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 C.1 beginning on page 183). 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 9 – 12 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 of the figures are presented with log-transformed y-axes as the data are typically positively skewed. 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 present 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. 9-12a). The AIMS FLNTU logger data (Fig. 9-12b) 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 are unlikely to do so. Furthermore, whilst the AIMS insitu data are predominantly collected during the dry season, the AIMS FLNTU loggers collect data across the entire year and are therefore likely to record a greater proportion of the full variation in conditions. Of course it is important when interpreting these diagnostic plots to focus mainly on the violin plots and less on the dots (representing individual observations). This is because the dots do not provide an indication of the density and it is easy to allow outliers to distort out impression of the variability of the data.

Similarly, the scale of the range eReefs and eReefs926 data (Fig. 9-12d-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. 9-12c).

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

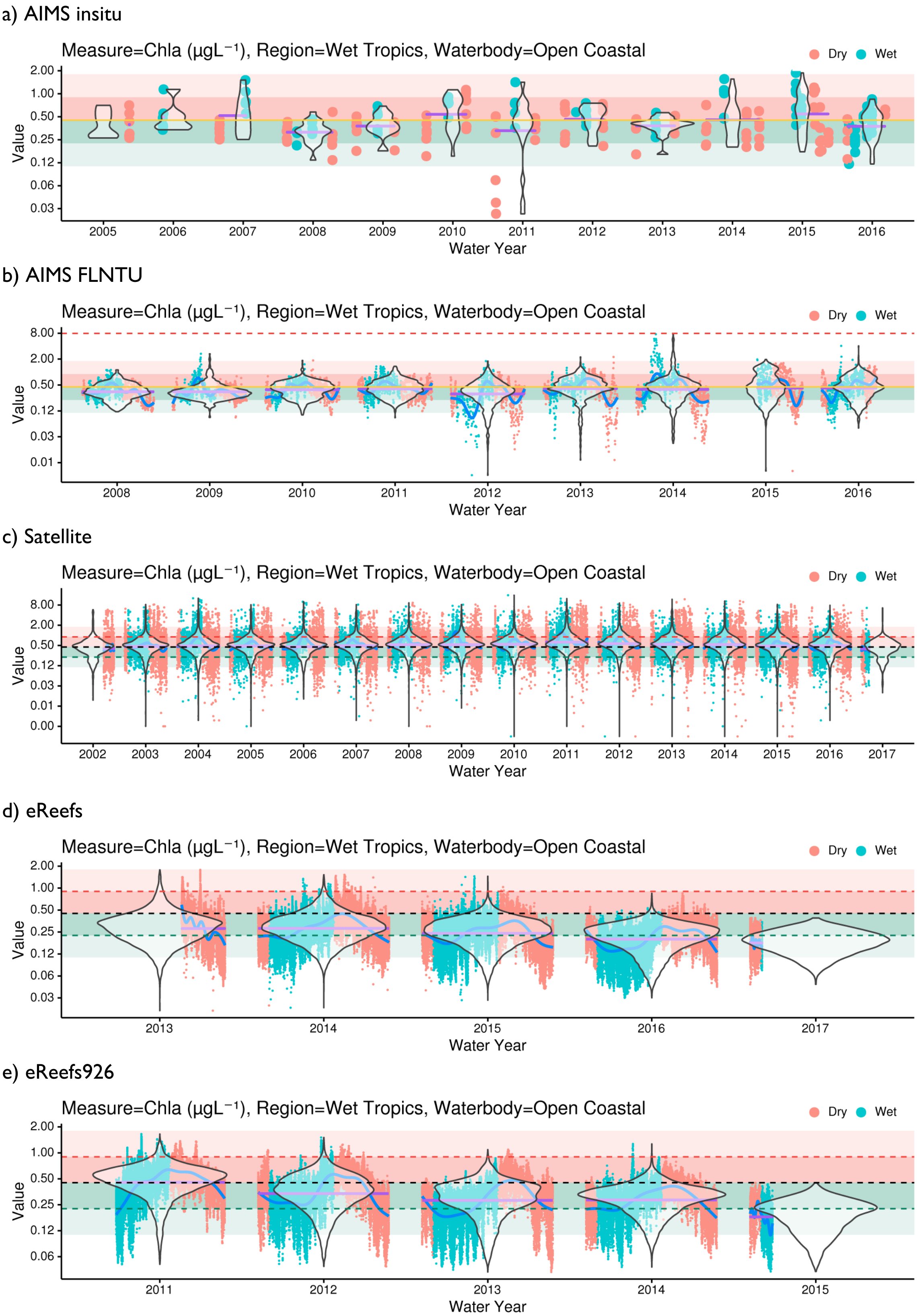


Figure 9: Observed (logarithmic axis with violin plot overlay) Chlorophyll-a data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU, c) Satellite, d) eReefs and e) eReefs926. Observations are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red 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: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

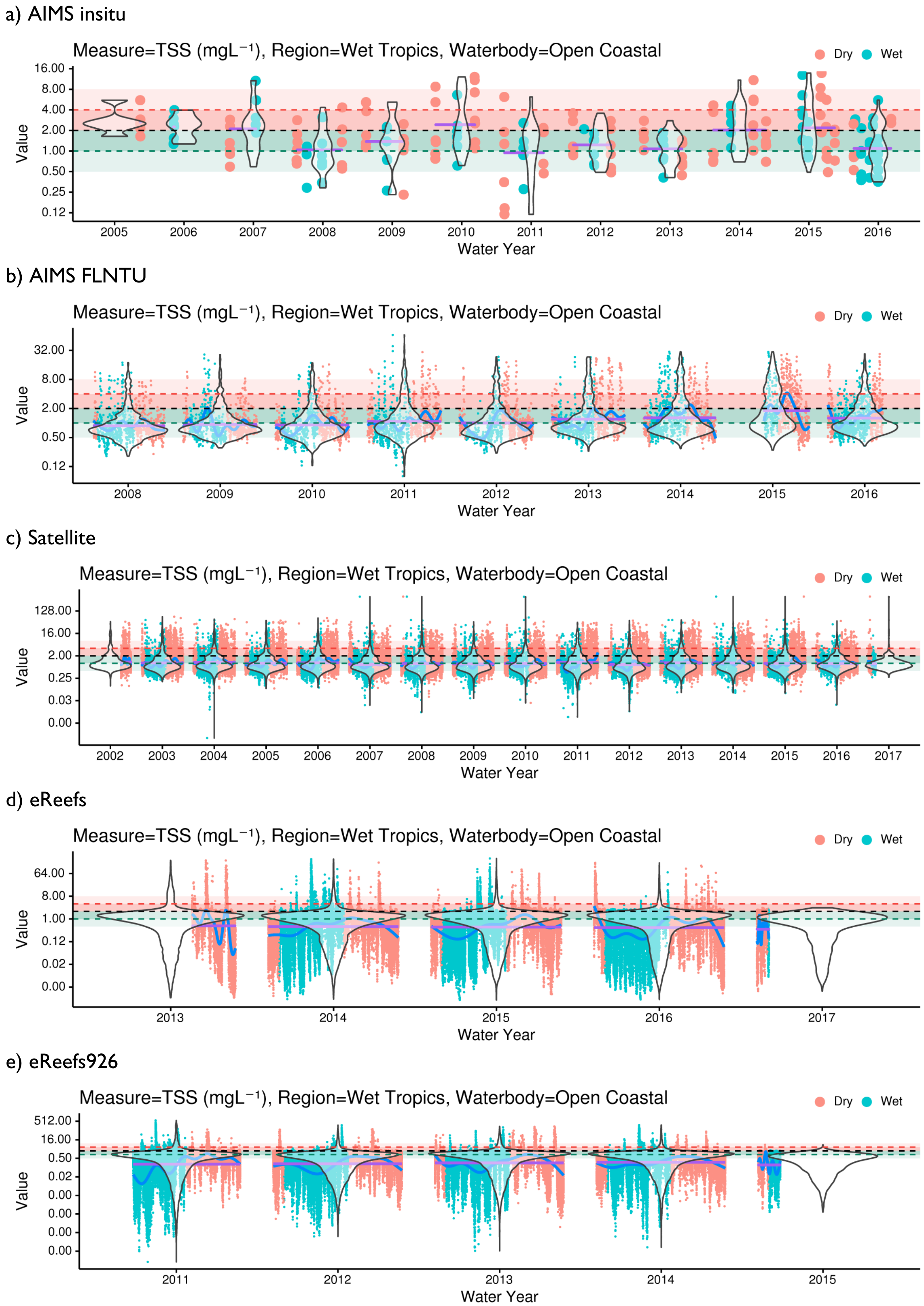
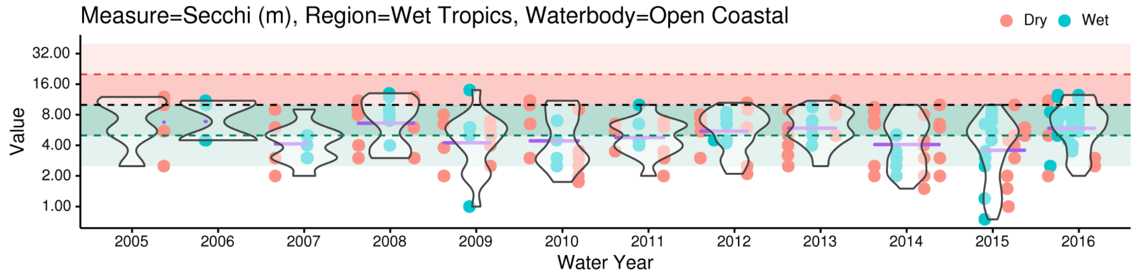
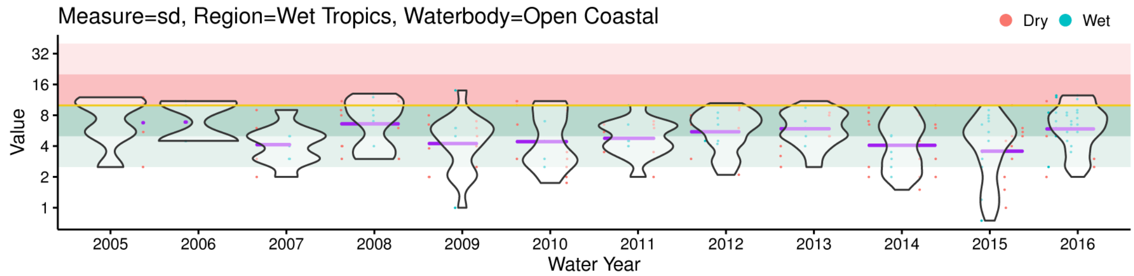


