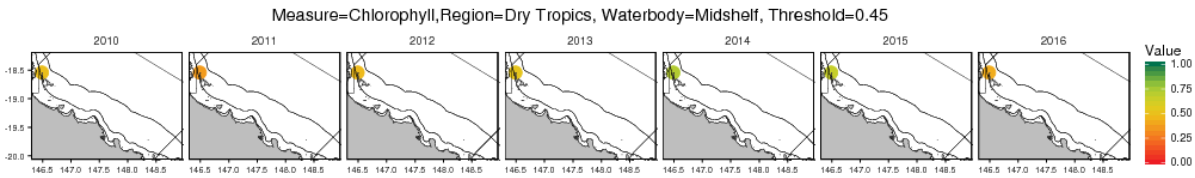
a) AIMS FLNTU raw data



b) AIMS FLNTU Binary



c) AIMS FLNTU fsMAMP



d) AIMS FLNTU fsMAMP4

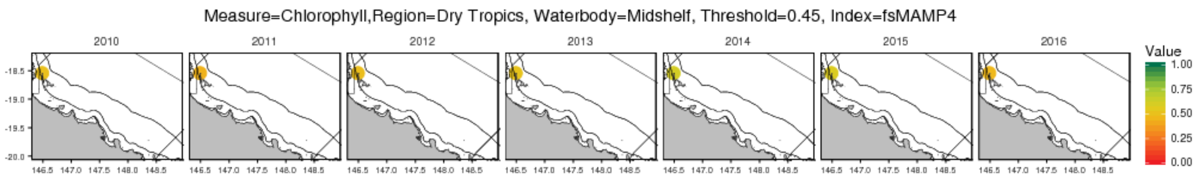
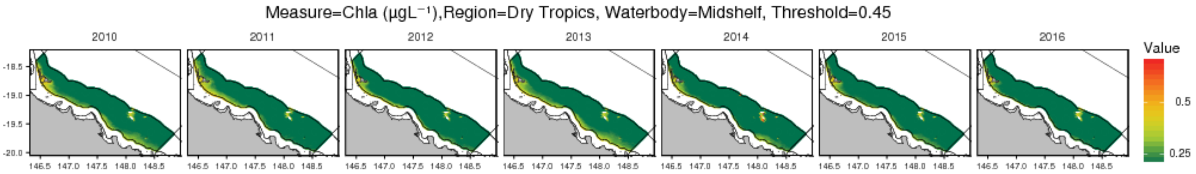
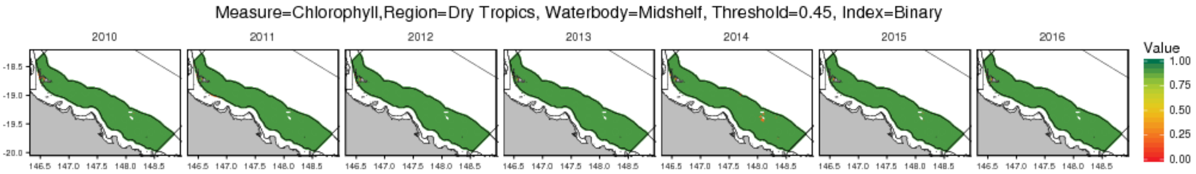


Figure 44: Spatial distribution of AIMS FLNTU Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and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

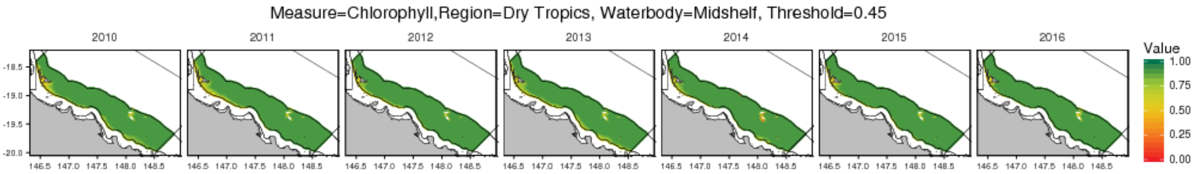
a) Satellite raw data



b) Satellite Binary



c) Satellite fsMAMP



d) Satellite fsMAMP4

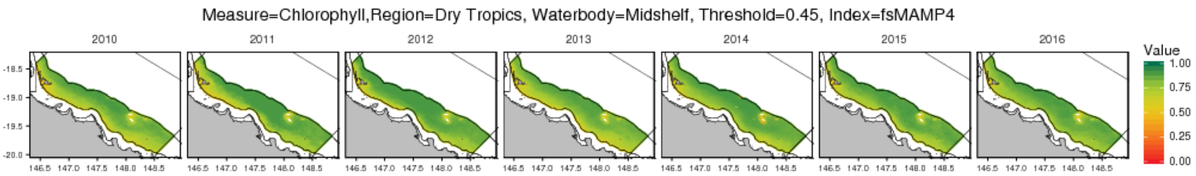
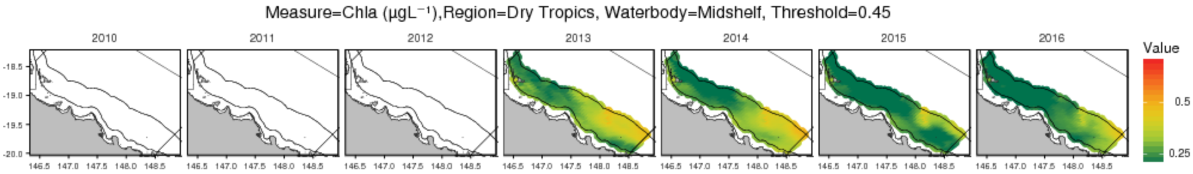
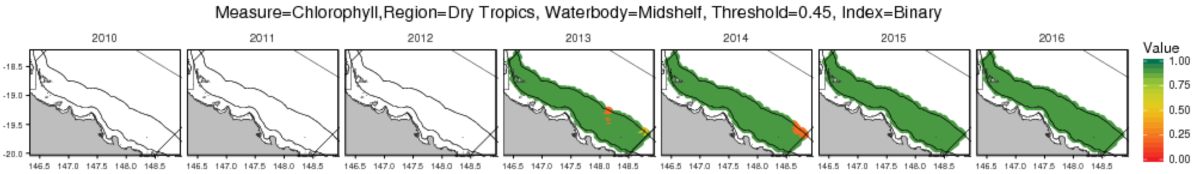


Figure 45: Spatial distribution of Satellite Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and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

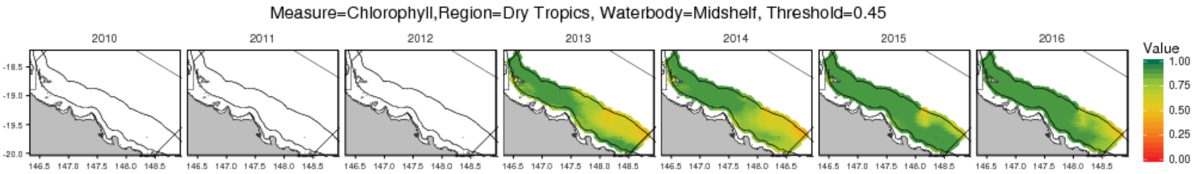
a) eReefs raw data



b) eReefs Binary



c) eReefs fsMAMP



d) eReefs fsMAMP4

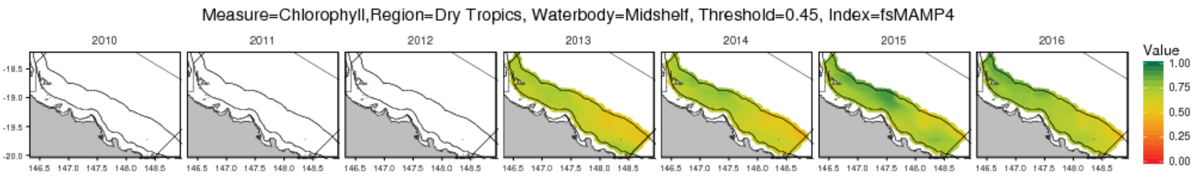
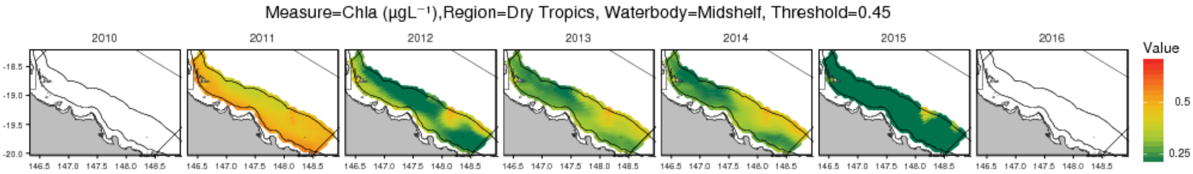
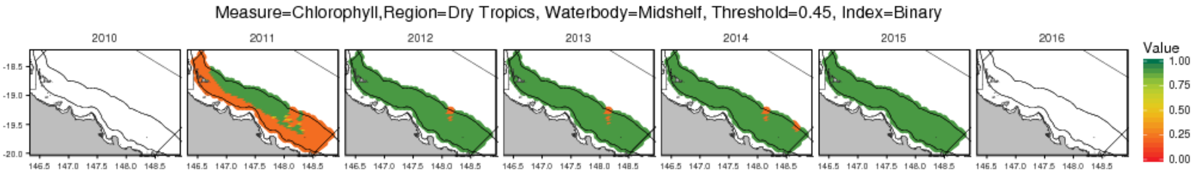


Figure 46: Spatial distribution of eReefs Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

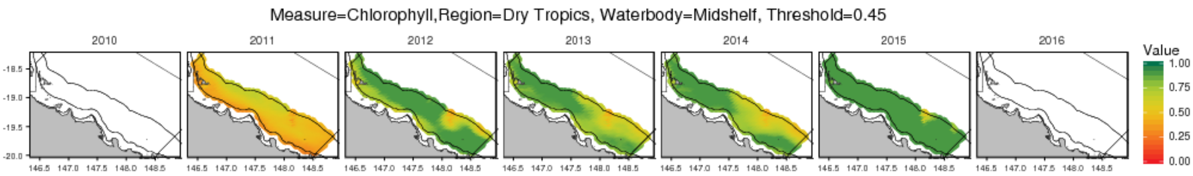
a) eReefs926 raw data



b) eReefs926 Binary



c) eReefs926 fsMAMP



d) eReefs926 fsMAMP4

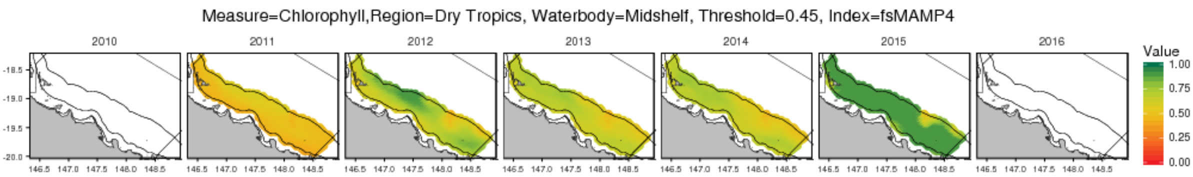


Figure 47: Spatial distribution of eReefs926 Chlorophyll-a a) samples and associated b) Binary, c) fsMAMP and d) fsMAMP4 index formulations for the Dry Tropics Midshelf zone. Color bars scaled to half (green) and twice (red) threshold value for raw data and 1 (green) and 0 (red) for Binary, fsMAMP and fsMAMP4.

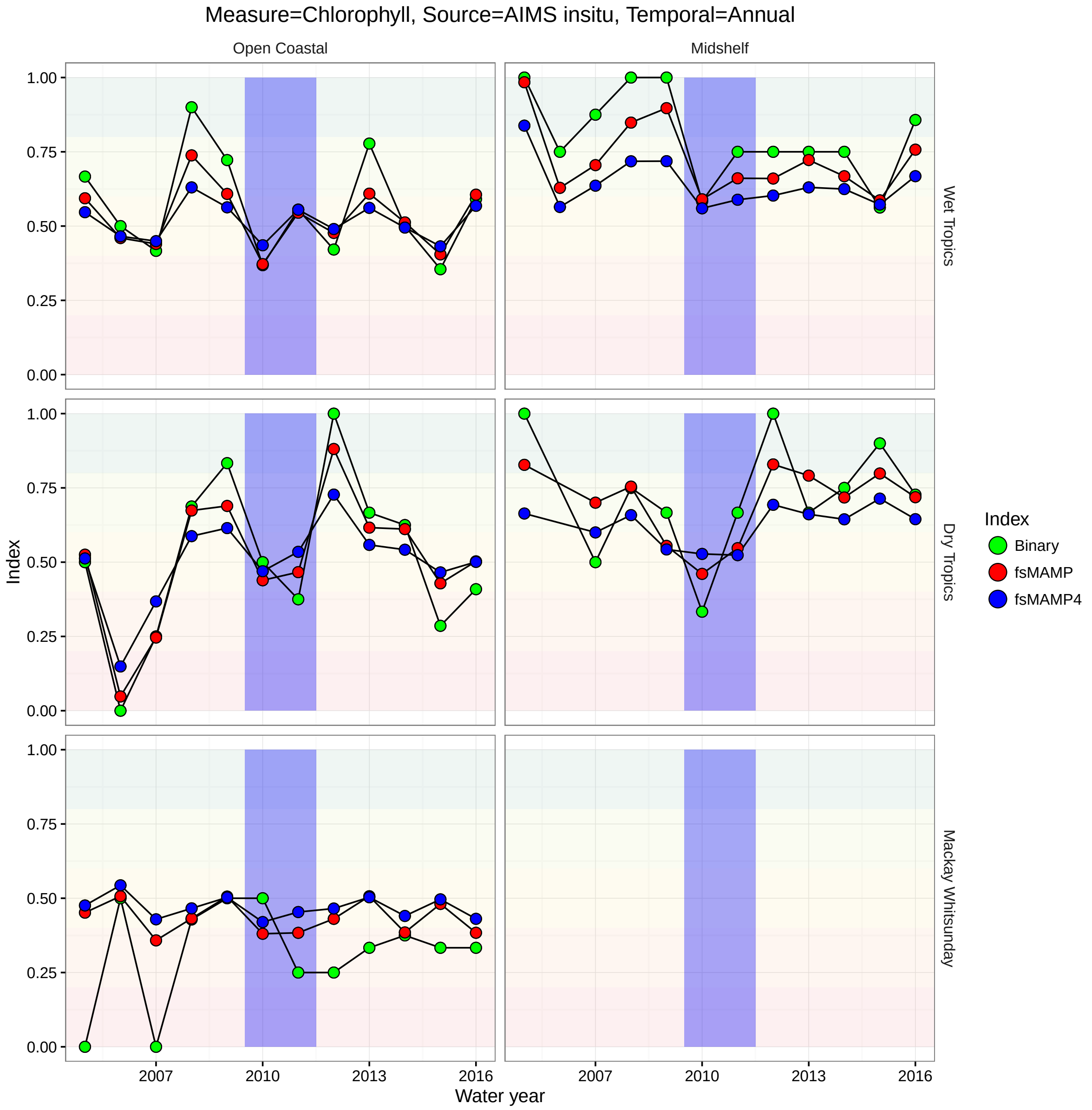


Figure 48: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for AIMS insitu Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

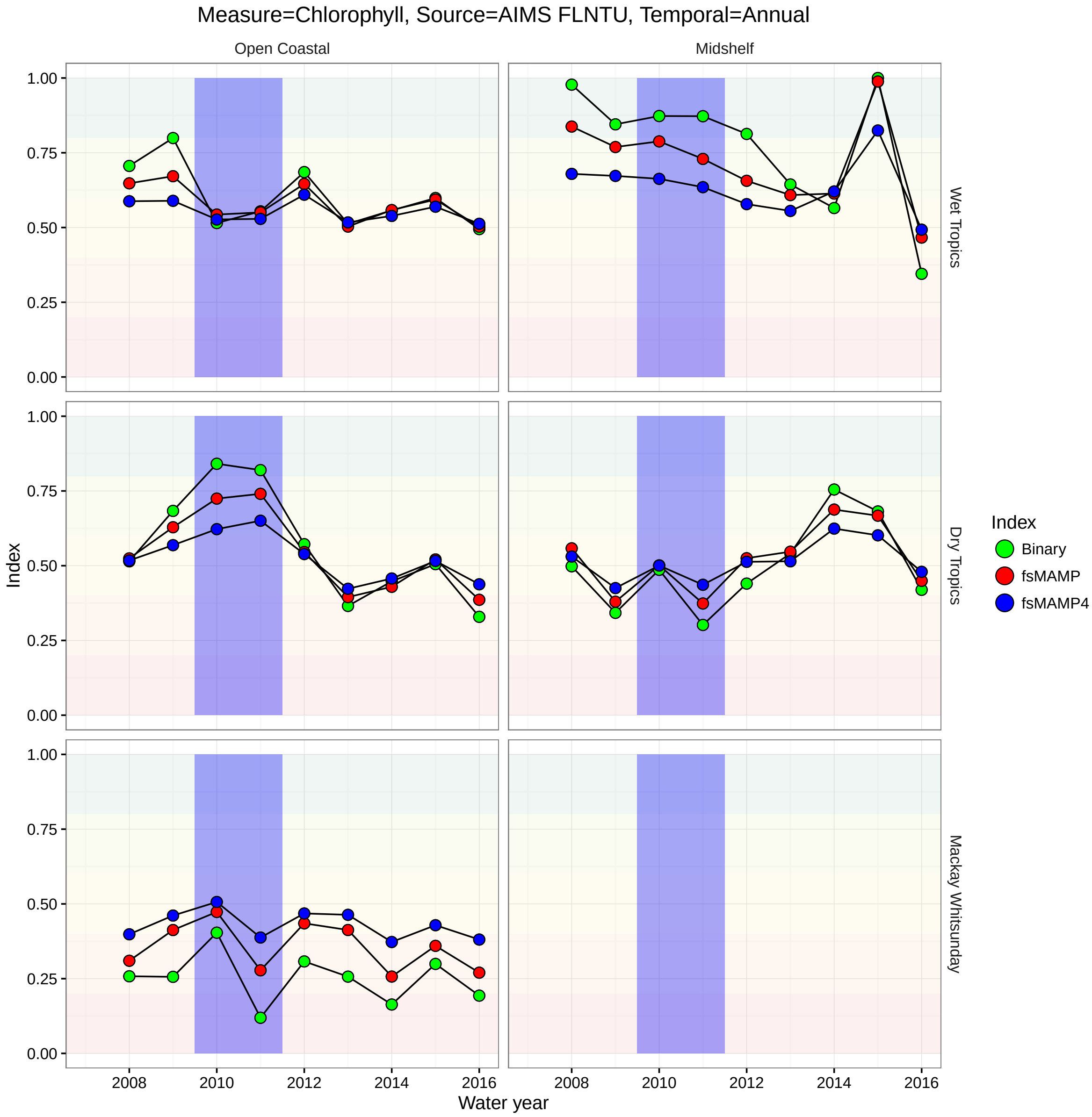


Figure 49: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for AIMS FLNTU Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

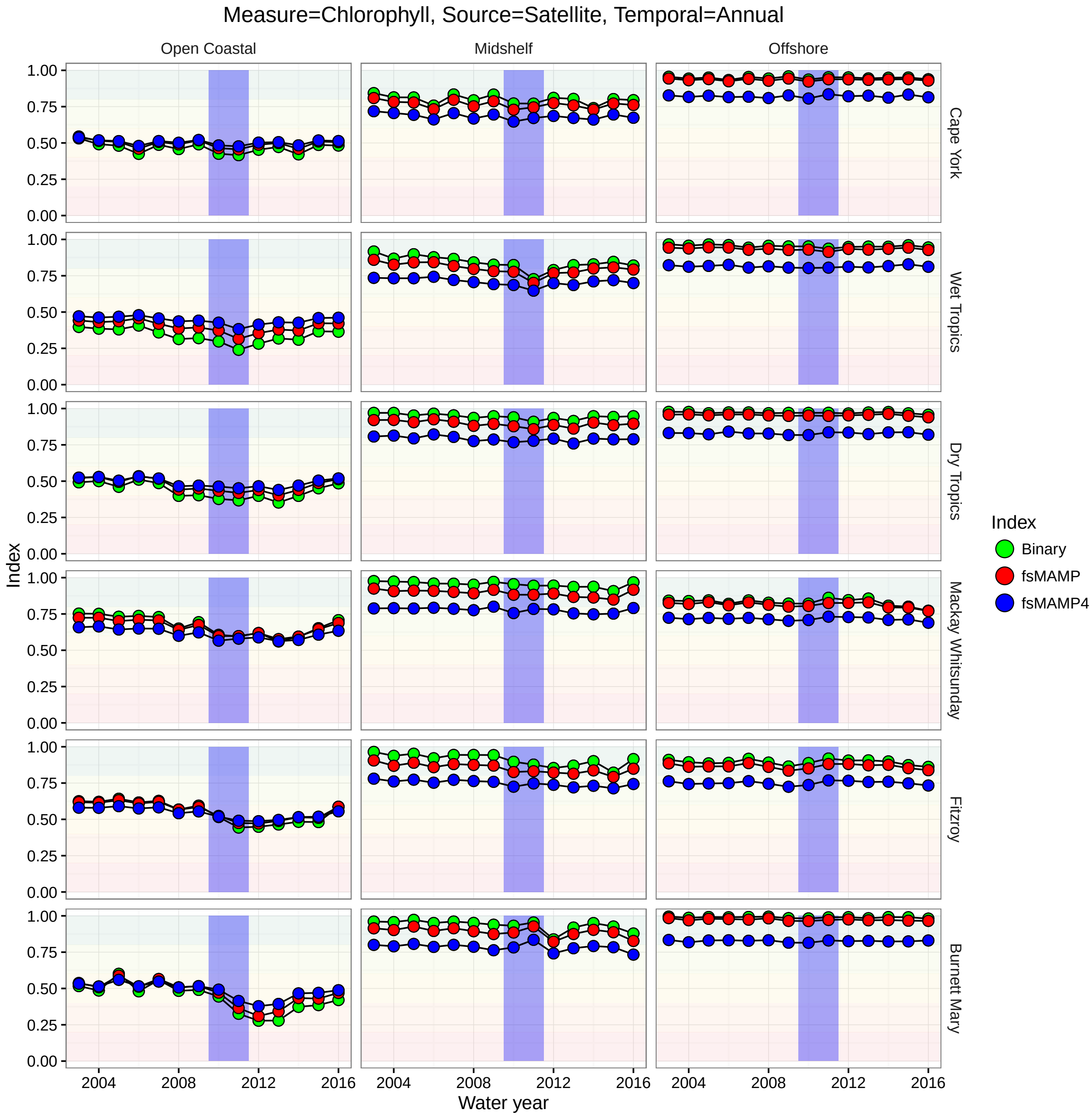


Figure 50: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for Satellite Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

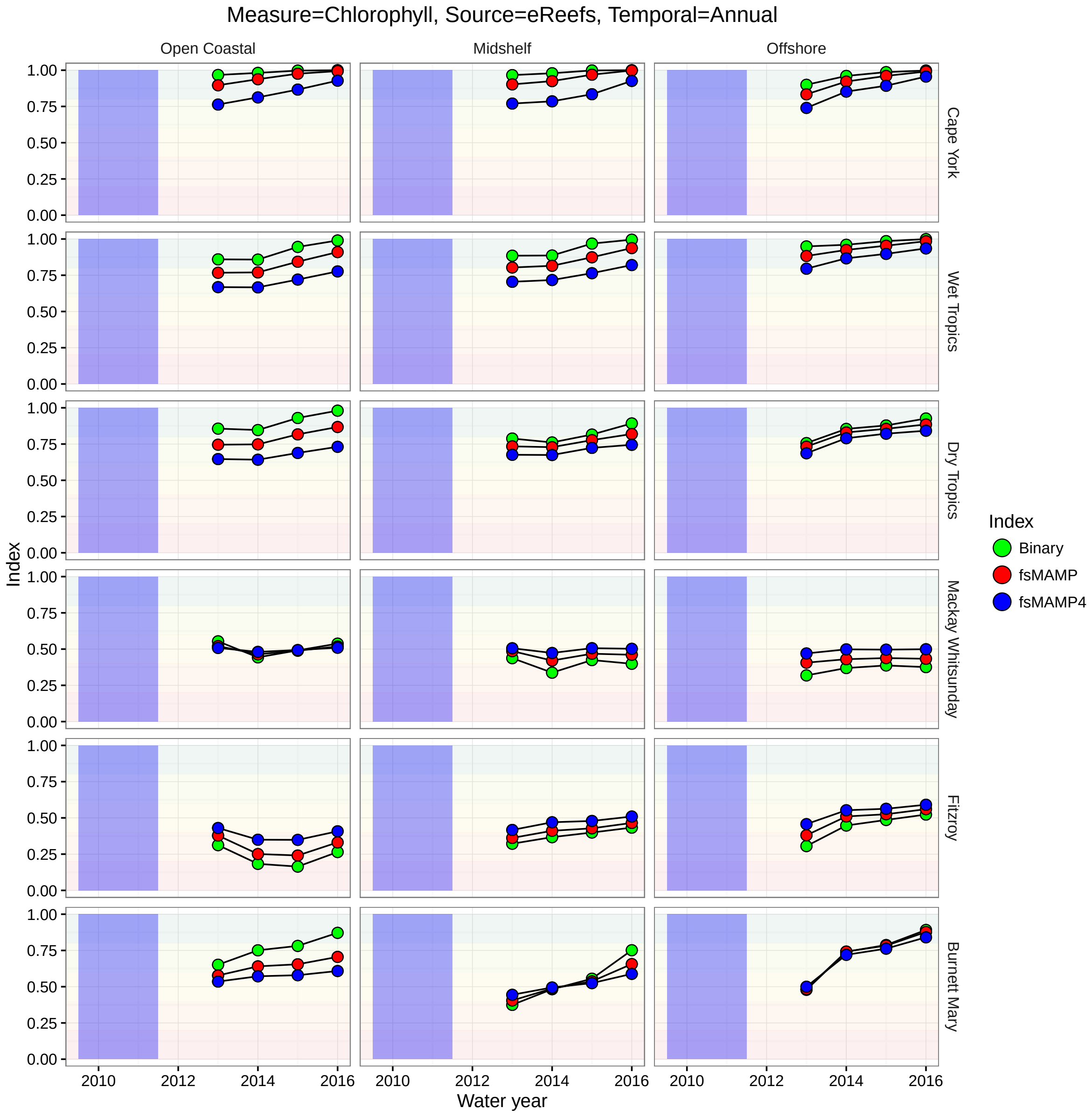


Figure 51: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for eReefs Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

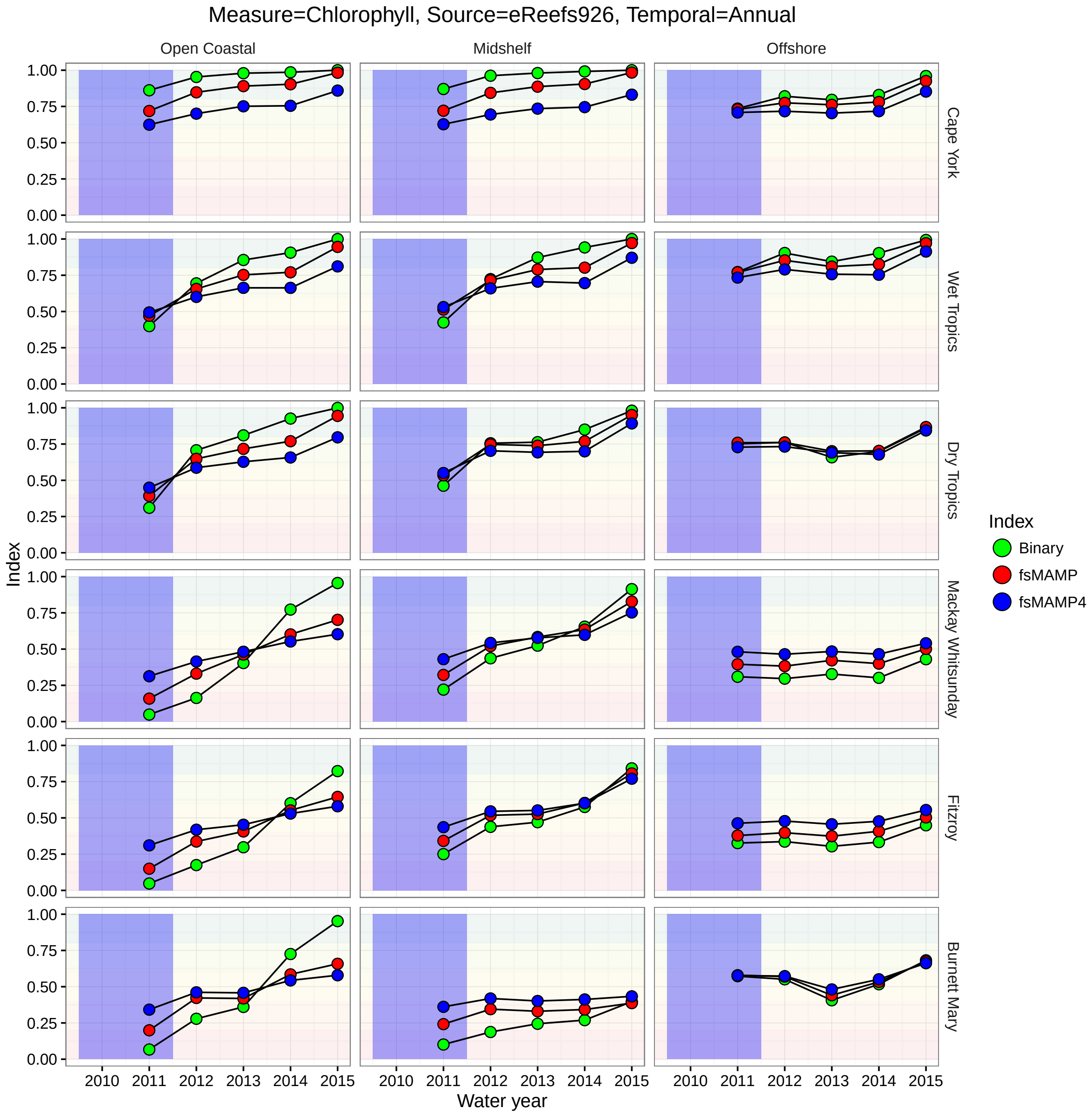


Figure 52: Time series of annually aggregated Binary, fsMAMP and fsMAMP4 index formulations for eReefs926 Chlorophyll-a across each of the Regions and Water bodies. The blue vertical bar spans from mid 2009 to mid 2011.