Figure 10: Observed (logarithmic axis with violin plot overlay) TSS data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU, c) Satellite, d) eReefs and e) eReefs926. Observations are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). 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: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

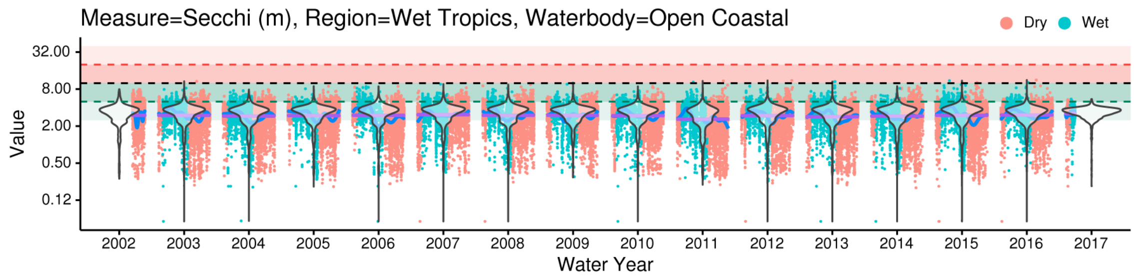
a) AIMS insitu



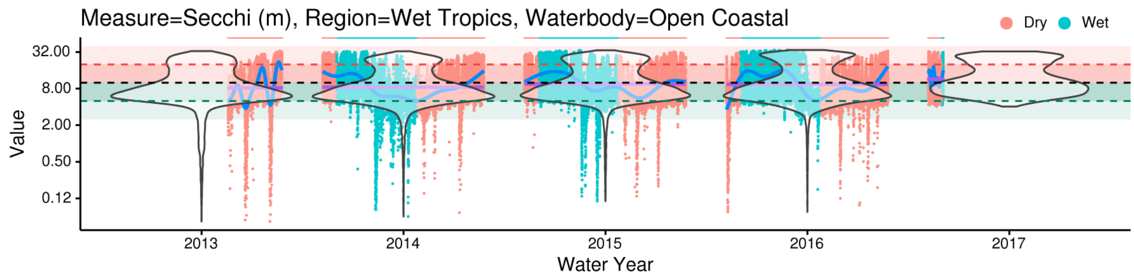
b) AIMS FLNTU



c) Satellite



d) eReefs



e) eReefs926

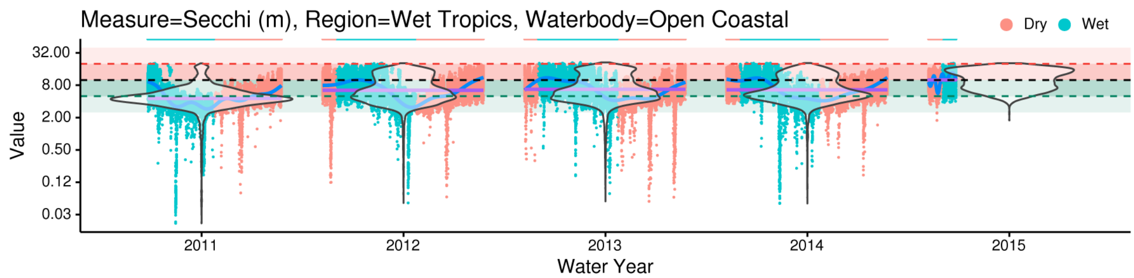
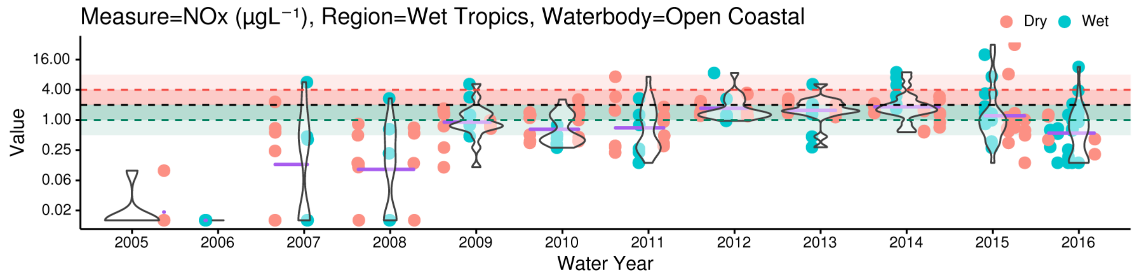
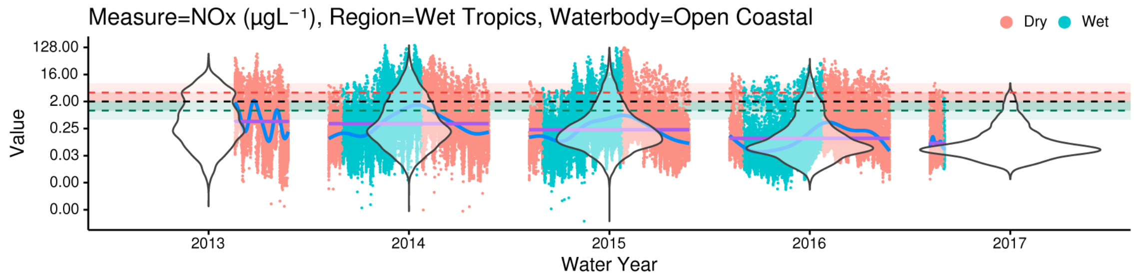