5.4.2 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 the AIMS insitu data 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 53 – 56 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[6], 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

[6]less obvious for TSS and NOx

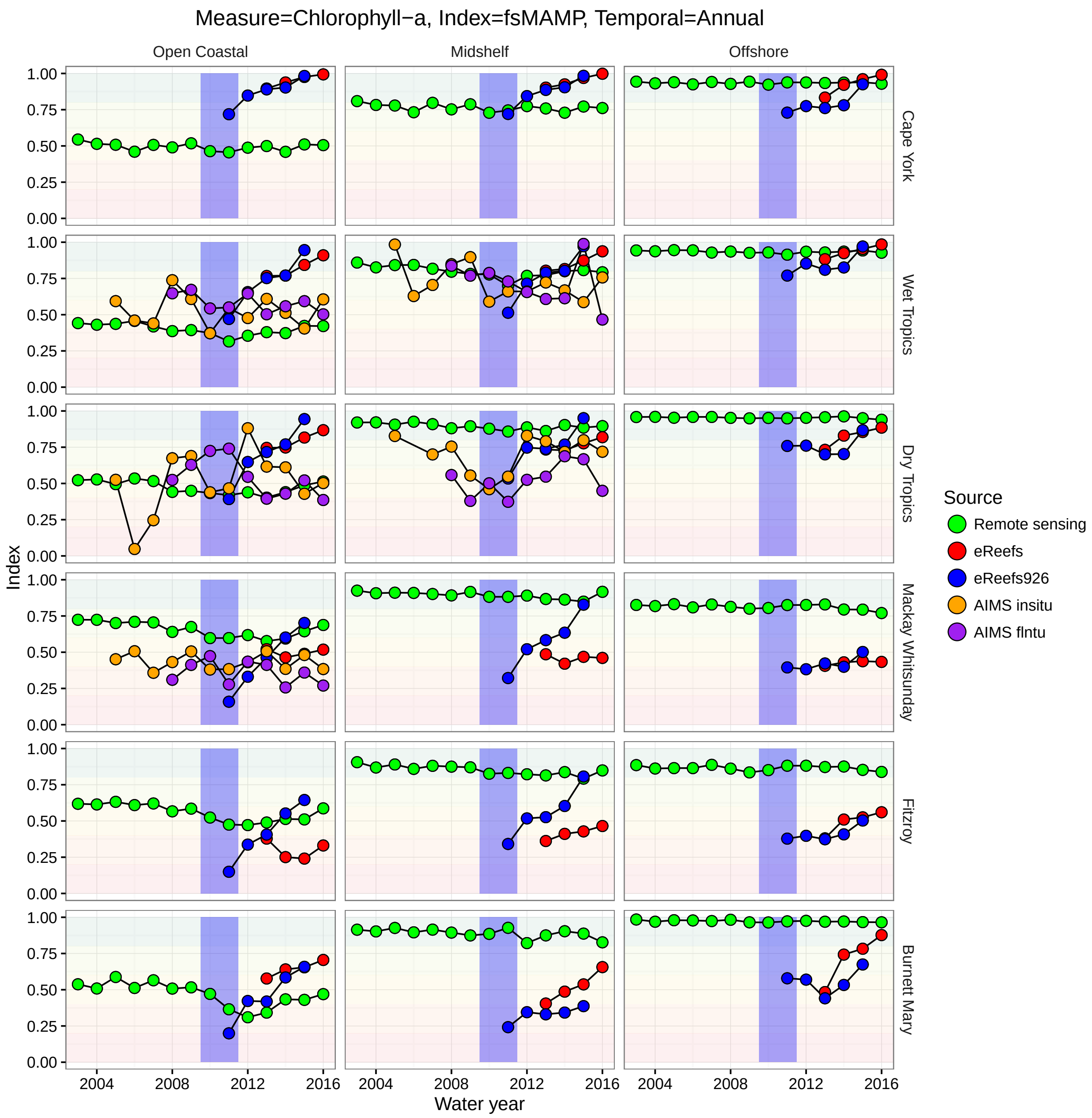


Figure 53: Time series of fsMAMP Chlorophyll-a index scores by zone for each data source. The blue vertical bar spans from mid 2009 to mid 2011.

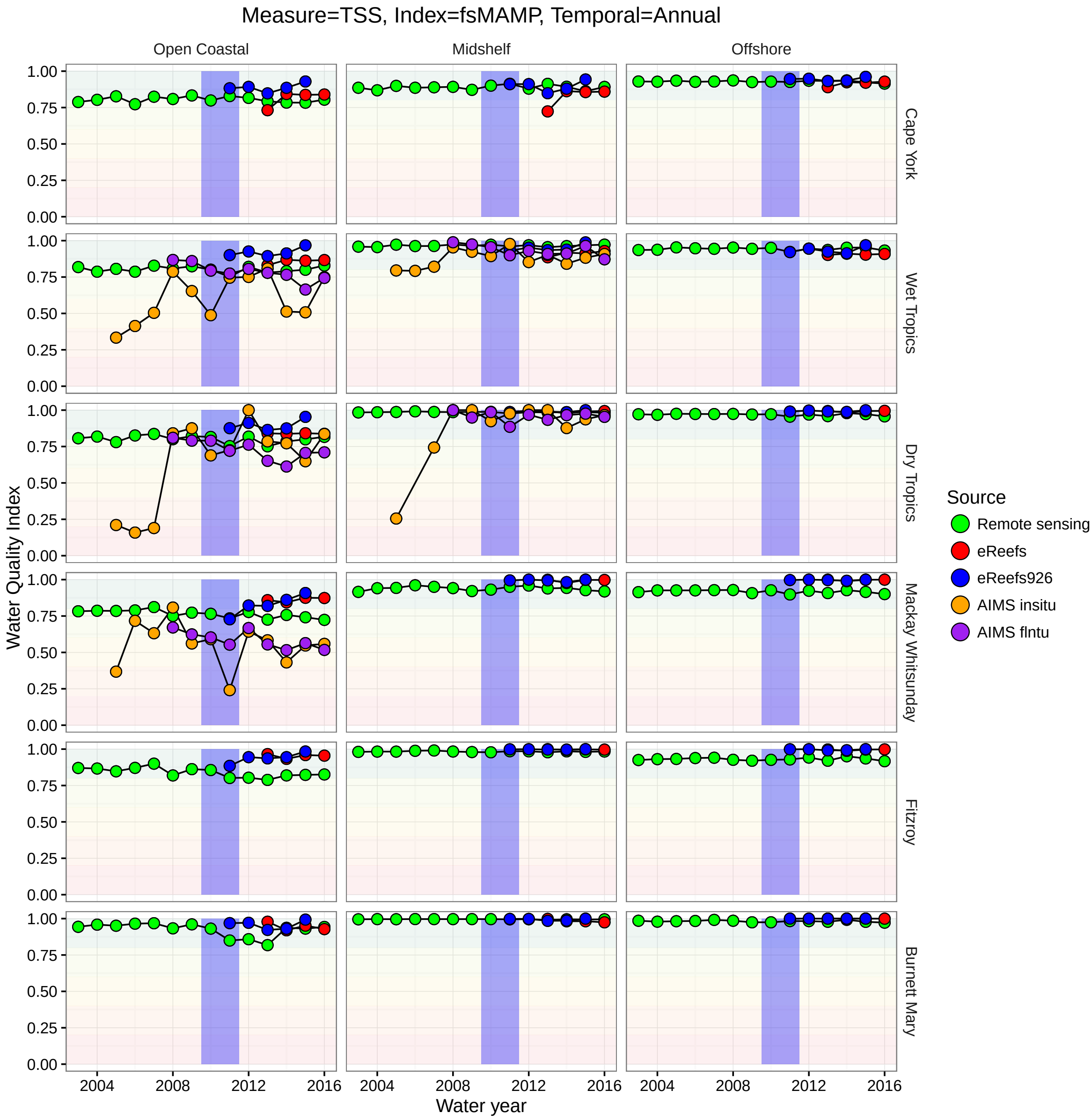


Figure 54: Time series of fsMAMP TSS index scores by zone for each data source. The blue vertical bar spans from mid 2009 to mid 2011.

Figure 55: 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.

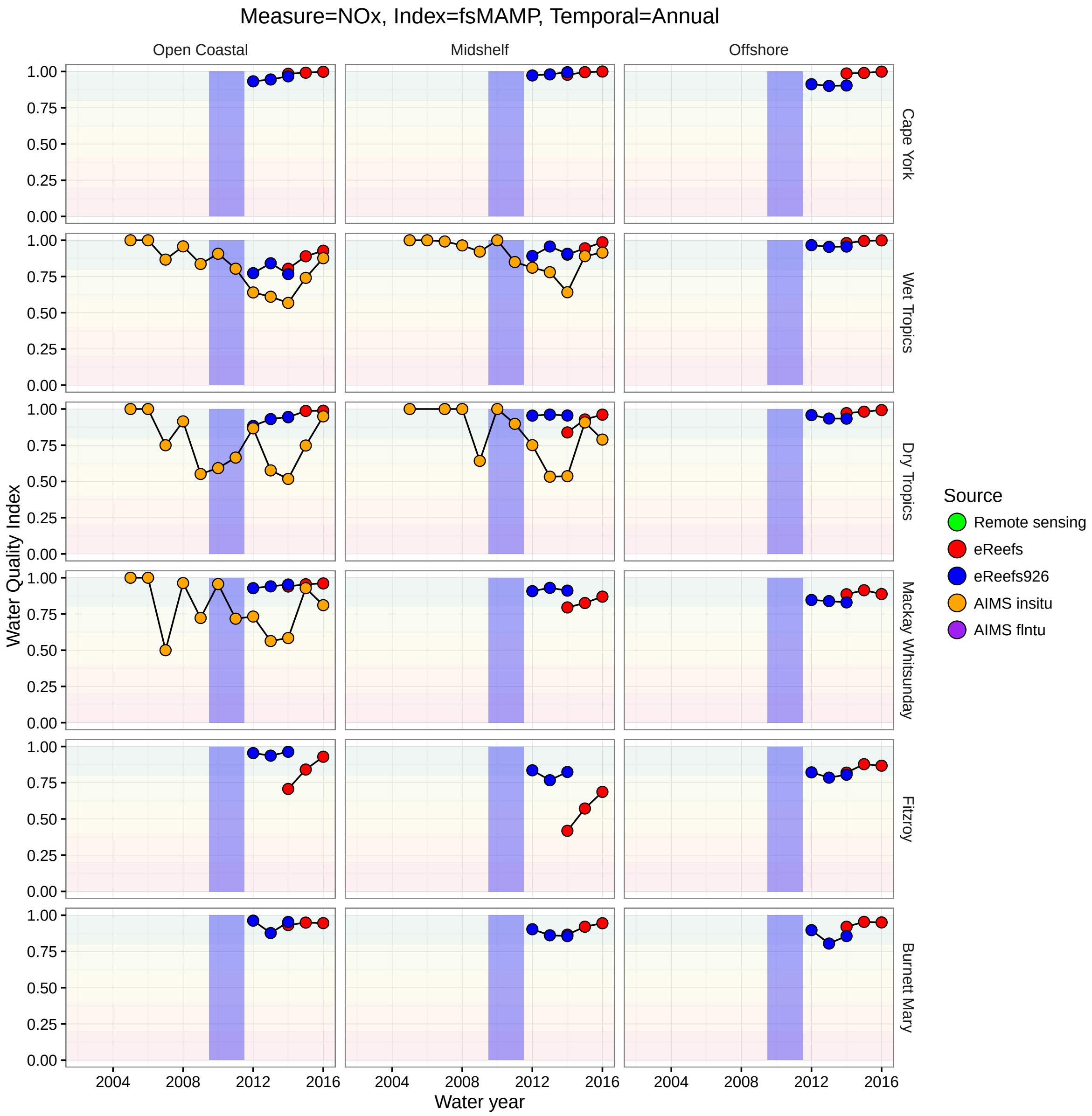


Figure 56: Time series of fsMAMP NOx index scores by zone for each data source. The blue vertical bar spans from mid 2009 to mid 2011.

5.4.3 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 57.

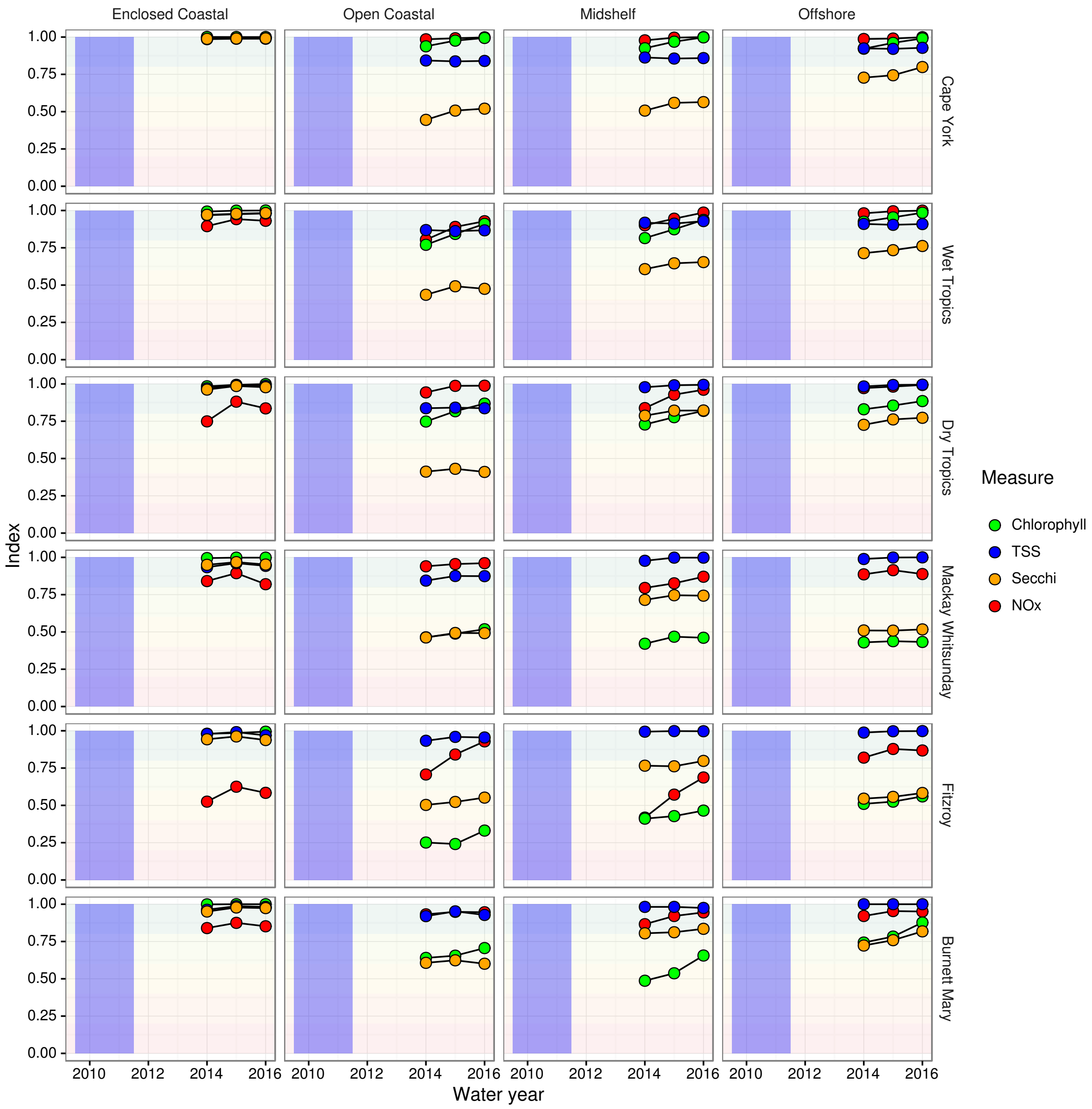


Figure 57: Time series of fsMAMP NOx index scores by zone for each data source. The blue vertical bar spans from mid 2009 to mid 2011.

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[7] 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 will be presented (see Figure 58).

Figure 58: Time series of eReefs fsMAMP Measure Index Scores by zone. The blue vertical bar spans from mid 2009 to mid 2011.

Water Quality Indicator Scores based on candidate combinations that include either Chl, TSS, SD and NOx or Chl and TSS are considered very similar. 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.