Figure 11: Observed (logarithmic axis with violin plot overlay) Secchi depth data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU, c) Satellite, d) eReefs and e) eReefs926. Observations are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red 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: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

a) AIMS insitu



b) eReefs



c) eReefs926

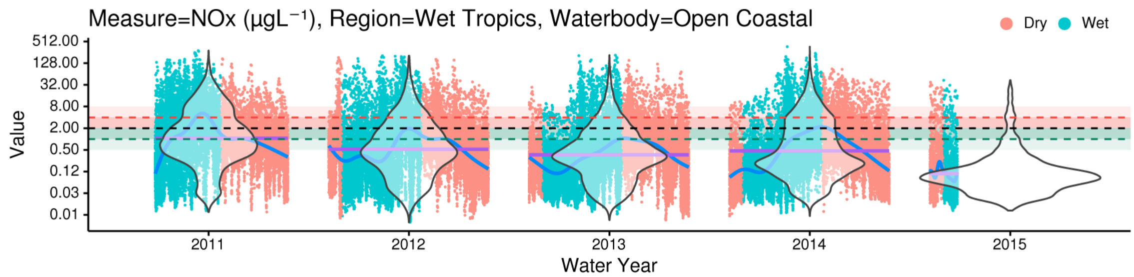


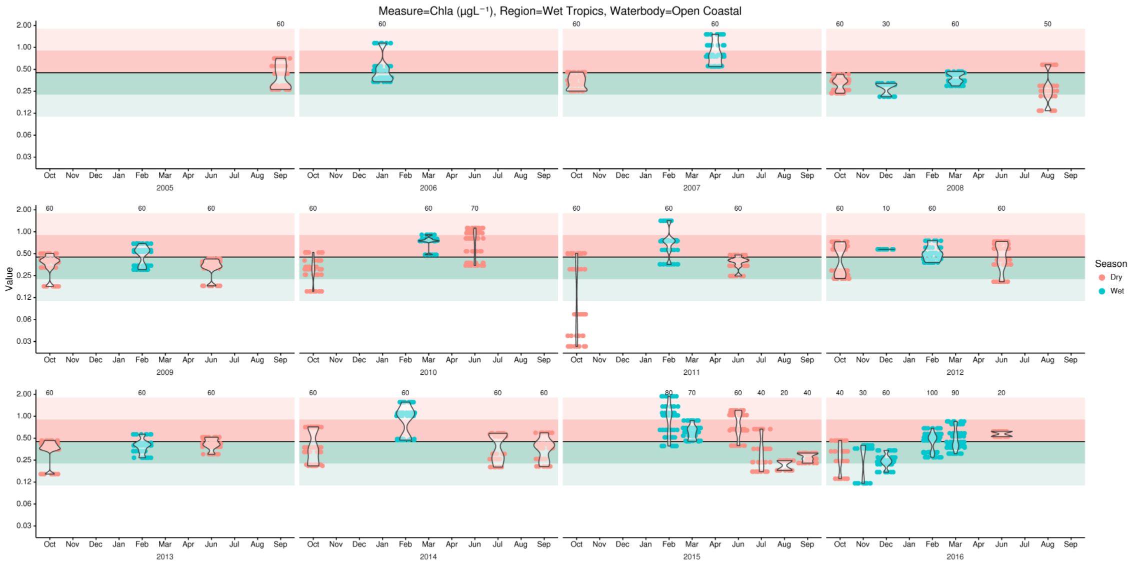
Figure 12: Observed (logarithmic axis with violin plot overlay) NOx data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) eReefs and c) eReefs926. Observations are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Blue smoother represents Generalized 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: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

4.3 Monthly data

Figures 13 – 18 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.

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 4.2 might be more easily appreciated.

a) AIMS insitu



b) AIMS FLNTU

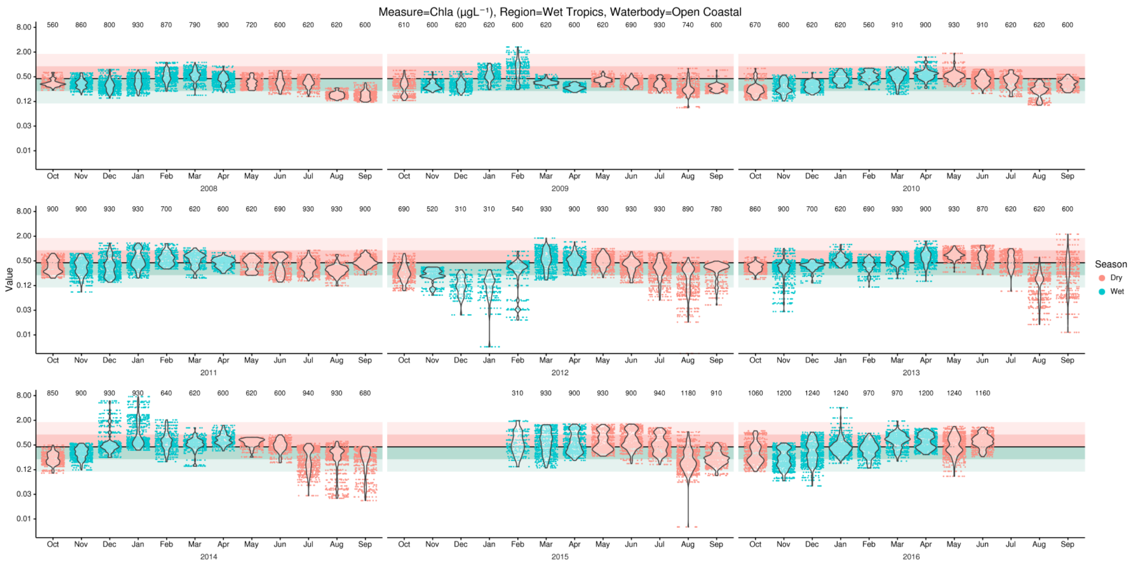
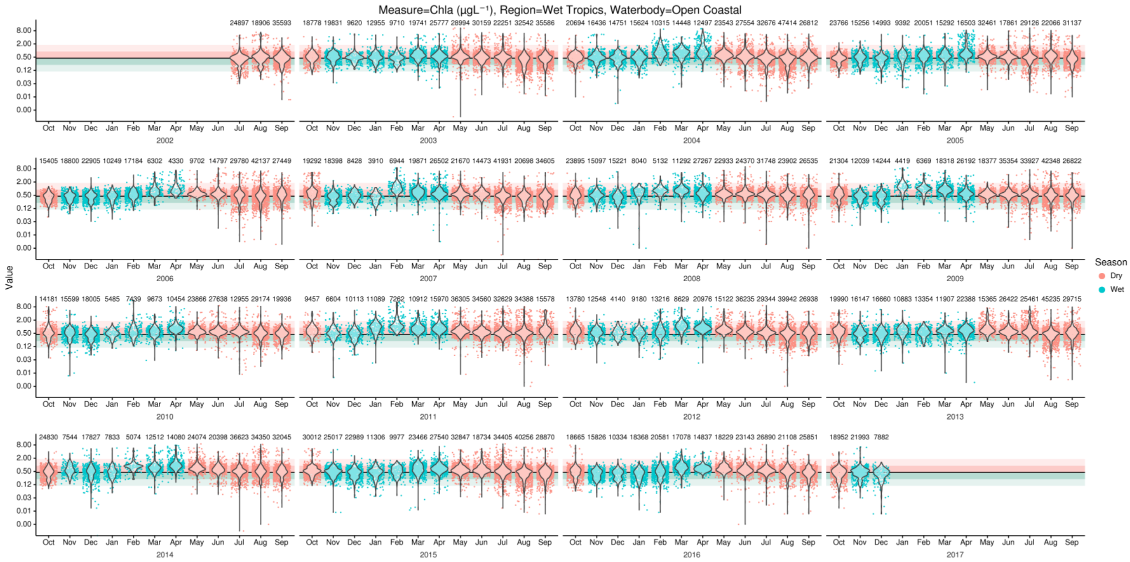


Figure 13: Observed (logarithmic axis with violin plot overlay) Chlorophyll-a data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

a) Satellite



b) eReefs

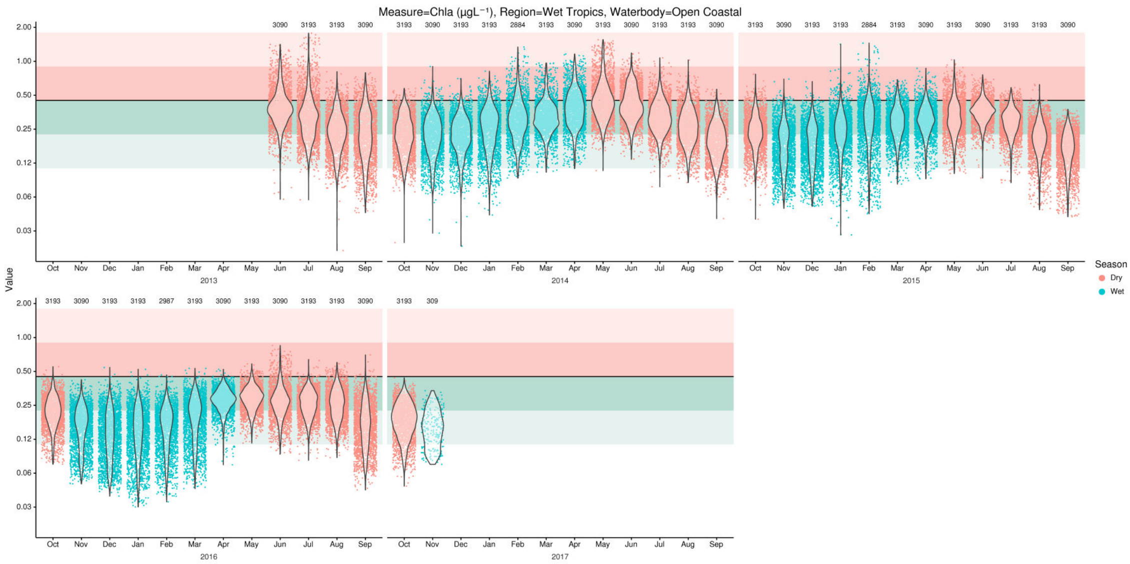
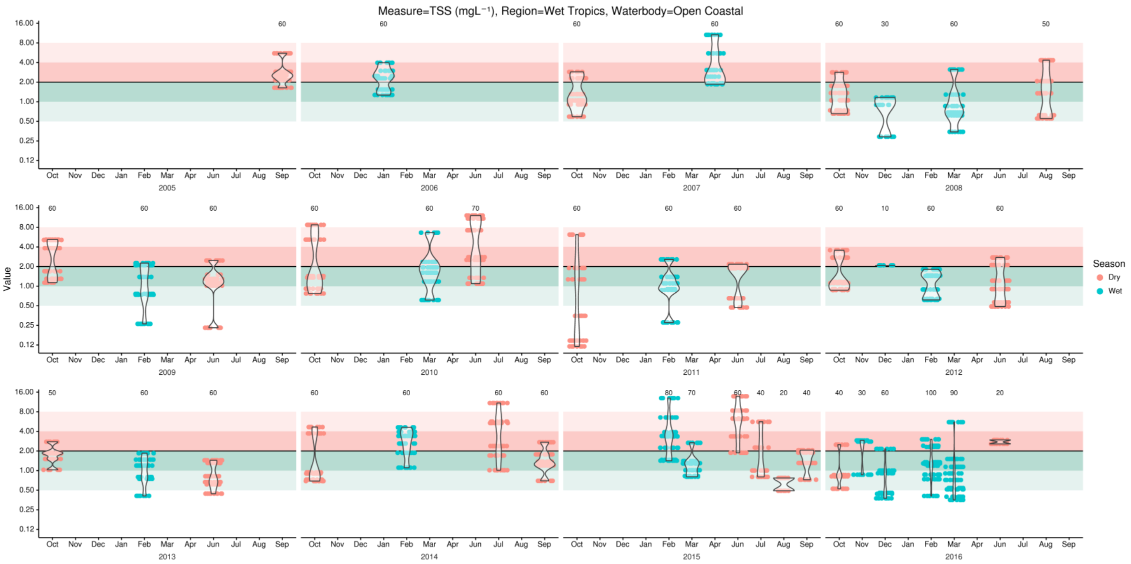


Figure 14: Observed (logarithmic axis with violin plot overlay) Chlorophyll-a data for the Wet Tropics Open Coastal Zone from a) Satellite, b) eReefs. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

a) AIMS insitu



b) AIMS FLNTU

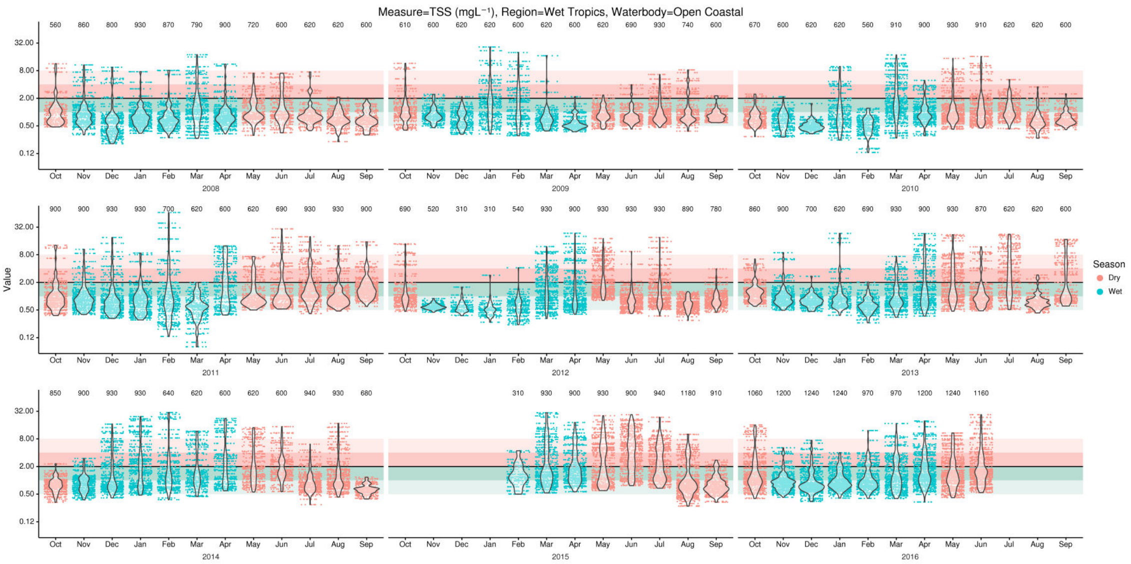
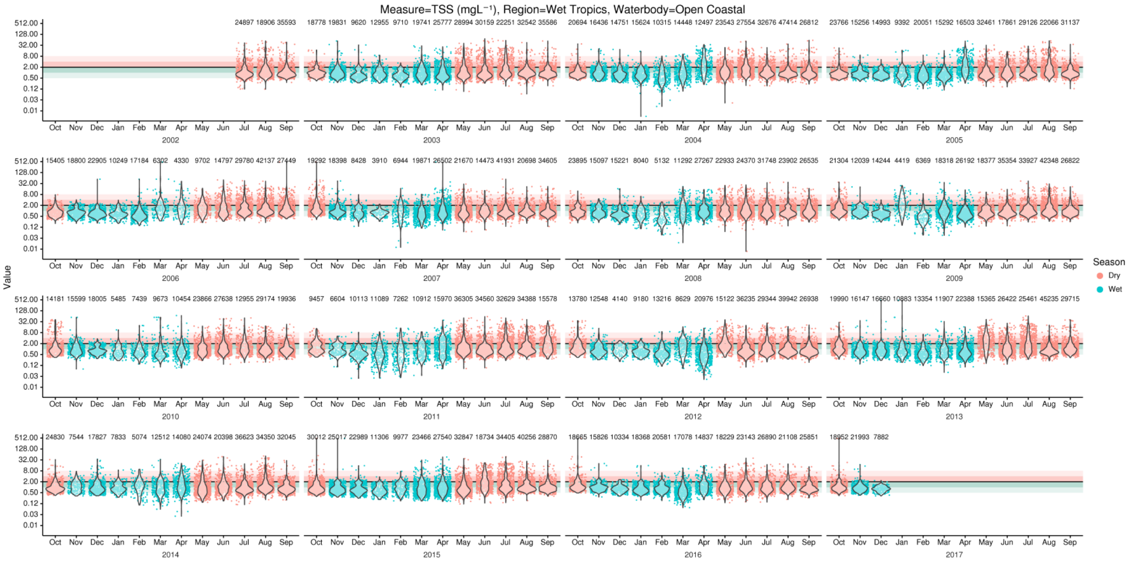