[7] These effectively act as weights

5.4.4 Measure/Site

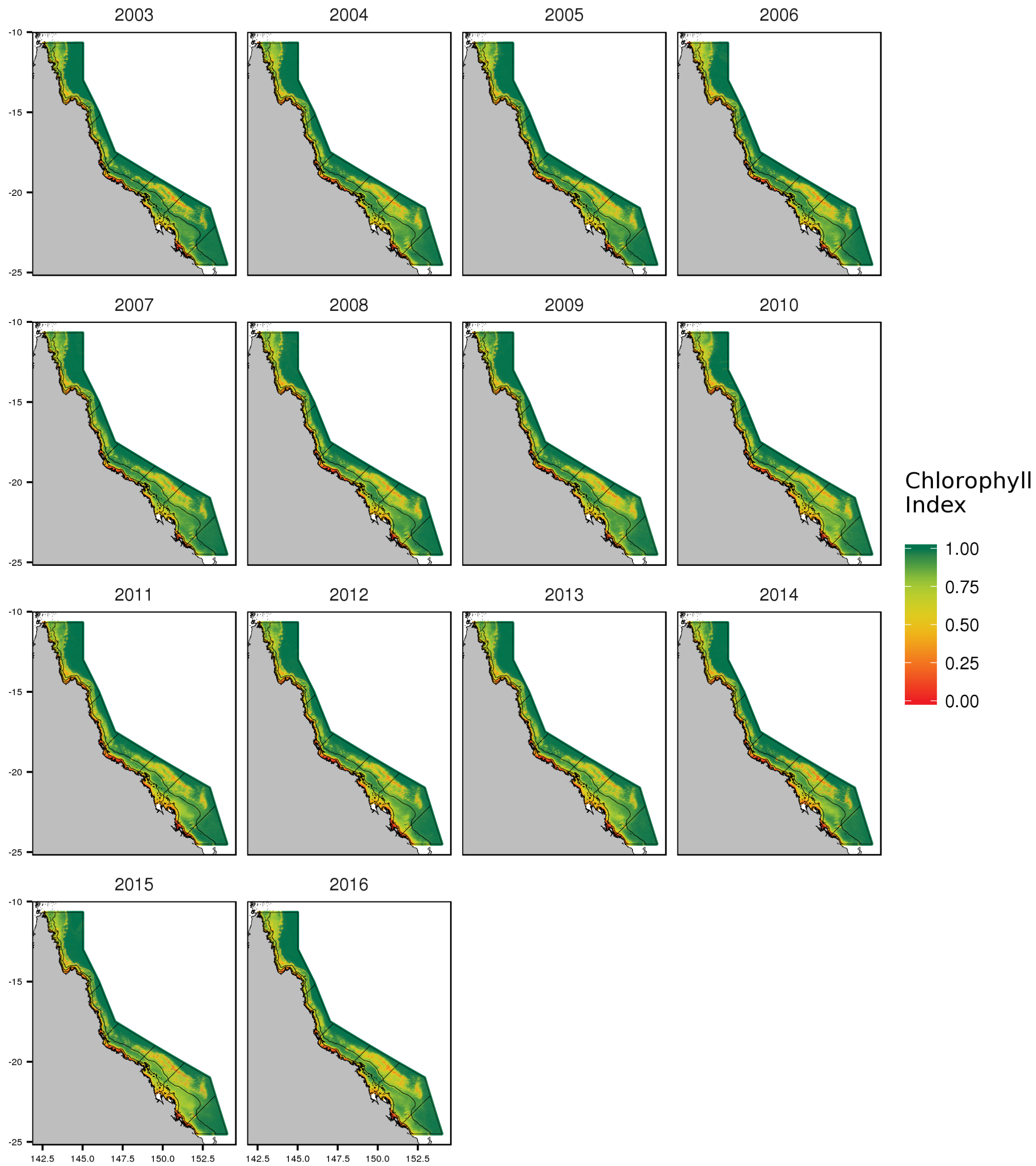


Figure 59: Spatio-temporal Satellite fsMAMP Chlorophyll-a index scores.

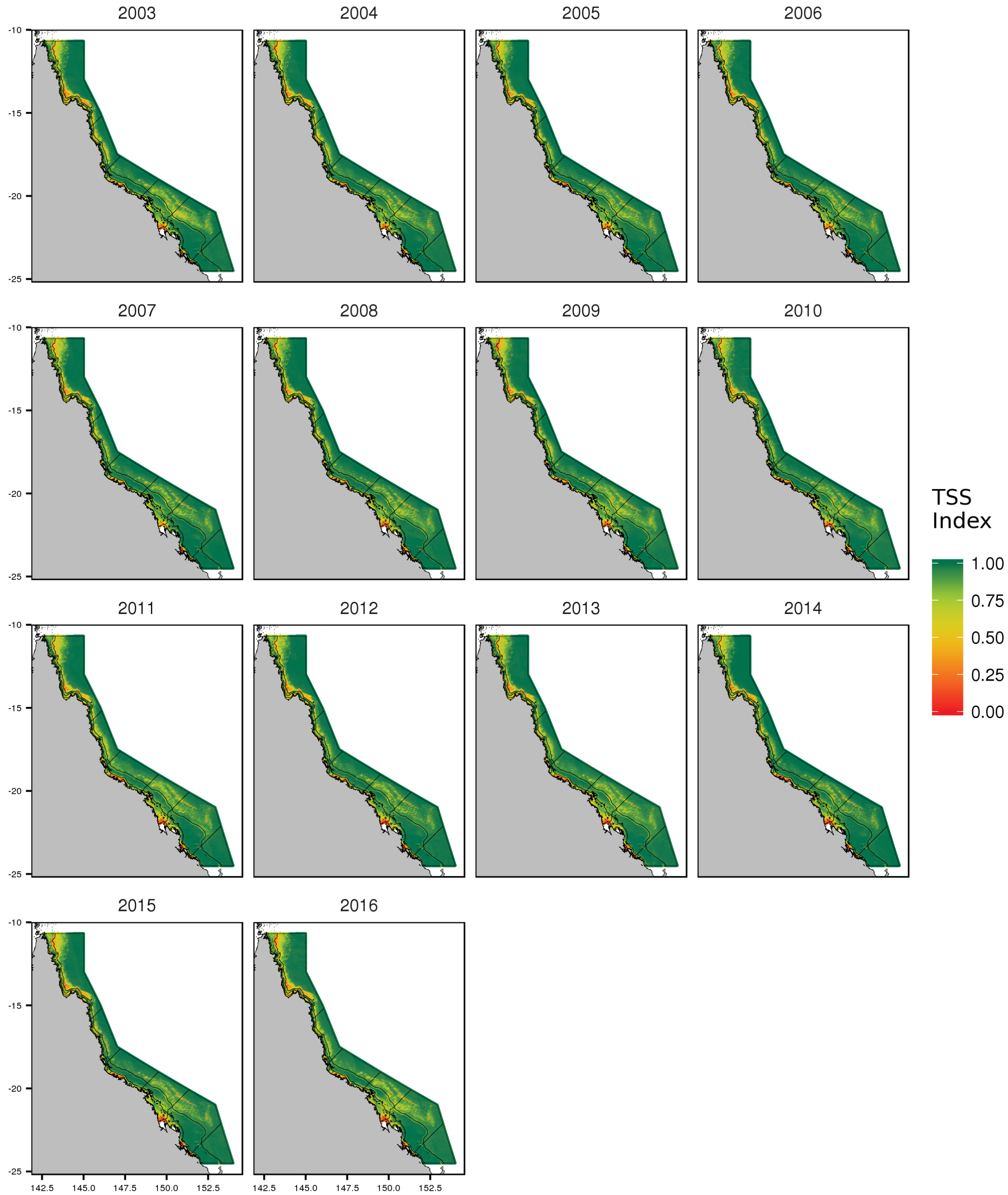


Figure 60: Spatio-temporal Satellite fsMAMP TSS index scores.

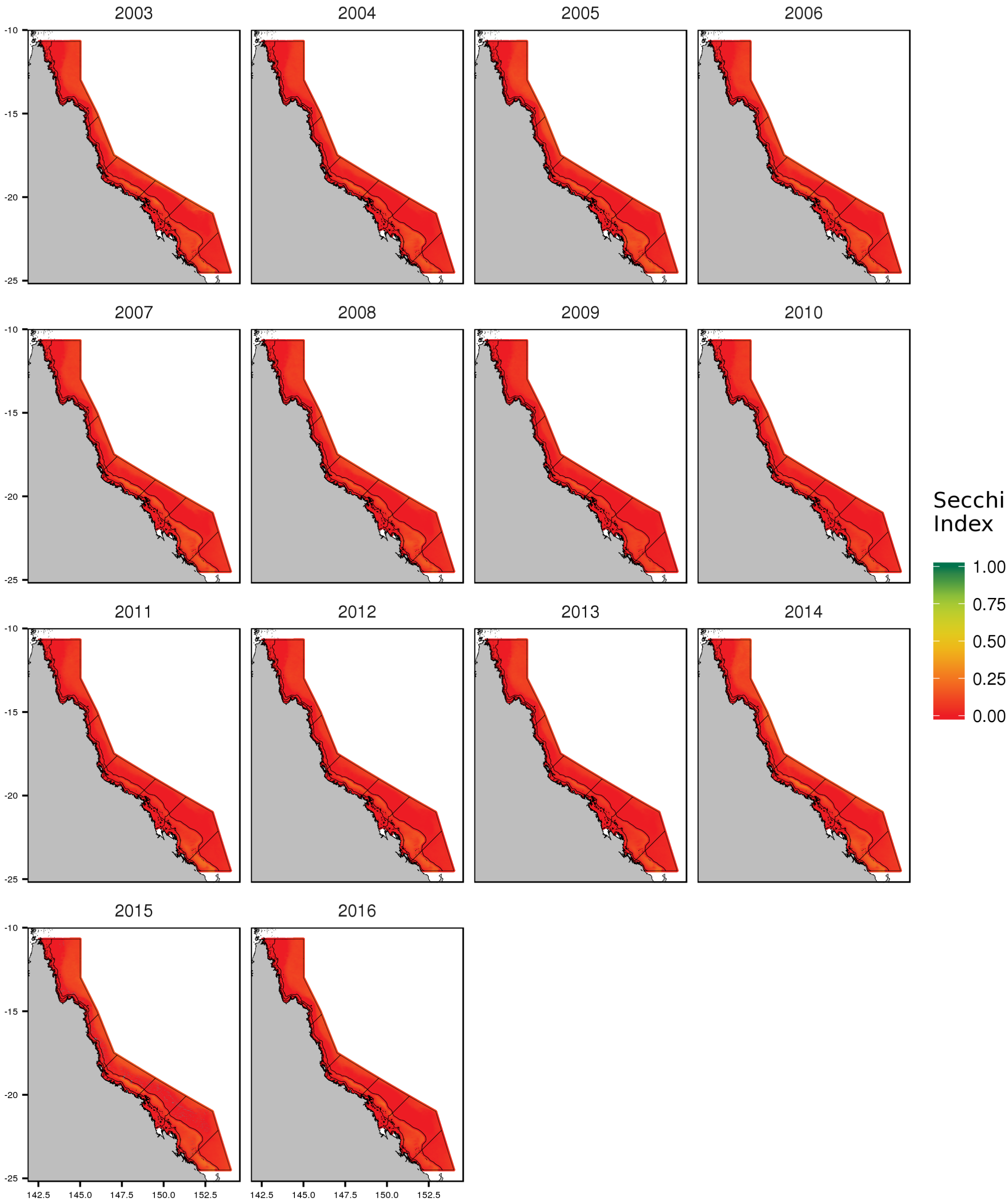


Figure 61: Spatio-temporal Satellite fsMAMP Secchi depth index scores.

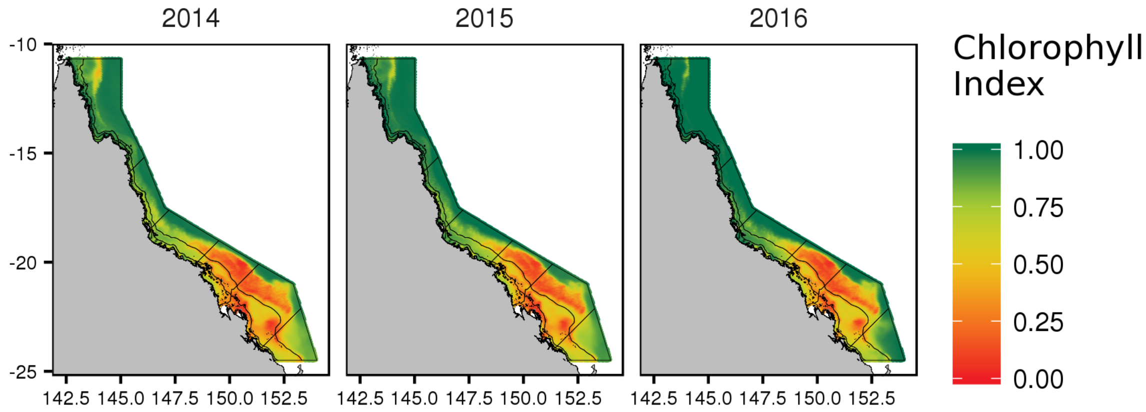


Figure 62: Spatio-temporal eReefs fsMAMP Chlorophyll-a index scores.

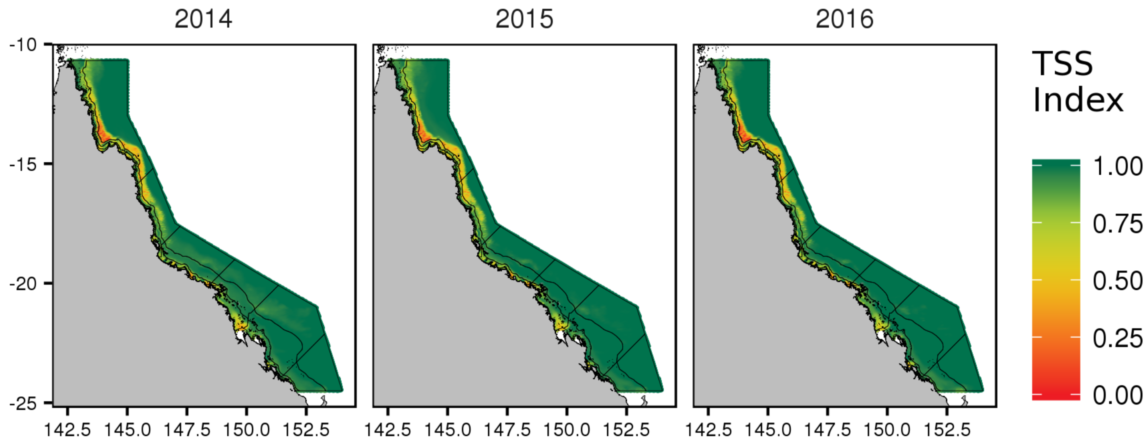


Figure 63: Spatio-temporal eReefs fsMAMP TSS index scores.

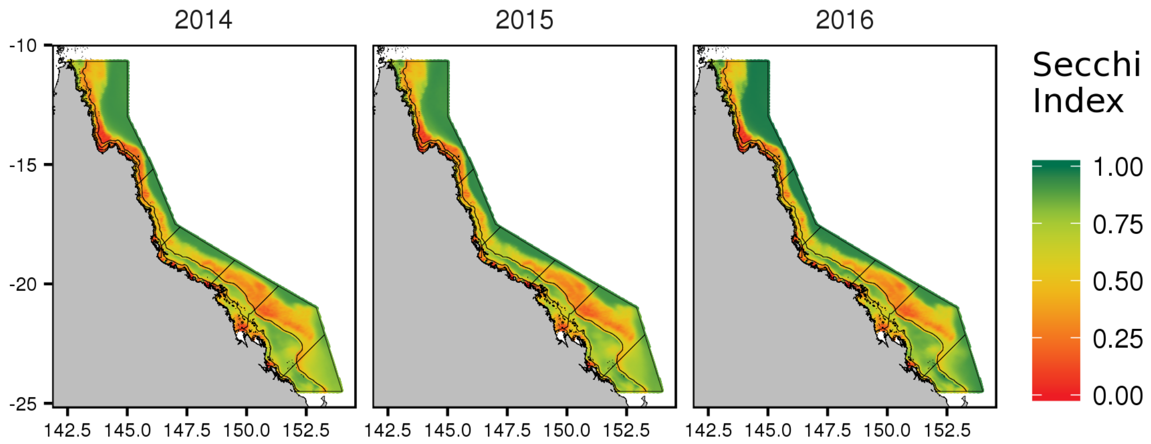


Figure 64: Spatio-temporal eReefs fsMAMP Secchi depth index scores.