Figure 15: Observed (logarithmic axis with violin plot overlay) TSS data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) AIMS FLNTU. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

a) Satellite



b) eReefs

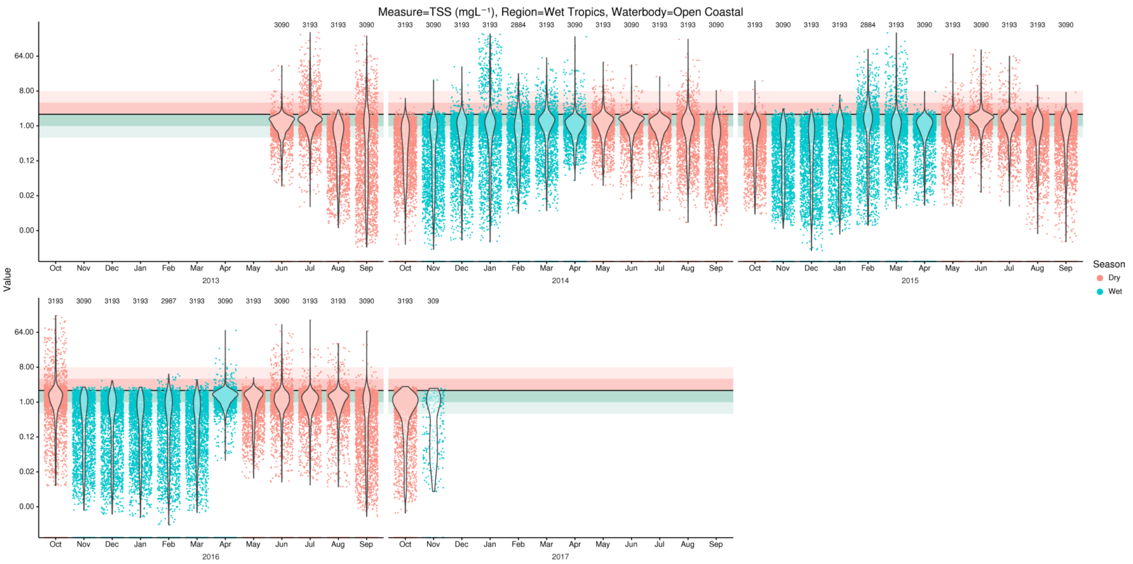
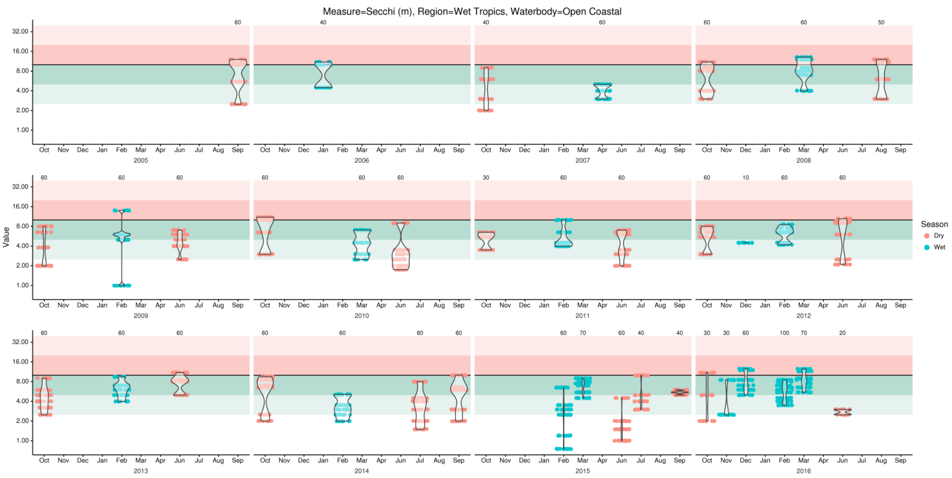
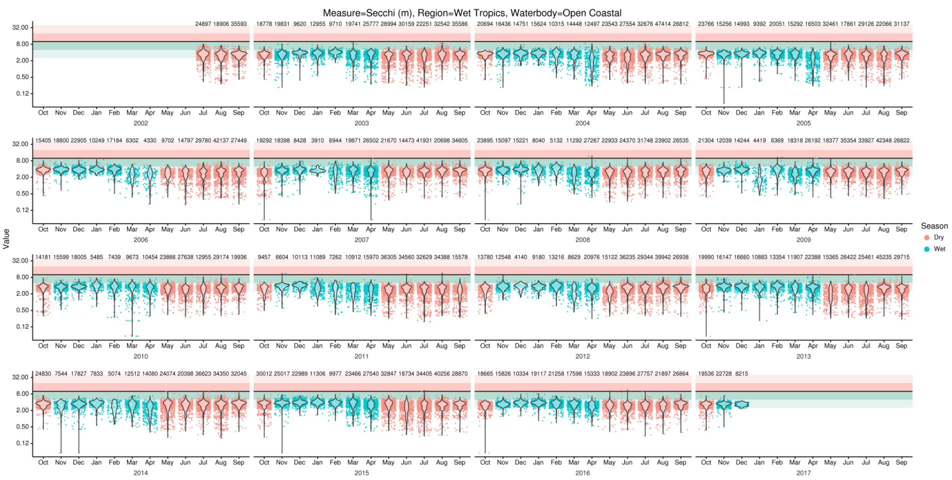


Figure 16: Observed (logarithmic axis with violin plot overlay) TSS data for the Wet Tropics Open Coastal Zone from a) Satellite, b) eReefs. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

a) AIMS insitu



b) Satellite



c) eReefs

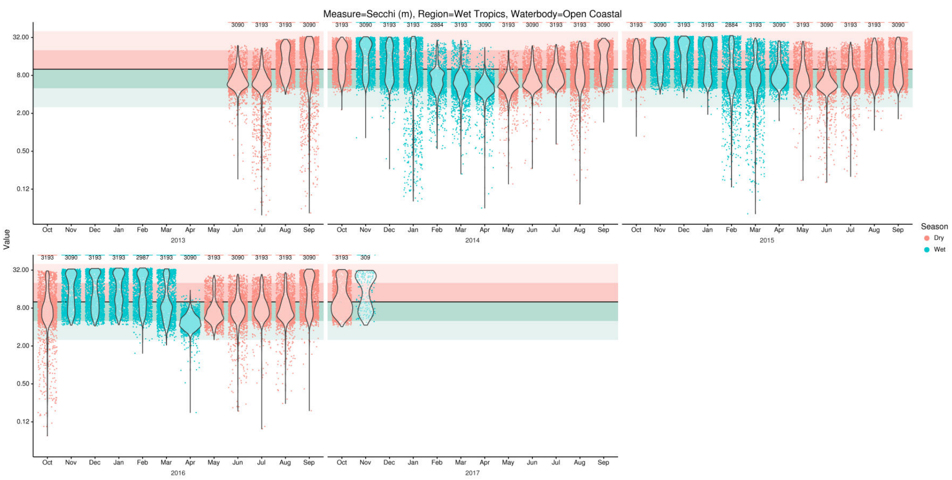
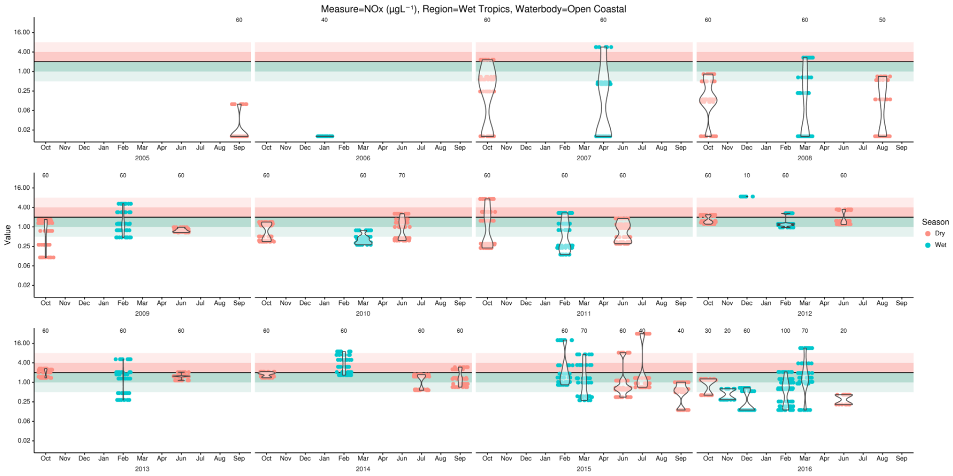
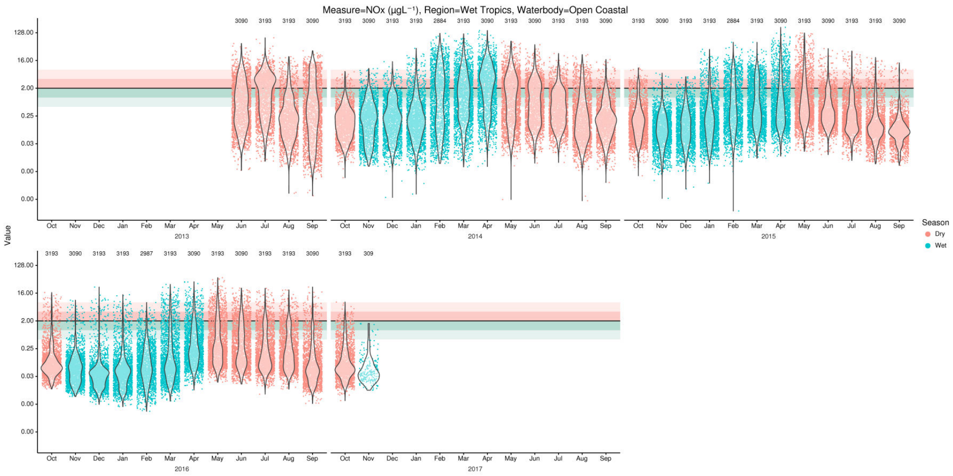


Figure 17: Observed (logarithmic axis with violin plot overlay) Secchi depth data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) Satellite and c) eReefs. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). 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: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

a) AIMS insitu



b) eReefs



c) eReefs926

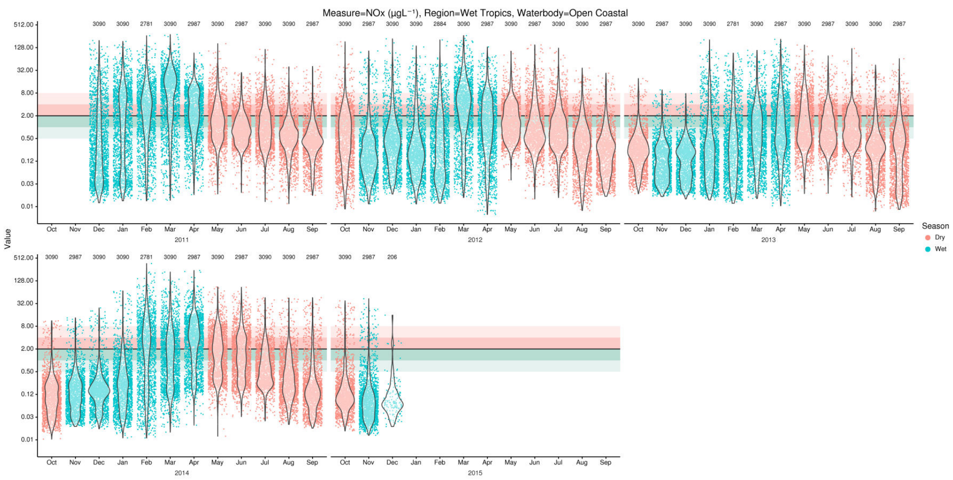


Figure 18: Observed (logarithmic axis with violin plot overlay) Secchi depth data for the Wet Tropics Open Coastal Zone from a) AIMS insitu, b) eReefs c) eReefs926. Observations grouped into months are ordered over time and colored conditional on season as Wet (blue symbols) and Dry (red symbols). Sample sizes represented as numbers above violins and horizontal black dashed line denotes threshold value. Red and green background shading indicates the range (10% shade: x4,/4; 30% shade: x2,/2) above and below threshold respectively.

4.4 Spatial data

Figures 19 – 26 explore the spatio-temporal patterns in observed data from a finer spatial perspective (again focussing on just the Wet Tropics Open Coastal and Dry Tropics Midshelf Zones). Importantly, the colour scales have been mapped to a constant value range for each source for a given Measure. The lower and upper bounds of the constant range is respectively based on twice and half the threshold (see Table A1) value (except for Secchi depth which are half and twice respectively). Half and double the threshold was considered broadly appropriate for the Insitu data and thus should also be broadly appropriate for the other sources since they are intended to be indirect approximations of direct sampling.

These figures also 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[2].

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. 23–25b). Moreover, the spatial patterns of Satellite derived Chlorophyll-a and TSS appear relatively invariant between years (see Figs.19–25b).

The eReefs and eReefs926 do show some variability in spatial and temporal Chlorophyll-a and Secchi depth (see Figs. 19c-d,21c-d,23c-d and 25c-d), yet relatively little for TSS and NOx (at least for Dry Tropics Midshelf). Whilst this apparent lack of variability is largely an artefact of the colour scale mapping, the values of these Measures are constantly substantially below the threshold value and thus invariant on the scale considered appropriate for comparison against the associated thresholds..

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

[2] The remote sensing Satellite data span a temporal range of 2002 through to 2017, although only the range 2010-2016 is displayed

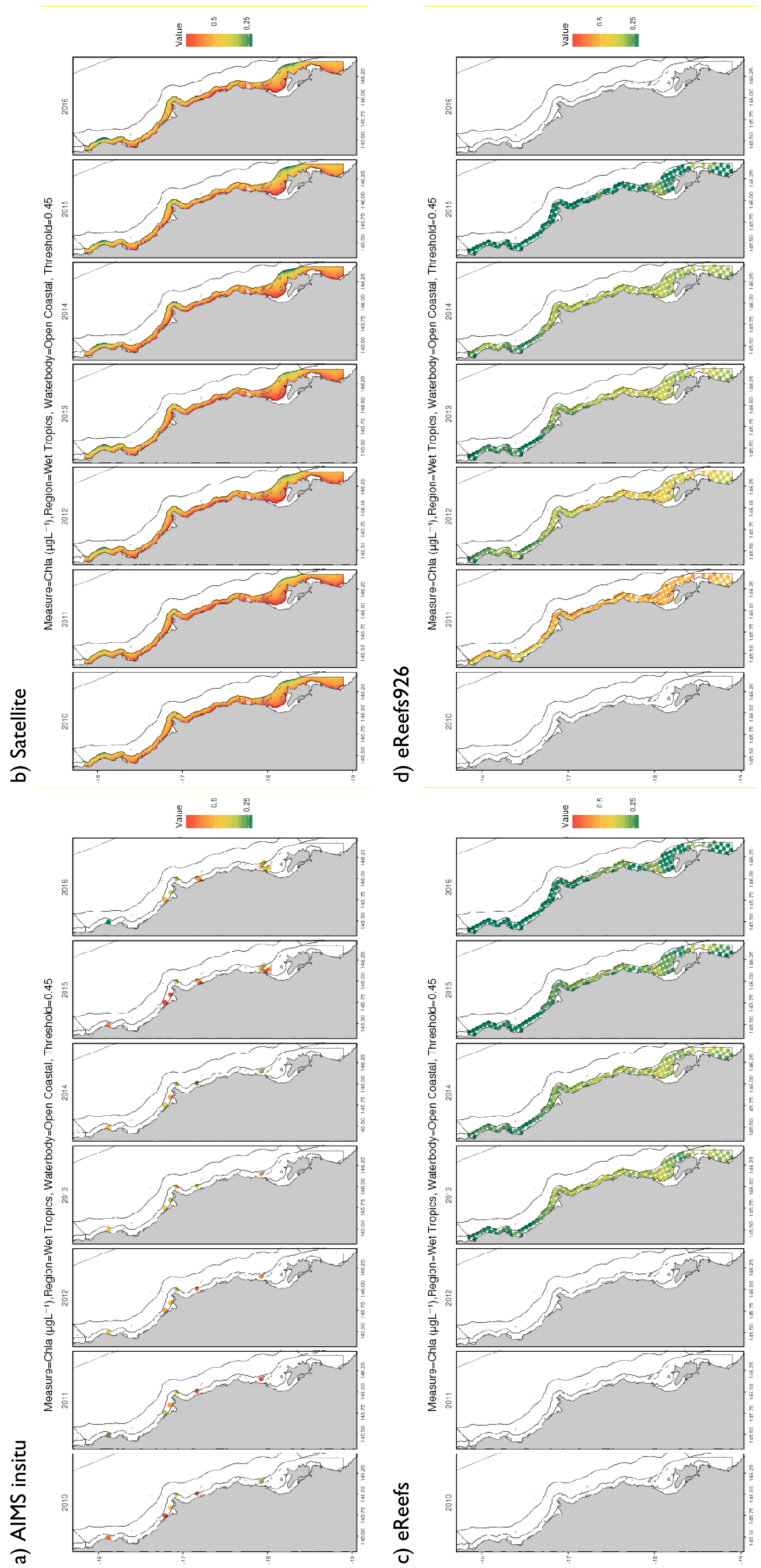


Figure 19: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 Chlorophyll-a (2009–2016) for the Wet Tropics Open Coastal Zone.