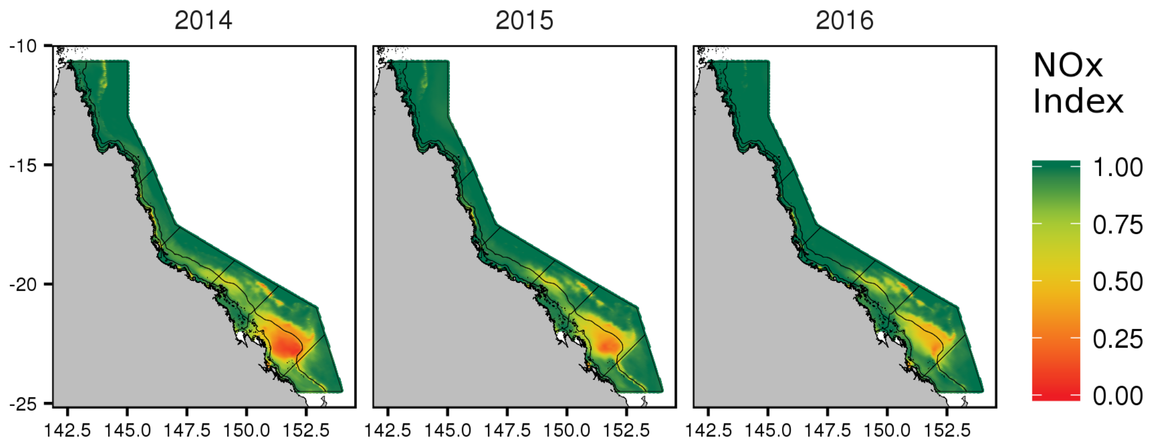


Figure 65: Spatio-temporal eReefs fsMAMP NOx index scores.

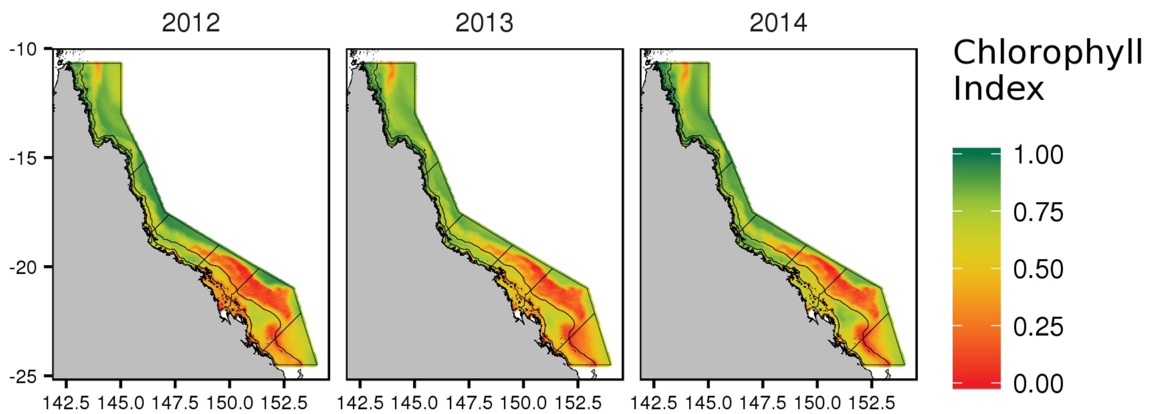


Figure 66: Spatio-temporal eReefs926 fsMAMP Chlorophyll-a index scores.

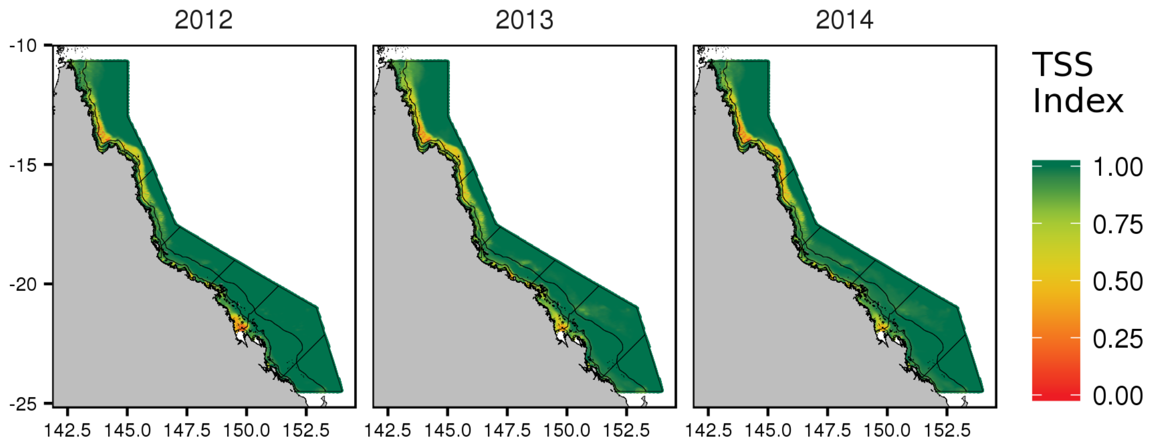


Figure 67: Spatio-temporal eReefs926 fsMAMP TSS index scores.

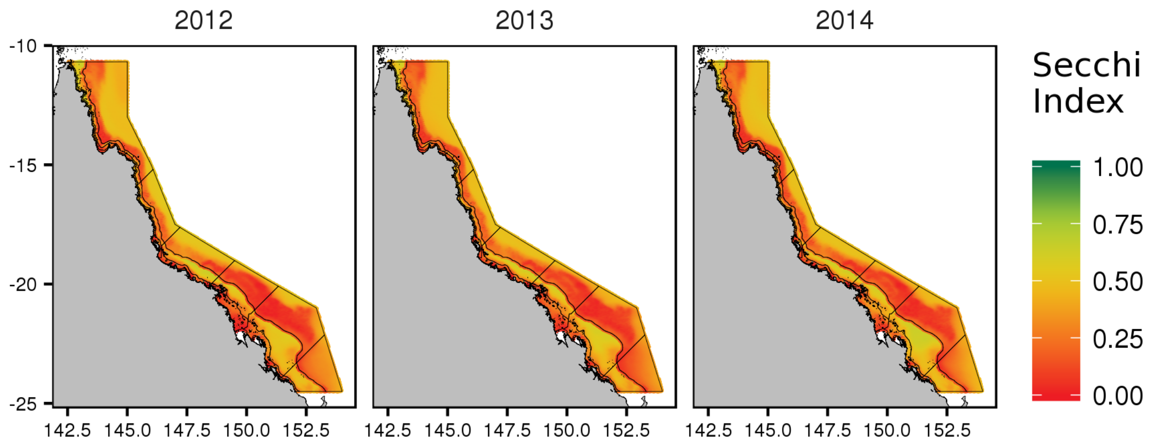


Figure 68: Spatio-temporal eReefs926 fsMAMP Secchi depth index scores.

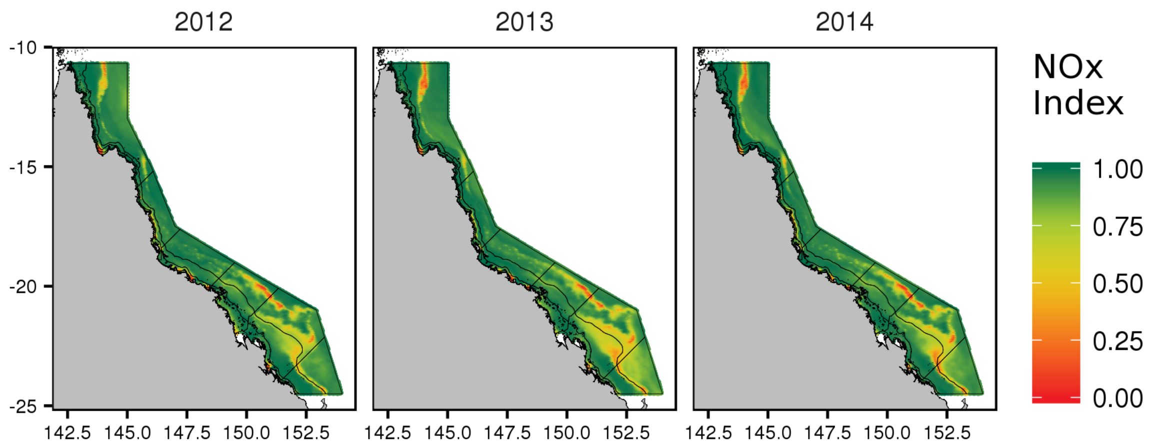


Figure 69: Spatio-temporal eReefs926 fsMAMP NOx index scores.

5.5 Summary of recommendations

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 70).

Figure 70: 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..

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 71.

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

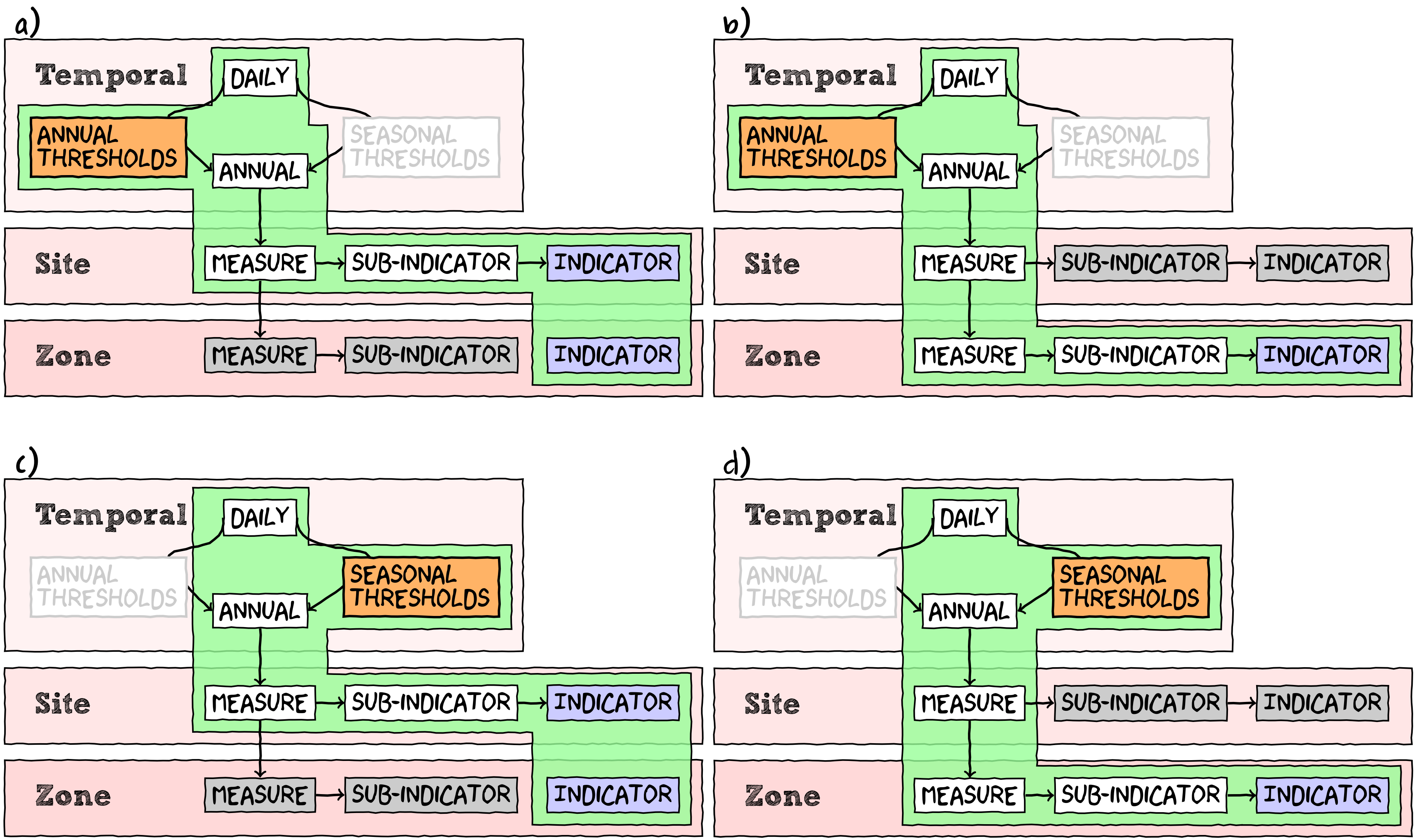


Figure 71: Schematic illustrating four possible aggregation routines through the combination of Temporal (Daily, Seasonal and Annual), Spatial (Site, Zone) and Measure (Measure, Sub-indicator, Indicator) nodes of the triple hierarchical aggregation routine associated with the GBR Report Card. Aggregation 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[8] 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 72). 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.

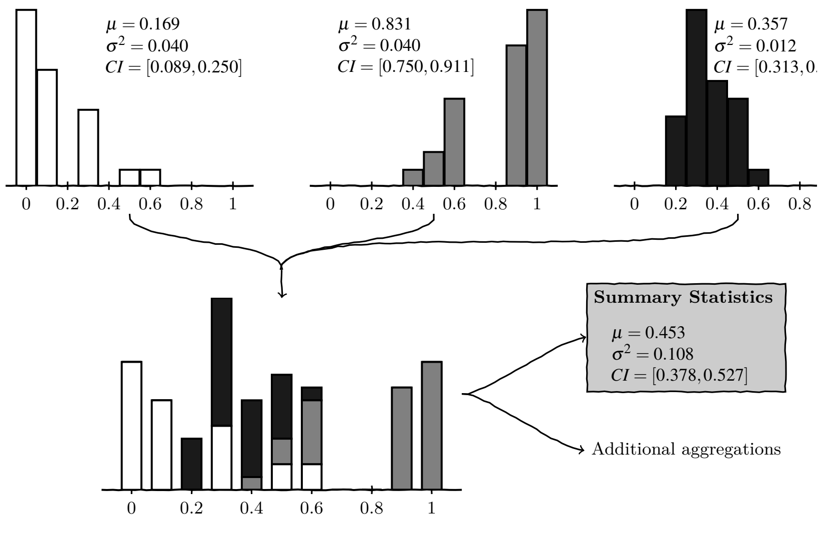


Figure 72: Illustration of Bootstrapped aggregation of three distributions. Simple summary statistics (mean, variance and 95% confidence interval presented for each distribution).