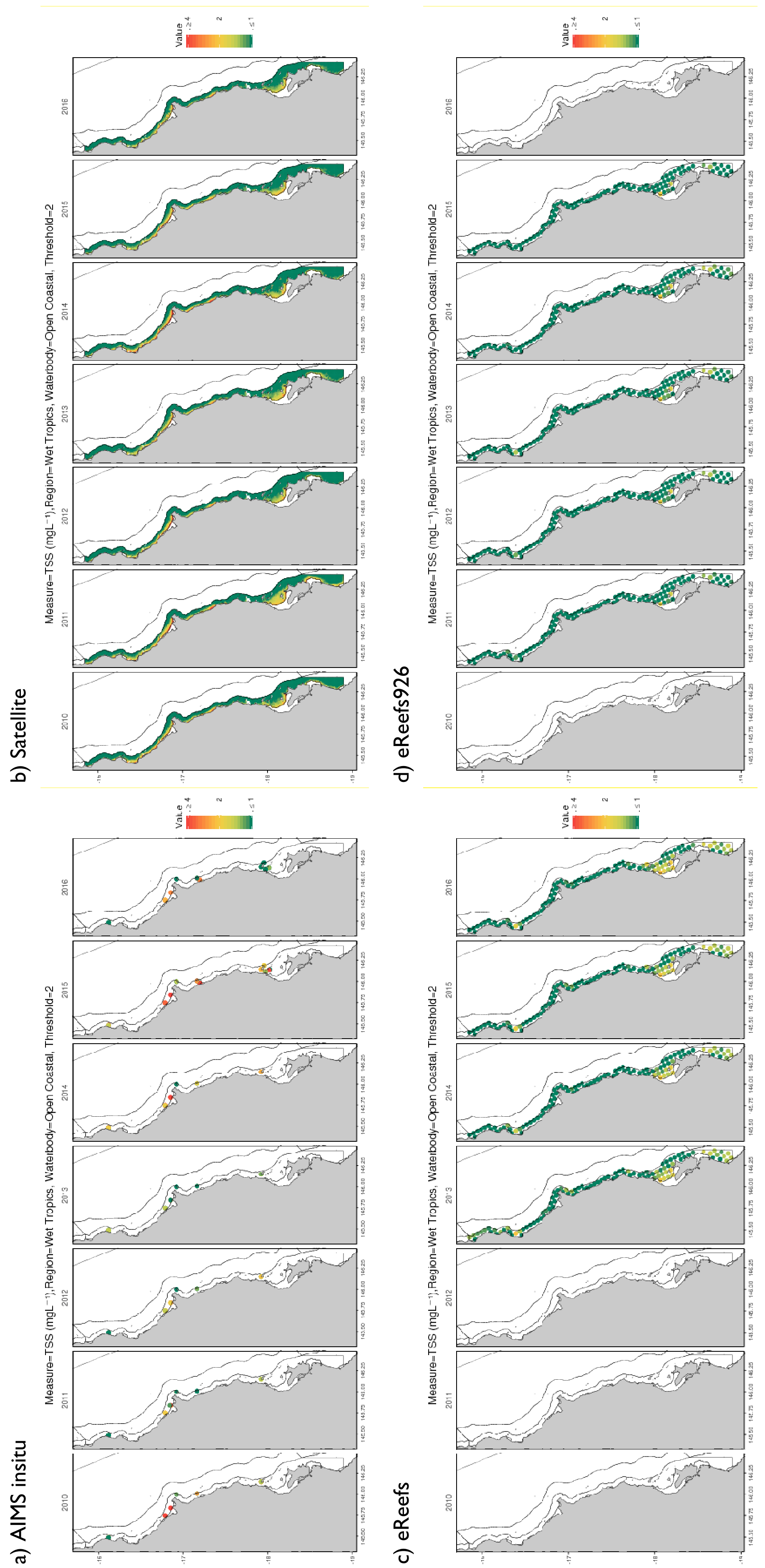


Figure 20: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 TSS (2009–2016) for the Wet Tropics Open Coastal Zone.

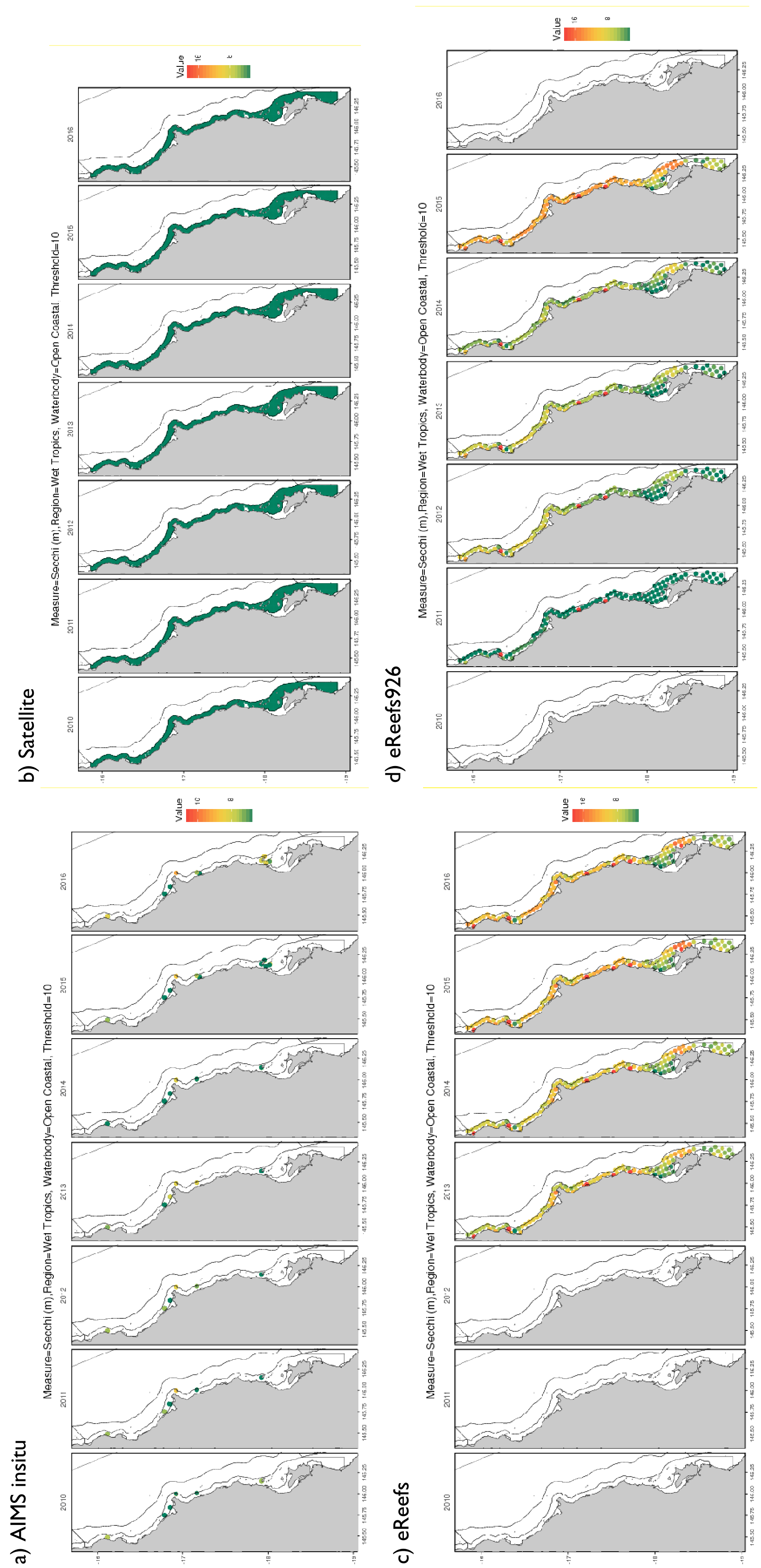


Figure 21: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 Secchi depth (2009–2016) for the Wet Tropics Open Coastal Zone.