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-

[8]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[9]. Bootstrapped distributions can be aggregated (added together) to yield accumulated child distributions that retain the combined properties of both parents (see Figure 72). 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)[10]). 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[11]. 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[12]. 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

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

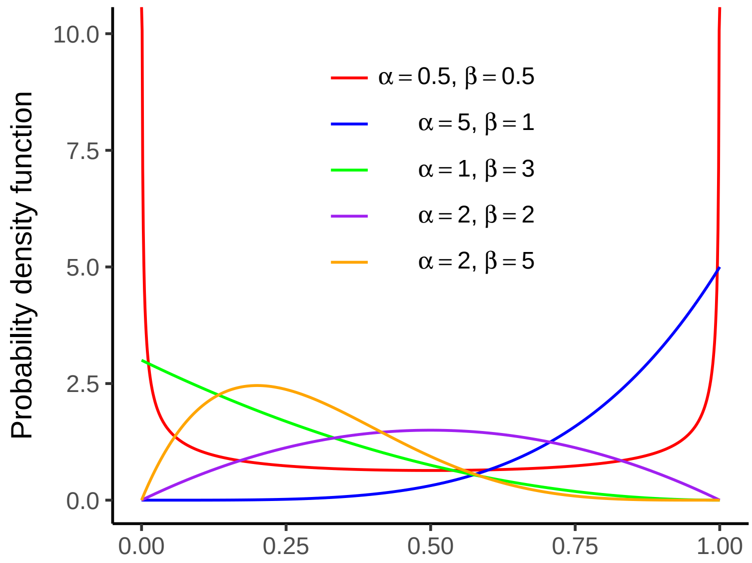


Figure 73: Beta probability densities

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[13] (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.

[13] Since aggregation occurs across two hierarchies (the Measure hierarchy and the Spatial hierarchy - see Figures 70 and 71), 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 74).

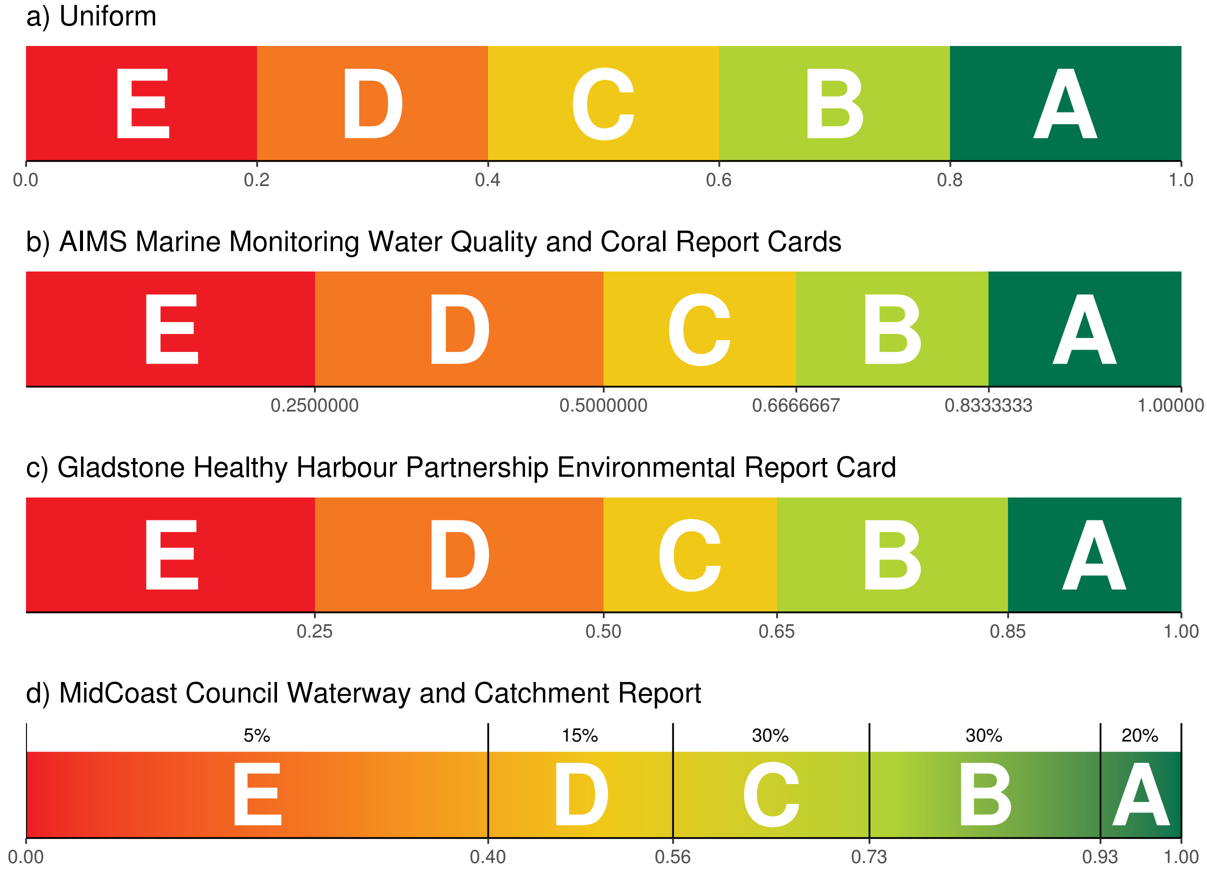


Figure 74: Score to grade conversion control charts. In each case, the scale along the base defines the grade boundaries.

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. 74d). 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. 74a).

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 74 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	0.824	B	A	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
Cape York	Midshelf	2015	0.764	B	B	B

..continued from previous page

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
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
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

..continued from previous page

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
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
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

..continued from previous page

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
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
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

..continued from previous page

Region	Water Body	Water Year	Score	Grade (MMP)	Grade (Uniform)	Grade (GHHP)
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
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[14] 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 [15] mean of the values. The 95% CI is thus calculated as the quantiles (p=0.025 and p=0.975) of the means.

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)

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

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

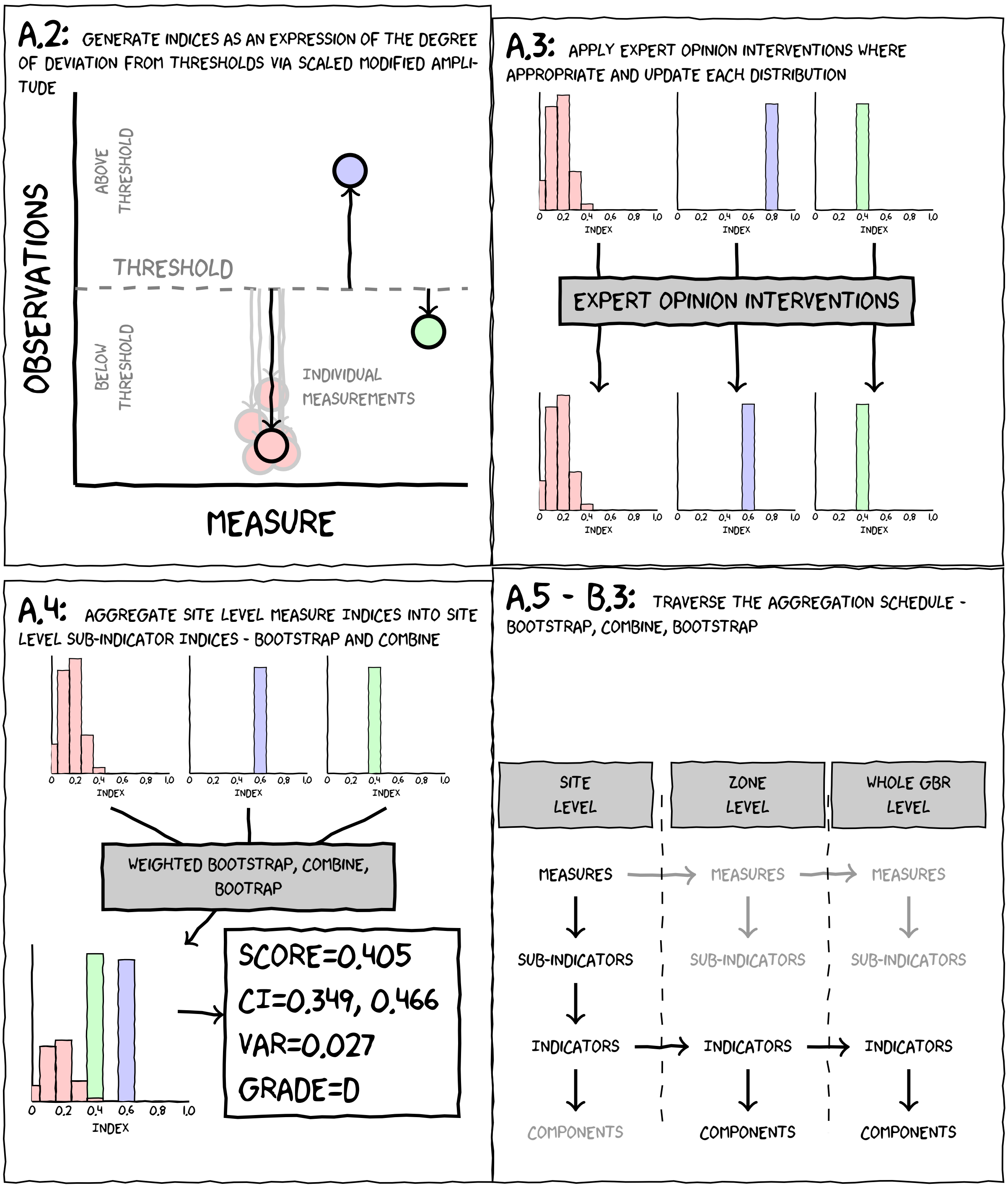
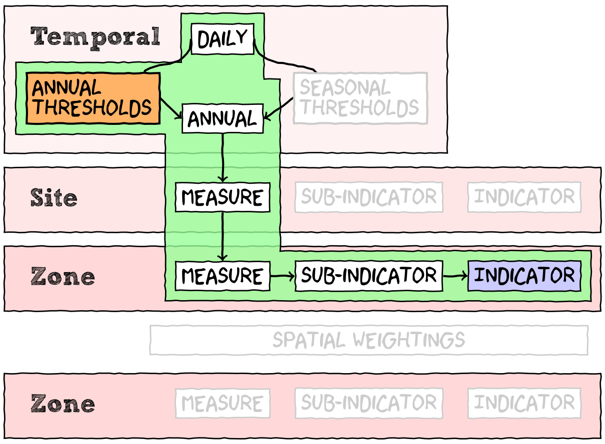
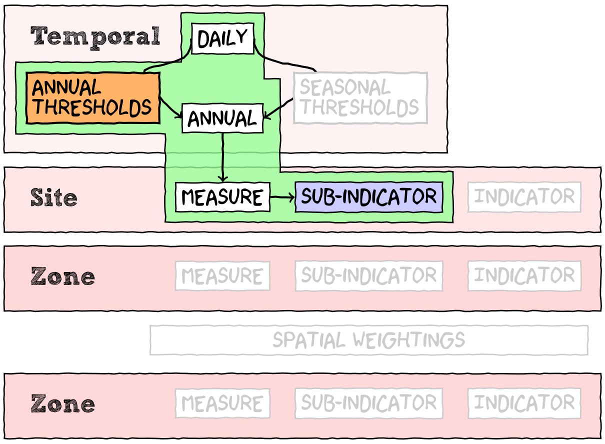


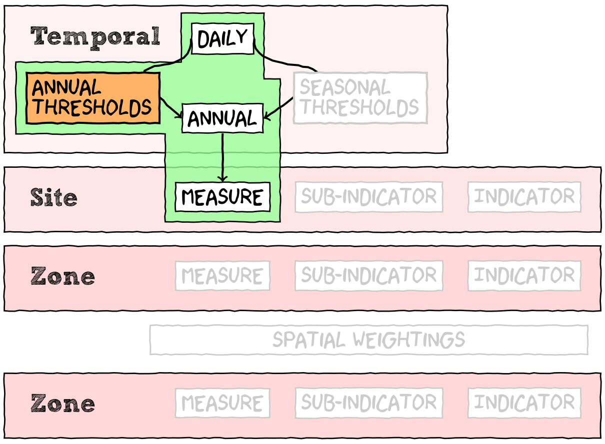
Figure 75: Schematic illustrating the major steps of the GBR Report Card. In this fabricated example, there are three Measures (Red, Green and Blue). Each of the Blue and Green Measures are represented by a single discrete observation, whereas the Red Measure is represented by a large collection of observations. Expert option intervened to lower the blue Measure distribution from observed values at 0.8 to 0.6.

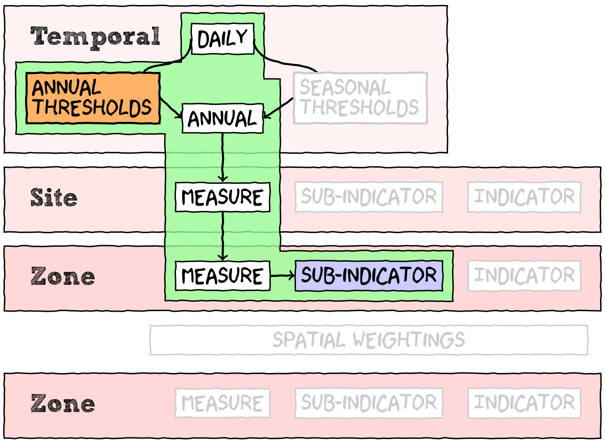
6.3 Aggregation summaries

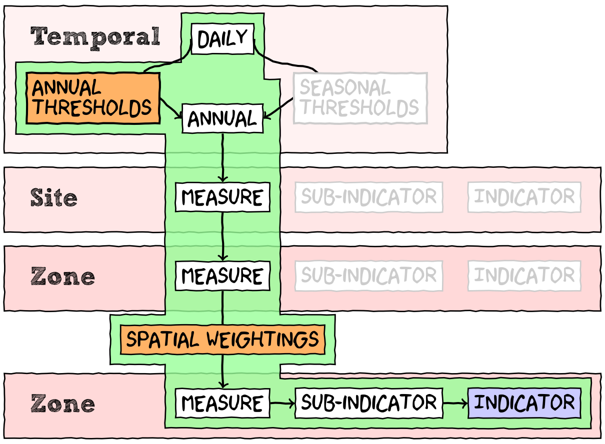
ISP have indicated that metric should be based purely on eReefs fsMAMP indexed Chlorophyll-a and Secchi Depth for Open Coastal, Midshelf and Offshore. Other aggregation combinations can be found in Appendix ??

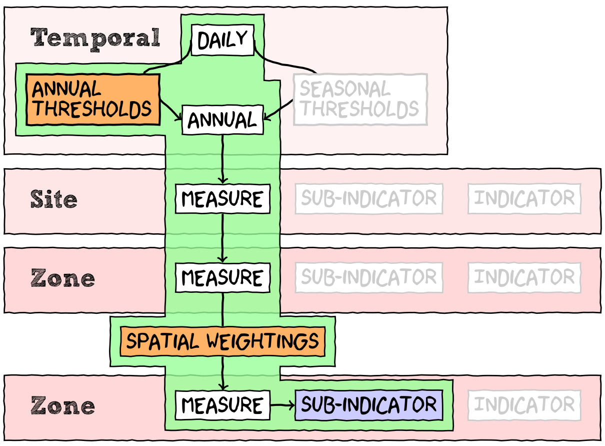


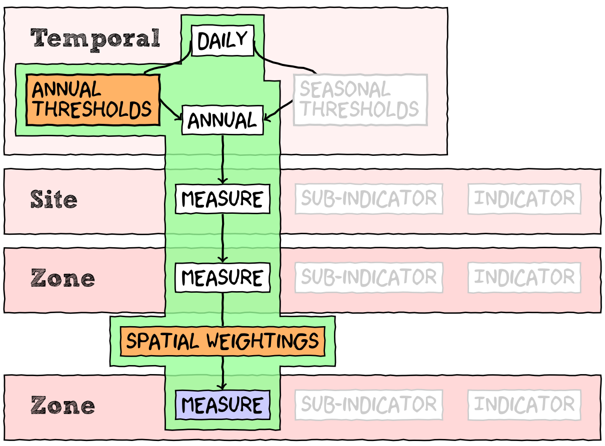




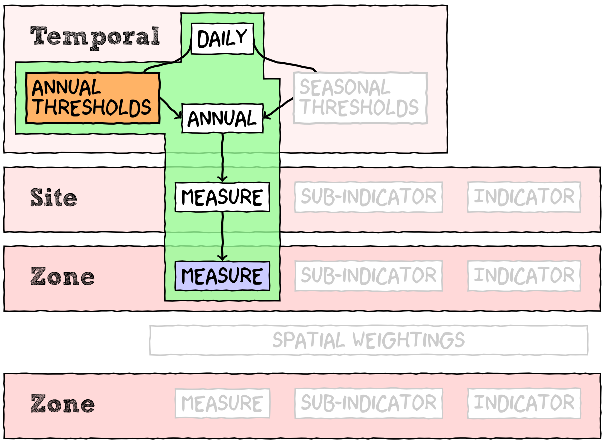

Temporal
DAILY
ANNUAL THRESHOLDS
SEASONAL THRESHOLDS
ANNUAL
Site
MEASURE
SUB-INDICATOR
INDICATOR
Zone
MEASURE
SUB-INDICATOR
INDICATOR
SPATIAL WEIGHTINGS
Zone
MEASURE
SUB-INDICATOR
INDICATOR


Temporal
DAILY
ANNUAL THRESHOLDS
SEASONAL THRESHOLDS
ANNUAL
Site
MEASURE
SUB-INDICATOR
INDICATOR
Zone
MEASURE
SUB-INDICATOR
INDICATOR
SPATIAL WEIGHTINGS
Zone
MEASURE
SUB-INDICATOR
INDICATOR


Temporal
DAILY
ANNUAL THRESHOLDS
SEASONAL THRESHOLDS
ANNUAL
Site
MEASURE
SUB-INDICATOR
INDICATOR
Zone
MEASURE
SUB-INDICATOR
INDICATOR
SPATIAL WEIGHTINGS
Zone
MEASURE
SUB-INDICATOR
INDICATOR


Temporal
DAILY
ANNUAL THRESHOLDS
SEASONAL THRESHOLDS
ANNUAL
Site
MEASURE
SUB-INDICATOR
INDICATOR
Zone
MEASURE
SUB-INDICATOR
INDICATOR
SPATIAL WEIGHTINGS
Zone
MEASURE
SUB-INDICATOR
INDICATOR

6.3.1 Measure/Zone



6.3.1.1 Simple time series

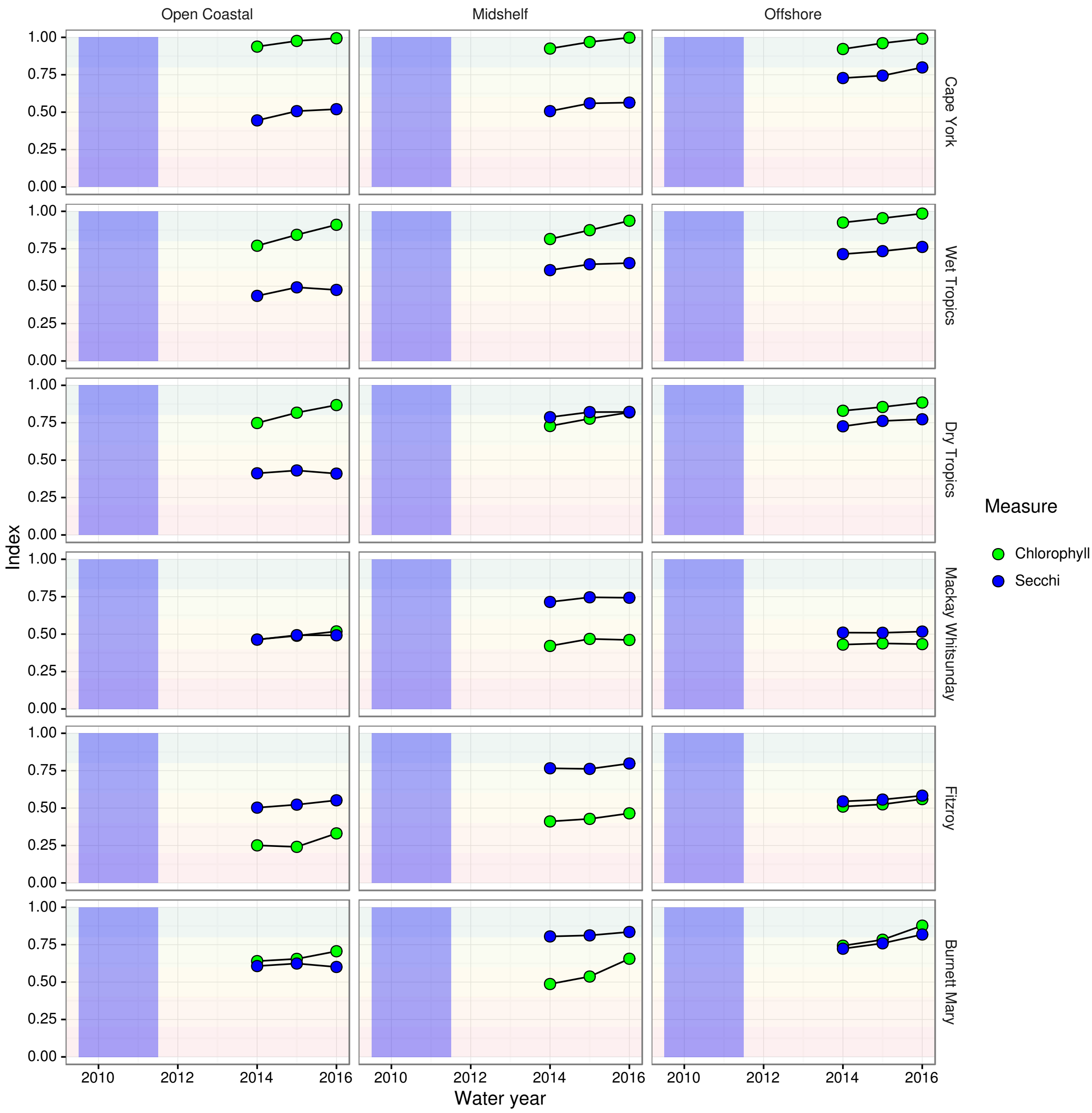


Figure 76: Time series of fsMAMP Chlorophyll-a and Secchi Depth index scores by zone. The blue vertical bar spans from mid 2009 to mid 2011. Faint colored horizontal bands represent Uniform grade ranges.

6.3.1.2 *Flat map*

Figure 77: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Chlorophyll-a index grades (Uniform grade type control chart applied).

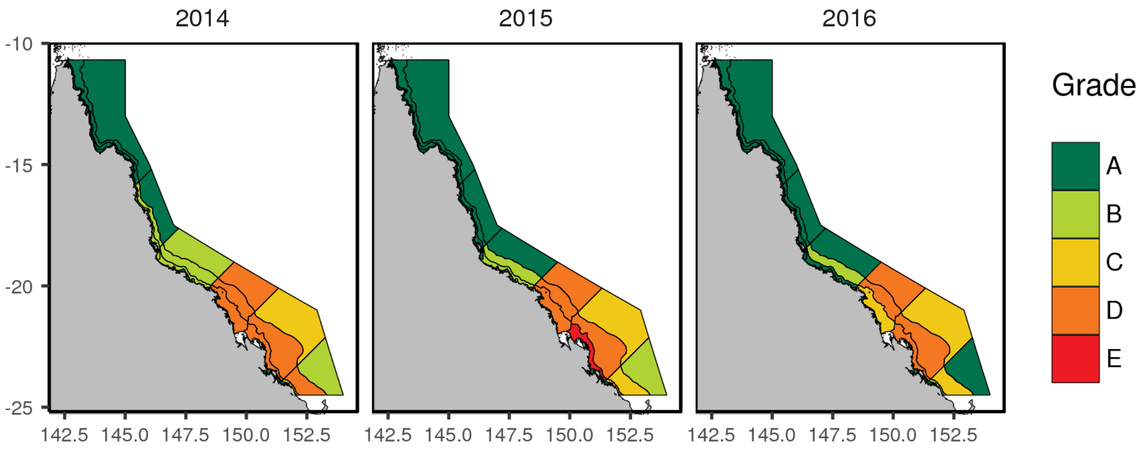


Figure 78: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Chlorophyll-a index grades (MMP-like grade type control chart applied).

Figure 79: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Secchi Depth index grades (Uniform grade type control chart applied).

Figure 80: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Secchi Depth index grades (MMP-like grade type control chart applied).

6.3.1.3 Mosaic plots

Figure 81: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Chlorophyll-a index grades (Uniform grade type control chart applied).

Figure 82: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Chlorophyll-a index grades (Uniform grade type control chart applied).

6.3.2 Subindicator/Zone

6.3.2.1 Simple time series

Figure 83: 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.

6.3.3 Indicator/Site

Figure 84: Spatio-temporal Satellite fsMAMP Water Quality index scores for each data source.

Figure 85: Spatio-temporal eReefs fsMAMP Water Quality index scores for each data source.

Figure 86: Spatio-temporal eReefs926 fsMAMP Water Quality index scores for each data source.

6.3.4 Indicator/Zone

Figure 87: Time series of Satellite fsMAMP Water Quality Index grades by zone. The blue vertical bar spans from mid 2009 to mid 2011.

Figure 88: Time series of eReefs fsMAMP Water Quality index scores by zone. The blue vertical bar spans from mid 2009 to mid 2011.

Figure 89: Time series of eReefs926 fsMAMP Water Quality index scores by zone. The blue vertical bar spans from mid 2009 to mid 2011.

Figure 90: Simplified (Zone mean) Satellite spatio-temporal fsMAMP Water Quality index scores.

Figure 91: Simplified (Zone mean) eReefs spatio-temporal fsMAMP Water Quality index scores.

Figure 92: Simplified (Zone mean) eReefs926 spatio-temporal fsMAMP Water Quality index scores.

Figure 93: Time series of Satellite fsMAMP Water Quality index scores by zone

Figure 94: Time series of eReefs fsMAMP Water Quality index scores by zone

Figure 95: Time series of eReefs926 fsMAMP Water Quality index scores by zone

Compare Measures aggregated to Zone level

- dependent on selection of sources etc..

6.4 Summary of recommendations

REFERENCES

Attard, M., M. Thompson, R. Kelly, and A. Locatelli, 2012. Tamar Estuary Ecosystem Health Assessment Program Monitoring Report 2012. Report to NRM North Tamar Estuary and Esk Rivers Program. Technical report, .

Australian and New Zealand Environment and Conservation Council, 2000. Australian Guidelines for Water Quality Monitoring and Reporting. Technical report, Environment Australia, Canberra.

Borja, A., D. M. Dauer, and A. Gremare. 2012. The importance of setting targets and reference conditions in assessing marine ecosystem quality. Ecological Indicators **12**:1 –7.

CCME, 2001. Canadian water quality guidelines for the protection of aquatic life: CCME Water Quality Index 1.0, Technical Report. Technical report, CCME, Winnipeg, Manitoba, Canada.

Coates, S., A. Waugh, A. Anwar, and M. Robson. 2007. Efficacy of multimetric fish index as an analysis tool for the transitional fish component of the water framework directive. Marine Pollution Bulletin **55**:255 – 240.

Connolly, R. M., S. Bunn, M. Campbell, B. Escher, J. Hunter, P. Maxwell, T. Page, S. Richmond, D. Rissik, A. Roiko, J. Smart, and P. Teasdale, 2013. Review of the use of report cards for monitoring ecosystem and waterway health. Report to: Gladstone Healthy Harbour Partnership. Technical report, , Queensland, Australia.

Dauvin, J. C., C. Fisson, J. Garnier, R. Lafite, T. Ruellet, J. Billen, G.and Deloffre, and R. Verney. 2008. A report card and quality indicators for the Seine estuary: From scientific approach to operational tool. Marine Pollution Bulletin **57**:187 – 201.

Dobbie, M. J., and D. Dail. 2013. Robustness and sensitivity of weighting and aggregation in constructing composite indices. Ecological Indicators .

Doran, G. T. 1981. There's a S.M.A.R.T. way to write management's goals and objectives. Management Review (AMA FORUM) **70**:35 – 36.

Emerson, J., A. Hsu, M. Levy, A. de Sherbinin, V. Mara, D. Esty, and M. Jaiteh, 2012. Environmental Performance Index and Pilot Trend Environmental Performance Index. Technical report, Yale Center for Environmental Law and Policy, New Haven.

EPA/EC, 1995. State of the Great Lakes 1995. Washington (DC). Technical report, Environmental Protection Agency and Environment Canada,.

Flint, N., J. Rolfe, C. Jones, C. Sellens, A. Rose, and L. Fabbro, 2012. Technical review for the development of an ecosystem health index and report card for the Fitzroy Partnership for river health. Part A: Review of ecosystem health indicators for the Fitzroy Basin. Technical report, Centre for Environmental Management, Central Queensland University.

Fox, D., 2013. Statistical issues associated with the development of an ecosystem report card. Technical report, Environmetrics Australia.

Gladstone Healthy Harbour Partnership, 2016. Technical Report, Gladstone Harbour Report Card 2016, GHHP Technical Report No.3. Technical report, Gladstone Healthy Harbour Partnership, Gladstone.

Hallett, C. S., F. J. Valesini, K. R. Clarke, S. A. Hesp, and S. D. Hoeksema. 2012. Development and validation of fish-based, multimetric indices for assessing the ecological health of Western Australian estuaries. Estuarine, Coastal and Shelf Science **104** – **105**:102 – 113.

Harwell, M., V. Myers, T. Young, A. Bartuska, N. Gassman, J. H. Gentile, S. Appelbaum, J. Barko, B. Causey, C. Johnson, A. McLean, R. Smola, P. Templet, and S. Tosini. 1999. A Framework for an Ecosystem Integrity Report Card: Examples from South Florida show how an ecosystem report card links societal values and scientific information. BioScience **49**:543–556. URL `http://www.jstor.org/stable/10.1525/bisi.1999.49.7.543`.

Hijuelos, A., and D. Reed, 2013. Methodology for Producing a Coastal Louisiana Report Card, September 13, 2013. The Water Institute of the Gulf. Technical report, .

Hurley, T., R. Sadiq, and A. Mazumder. 2012. Adaptation and evaluation of the Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) for use as an effective tool to characterize drinking source water quality. Water Resources **46**:3544 – 3552.

James, A., C, J. Kershner, J. Samhouri, S. O'Neil, and P. S. Levin. 2012. A Methodology for Evaluating and Ranking Water Quantity Indicators in Support of Ecosystem-Based Management. Envirnmental Management **49**:703 – 719.

Jones, C., J. Flint, Rolfe, C. Sellens, and L. Fabbro, 2013. Technical review for the development of an ecosystem health index and report card for the Fitzroy Partnership for river health. Part B: Methodology and data analysis to support an ecosystem health index and report card for the Fitzroy Basin. Technical report, Centre for Environmental Management, Central Queensland University.

Lønborg, C., M. Devlin, R. Brinkman, P. Costello, E. da Silva, J. Davidson, K. Gunn, M. Logan, C. Petus, B. Schaffelke, M. Skuza, H. Tonin, D. Tracey, M. Wright, and I. Zagorskis, 2016. Reef Rescue Marine Monitoring Program. Annual Report of AIMS and JCU Activities 2014 to 2015– Inshore water quality monitoring. Report for the Great Barrier Reef Marine Park Authority. Technical report, Australian Institute of Marine Science and JCU TropWATER, Townsville. 170pp.

MidCoast Council, 2016. MidCoast Council 2016 Waterway and Catchment Report. Technical report, MidCoast Council, Natural Systems and Estuaries Section.

Miller, W. W., H. M. Joung, C. N. Mahannah, and J. R. Garret. 1986. Identification of Water Quality Differences in Nevada Through Index Application1. Journal of Environmental Quality **15**.

Neary, B. P., 2012. A sensitivity analysis of the Canadian Water Quality Index. A report for CCME prepared by Gartner Lee Limited, Ontario, Canada. Technical report, .

Walsh, P., and W. Wheeler, 2012. Water quality index aggregation and cost benefit analysis. Working Paper 12-05. Technical report, U.S. Environmental Protection Agency, National Center for Environmental Economics, Washington, DC.

Whittaker, G., S. Lautenbach, and M. Volk, 2012. What is a good index? Problems with statistically based indicators and the Malmquist index as alternative. *in* R. Seppelt, A. A. Voinov, S. Lange, and D. Bankamp, editors. International Environmental Modelling and Software Society (iEMSs) 2012 International Congress on Environmental Modelling and Software Managing Resources of a Limited Planet, Sixth Biennial Meeting. , Leipzig, Germany.

Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC.

Appendicies

A. THRESHOLDS

Table A1: Water Quality Threshold values for each Measure in each Zone (Region/Water Body). Thresholds values are similar to annual Guideline values. Wet and Dry represent Wet and Dry season thresholds respectively. Direction of Failure indicates whether a values higher ('H') or lower ('L') than a Threshold would constitute an exceedence. Range From and Range To represent Thresholds for Measures that have a range of optimum values (such as dissolved oxygen or pH).

Measure	Units	Water Body	Region	Threshold Annual	Threshold Dry	Threshold Wet	Direction of Faiure	Justification
chl	μgL^{-1}	Enclosed Coastal	Cape York	2.00	2.00	2.00	H	QLD WQ guidelines
chl	μgL^{-1}	Enclosed Coastal	Wet Tropics	1.10	1.10	1.10	H	There is no seasonal adjustment
chl	μgL^{-1}	Enclosed Coastal	Dry Tropics	1.00	1.00	1.00	H	
chl	μgL^{-1}	Enclosed Coastal	Mackay Whitsunday	1.30	1.30	1.30	H	
chl	μgL^{-1}	Enclosed Coastal	Fitzroy	2.00	2.00	2.00	H	
chl	μgL^{-1}	Enclosed Coastal	Burnett Mary	2.00	2.00	2.00	H	
chl	μgL^{-1}	Open Coastal	Cape York	0.45	0.63	0.32	H	GBRMPA WQ guidelines
chl	μgL^{-1}	Open Coastal	Wet Tropics	0.45	0.63	0.32	H	40% higher in summer, 30% lower in winter
chl	μgL^{-1}	Open Coastal	Dry Tropics	0.45	0.63	0.32	H	Here summer is taken as Wet Season
chl	μgL^{-1}	Open Coastal	Mackay Whitsunday	0.45	0.63	0.32	H	and winter is taken as Dry Season
chl	μgL^{-1}	Open Coastal	Fitzroy	0.45	0.63	0.32	H	
chl	μgL^{-1}	Open Coastal	Burnett Mary	0.45	0.63	0.32	H	
chl	μgL^{-1}	Midshelf	Cape York	0.45	0.63	0.32	H	GBRMPA WQ guidelines
chl	μgL^{-1}	Midshelf	Wet Tropics	0.45	0.63	0.32	H	40% higher in summer, 30% lower in winter
chl	μgL^{-1}	Midshelf	Dry Tropics	0.45	0.63	0.32	H	Here summer is taken as Wet Season
chl	μgL^{-1}	Midshelf	Mackay Whitsunday	0.45	0.63	0.32	H	and winter is taken as Dry Season
chl	μgL^{-1}	Midshelf	Fitzroy	0.45	0.63	0.32	H	
chl	μgL^{-1}	Midshelf	Burnett Mary	0.45	0.63	0.32	H	
chl	μgL^{-1}	Offshore	Cape York	0.40	0.56	0.28	H	GBRMPA WQ guidelines
chl	μgL^{-1}	Offshore	Wet Tropics	0.40	0.56	0.28	H	40% higher in summer, 30% lower in winter
chl	μgL^{-1}	Offshore	Dry Tropics	0.40	0.56	0.28	H	Here summer is taken as Wet Season
chl	μgL^{-1}	Offshore	Mackay Whitsunday	0.40	0.56	0.28	H	and winter is taken as Dry Season
chl	μgL^{-1}	Offshore	Fitzroy	0.40	0.56	0.28	H	
chl	μgL^{-1}	Offshore	Burnett Mary	0.40	0.56	0.28	H	
nap	mgL^{-1}	Enclosed Coastal	Cape York	10.00	10.00	10.00	H	QLD WQ guidelines
nap	mgL^{-1}	Enclosed Coastal	Wet Tropics	10.00	10.00	10.00	H	There is no seasonal adjustment and
nap	mgL^{-1}	Enclosed Coastal	Dry Tropics	10.00	10.00	10.00	H	values for CY and WT are not determined
nap	mgL^{-1}	Enclosed Coastal	Mackay Whitsunday	10.00	10.00	10.00	H	Suggest applying same ratio as for turbidity
nap	mgL^{-1}	Enclosed Coastal	Fitzroy	15.00	15.00	15.00	H	between CY/WT and others, i.e (15*10)/6=25
nap	mgL^{-1}	Enclosed Coastal	Burnett Mary	15.00	15.00	15.00	H	NAP is taken as = TSS in this context
nap	mgL^{-1}	Open Coastal	Cape York	2.00	2.40	1.60	H	GBRMPA WQ guidelines
nap	mgL^{-1}	Open Coastal	Wet Tropics	2.00	2.40	1.60	H	20% higher in summer, 20% lower in winter
nap	mgL^{-1}	Open Coastal	Dry Tropics	2.00	2.40	1.60	H	Here summer is taken as Wet Season
nap	mgL^{-1}	Open Coastal	Mackay Whitsunday	2.00	2.40	1.60	H	and winter is taken as Dry Season
nap	mgL^{-1}	Open Coastal	Fitzroy	2.00	2.40	1.60	H	
nap	mgL^{-1}	Open Coastal	Burnett Mary	2.00	2.40	1.60	H	NAP is taken as = TSS in this context
nap	mgL^{-1}	Midshelf	Cape York	2.00	2.40	1.60	H	GBRMPA WQ guidelines
nap	mgL^{-1}	Midshelf	Wet Tropics	2.00	2.40	1.60	H	20% higher in summer, 20% lower in winter
nap	mgL^{-1}	Midshelf	Dry Tropics	2.00	2.40	1.60	H	Here summer is taken as Wet Season
nap	mgL^{-1}	Midshelf	Mackay Whitsunday	2.00	2.40	1.60	H	and winter is taken as Dry Season
nap	mgL^{-1}	Midshelf	Fitzroy	2.00	2.40	1.60	H	

...continued from previous page

Measure	Units	Water Body	Region	Threshold Annual	Threshold Dry	Threshold Wet	Direction of Faiure	Justification
nap	mgL^{-1}	Midshelf	Burnett Mary	2.00	2.40	1.60	H	NAP is taken as = TSS in this context
nap	mgL^{-1}	Offshore	Cape York	0.70	0.84	0.56	H	GBRMPA WQ guidelines
nap	mgL^{-1}	Offshore	Wet Tropics	0.70	0.84	0.56	H	20% higher in summer, 20% lower in winter
nap	mgL^{-1}	Offshore	Dry Tropics	0.70	0.84	0.56	H	Here summer is taken as Wet Season
nap	mgL^{-1}	Offshore	Mackay Whitsunday	0.70	0.84	0.56	H	and winter is taken as Dry Season
nap	mgL^{-1}	Offshore	Fitzroy	0.70	0.84	0.56	H	
nap	mgL^{-1}	Offshore	Burnett Mary	0.70	0.84	0.56	H	NAP is taken as = TSS in this context
ntu	NTU	Enclosed Coastal	Cape York	4.00	4.00	4.00	H	QLD WQ guidelines
ntu	NTU	Enclosed Coastal	Wet Tropics	4.00	4.00	4.00	H	There is no seasonal adjustment
ntu	NTU	Enclosed Coastal	Dry Tropics	4.00	4.00	4.00	H	
ntu	NTU	Enclosed Coastal	Mackay Whitsunday	4.00	4.00	4.00	H	Unclear as to why the CY/WT guidelines
ntu	NTU	Enclosed Coastal	Fitzroy	6.00	6.00	6.00	H	are higher than Southern regions
ntu	NTU	Enclosed Coastal	Burnett Mary	6.00	6.00	6.00	H	
ntu	NTU	Open Coastal	Cape York	1.50	1.80	1.20	H	No guideline available but turbidity needed if logger
ntu	NTU	Open Coastal	Wet Tropics	1.50	1.80	1.20	H	data is to be integrated.
ntu	NTU	Open Coastal	Dry Tropics	1.50	1.80	1.20	H	MMP used a correlation between TSS and NTU
ntu	NTU	Open Coastal	Mackay Whitsunday	1.50	1.80	1.20	H	(based on whole of GBR data) which is also used here
ntu	NTU	Open Coastal	Fitzroy	1.50	1.80	1.20	H	
ntu	NTU	Open Coastal	Burnett Mary	1.50	1.80	1.20	H	Applied 20% higher in summer, 20% lower in winter
ntu	NTU	Midshelf	Cape York	1.50	1.80	1.20	H	No guideline available but turbidity needed if logger
ntu	NTU	Midshelf	Wet Tropics	1.50	1.80	1.20	H	data is to be integrated.
ntu	NTU	Midshelf	Dry Tropics	1.50	1.80	1.20	H	MMP used a correlation between TSS and NTU
ntu	NTU	Midshelf	Mackay Whitsunday	1.50	1.80	1.20	H	(based on whole of GBR data) which is also used here
ntu	NTU	Midshelf	Fitzroy	1.50	1.80	1.20	H	
ntu	NTU	Midshelf	Burnett Mary	1.50	1.80	1.20	H	Applied 20% higher in summer, 20% lower in winter
ntu	NTU	Offshore	Cape York	1.00	1.20	0.80	H	No guideline available but turbidity needed if logger
ntu	NTU	Offshore	Wet Tropics	1.00	1.20	0.80	H	data is to be integrated.
ntu	NTU	Offshore	Dry Tropics	1.00	1.20	0.80	H	MMP used a correlation between TSS and NTU
ntu	NTU	Offshore	Mackay Whitsunday	1.00	1.20	0.80	H	(based on whole of GBR data) which is also used here
ntu	NTU	Offshore	Fitzroy	1.00	1.20	0.80	H	
ntu	NTU	Offshore	Burnett Mary	1.00	1.20	0.80	H	Applied 20% higher in summer, 20% lower in winter
sd	m	Enclosed Coastal	Cape York	1.00	1.00	1.00	L	QLD WQ guidelines
sd	m	Enclosed Coastal	Wet Tropics	1.00	1.00	1.00	L	There is no seasonal adjustment
sd	m	Enclosed Coastal	Dry Tropics	1.50	1.50	1.50	L	
sd	m	Enclosed Coastal	Mackay Whitsunday	1.00	1.00	1.00	L	Unclear as to why the CY/WT guidelines
sd	m	Enclosed Coastal	Fitzroy	1.50	1.50	1.50	L	are lower than Southern regions
sd	m	Enclosed Coastal	Burnett Mary	1.50	1.50	1.50	L	
sd	m	Open Coastal	Cape York	10.00	10.00	10.00	L	GBRMPA WQ guidelines
sd	m	Open Coastal	Wet Tropics	10.00	10.00	10.00	L	There is no seasonal adjustment
sd	m	Open Coastal	Dry Tropics	10.00	10.00	10.00	L	
sd	m	Open Coastal	Mackay Whitsunday	10.00	10.00	10.00	L	
sd	m	Open Coastal	Fitzroy	10.00	10.00	10.00	L	
sd	m	Open Coastal	Burnett Mary	10.00	10.00	10.00	L	
sd	m	Midshelf	Cape York	10.00	10.00	10.00	L	GBRMPA WQ guidelines

...continued from previous page

Measure	Units	Water Body	Region	Threshold Annual	Threshold Dry	Threshold Wet	Direction of Faiure	Justification
sd	m	Midshelf	Wet Tropics	10.00	10.00	10.00	L	There is no seasonal adjustment
sd	m	Midshelf	Dry Tropics	10.00	10.00	10.00	L	
sd	m	Midshelf	Mackay Whitsunday	10.00	10.00	10.00	L	
sd	m	Midshelf	Fitzroy	10.00	10.00	10.00	L	
sd	m	Midshelf	Burnett Mary	10.00	10.00	10.00	L	
sd	m	Offshore	Cape York	17.00	17.00	17.00	L	GBRMPA WQ guidelines
sd	m	Offshore	Wet Tropics	17.00	17.00	17.00	L	There is no seasonal adjustment
sd	m	Offshore	Dry Tropics	17.00	17.00	17.00	L	
sd	m	Offshore	Mackay Whitsunday	17.00	17.00	17.00	L	
sd	m	Offshore	Fitzroy	17.00	17.00	17.00	L	
sd	m	Offshore	Burnett Mary	17.00	17.00	17.00	L	
NOx	μgL^{-1}	Enclosed Coastal	Cape York	10.00	10.00	10.00	H	Old MMP guidelines
NOx	μgL^{-1}	Enclosed Coastal	Wet Tropics	10.00	10.00	10.00	H	There is no seasonal adjustment
NOx	μgL^{-1}	Enclosed Coastal	Dry Tropics	3.00	3.00	3.00	H	
NOx	μgL^{-1}	Enclosed Coastal	Mackay Whitsunday	3.00	3.00	3.00	H	
NOx	μgL^{-1}	Enclosed Coastal	Fitzroy	3.00	3.00	3.00	H	
NOx	μgL^{-1}	Enclosed Coastal	Burnett Mary	3.00	3.00	3.00	H	
NOx	μgL^{-1}	Open Coastal	Cape York	2.00	2.00	2.00	H	Old MMP guidelines
NOx	μgL^{-1}	Open Coastal	Wet Tropics	2.00	2.00	2.00	H	There is no seasonal adjustment
NOx	μgL^{-1}	Open Coastal	Dry Tropics	3.00	3.00	3.00	H	
NOx	μgL^{-1}	Open Coastal	Mackay Whitsunday	3.00	3.00	3.00	H	
NOx	μgL^{-1}	Open Coastal	Fitzroy	3.00	3.00	3.00	H	
NOx	μgL^{-1}	Open Coastal	Burnett Mary	3.00	3.00	3.00	H	
NOx	μgL^{-1}	Midshelf	Cape York	2.00	2.00	2.00	H	Old guidelines
NOx	μgL^{-1}	Midshelf	Wet Tropics	2.00	2.00	2.00	H	There is no seasonal adjustment
NOx	μgL^{-1}	Midshelf	Dry Tropics	2.00	2.00	2.00	H	
NOx	μgL^{-1}	Midshelf	Mackay Whitsunday	2.00	2.00	2.00	H	
NOx	μgL^{-1}	Midshelf	Fitzroy	2.00	2.00	2.00	H	
NOx	μgL^{-1}	Midshelf	Burnett Mary	2.00	2.00	2.00	H	
NOx	μgL^{-1}	Offshore	Cape York	2.00	2.00	2.00	H	Old MMP guidelines
NOx	μgL^{-1}	Offshore	Wet Tropics	2.00	2.00	2.00	H	There is no seasonal adjustment
NOx	μgL^{-1}	Offshore	Dry Tropics	2.00	2.00	2.00	H	
NOx	μgL^{-1}	Offshore	Mackay Whitsunday	2.00	2.00	2.00	H	
NOx	μgL^{-1}	Offshore	Fitzroy	2.00	2.00	2.00	H	
NOx	μgL^{-1}	Offshore	Burnett Mary	2.00	2.00	2.00	H	

B. EREEFS MODELS

Table B2: eReefs regional biogeochemical simulation catalog.

Simulation name	Projects	Date range	Delivery	Notes/Improvements
GBR4_H1p85_B1p0_Cbas_Dhnd	SIEF	Jan 1, 2011 – Jun 30, 2014	Available on NCI	Simulation delivered as part of SIEF project (previously known as 926). Skill assessment available in SIEF report.