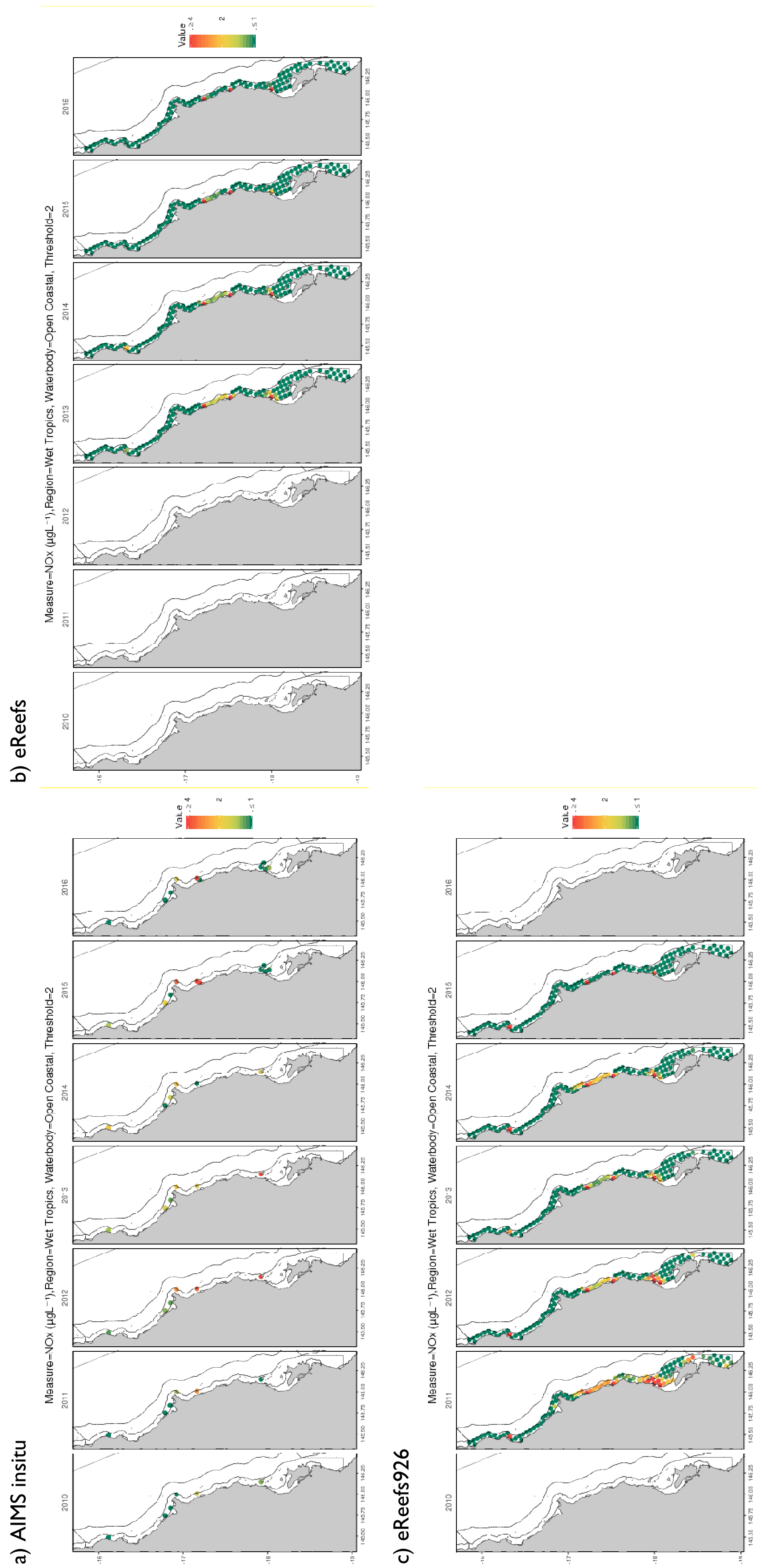
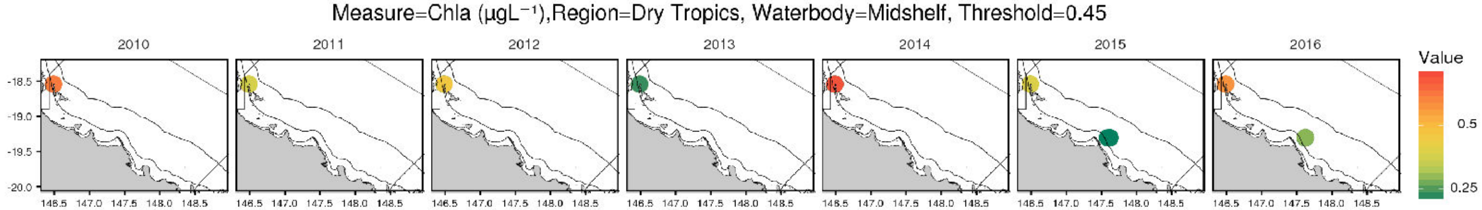
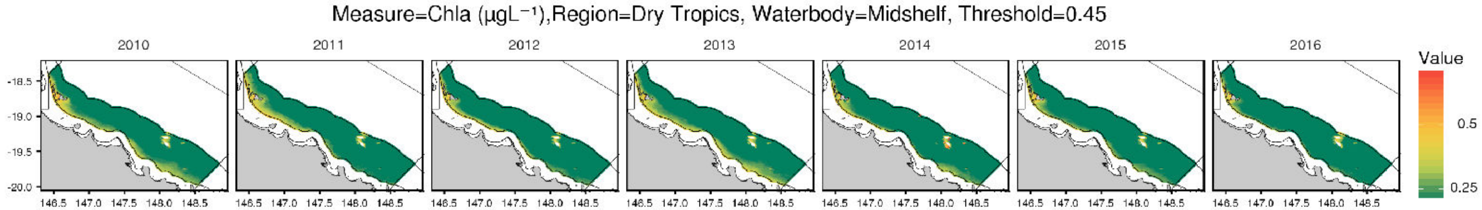


Figure 22: Spatial distribution of observed a) AIMS insitu, b) eReefs and c) eReefs926 NOx (2009–2016) for the Wet Tropics Open Coastal Zone.

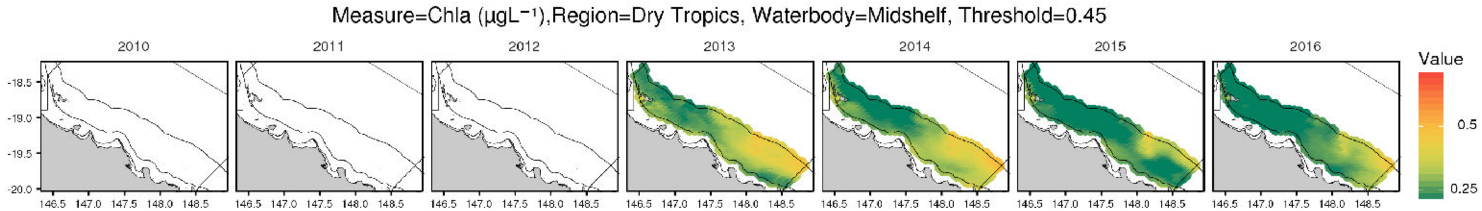
a) AIMS insitu



b) Satellite



c) eReefs



d) eReefs926

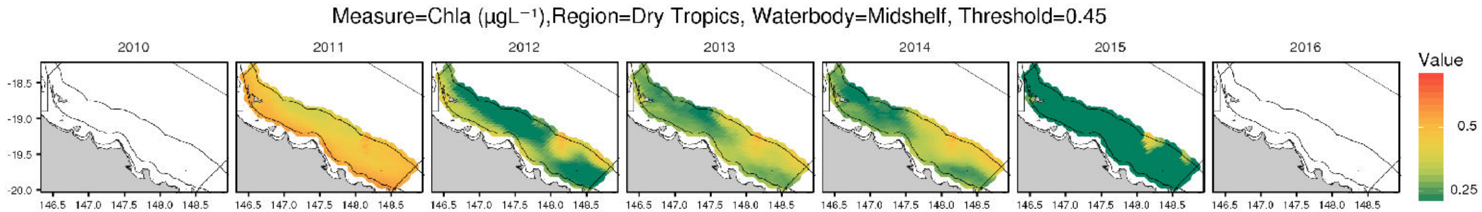


Figure 23: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 Chlorophyll-a (2009–2016) for the Dry Tropics Midshelf Zone.

a) AIMS insitu

b) Satellite

c) eReefs

d) eReefs926

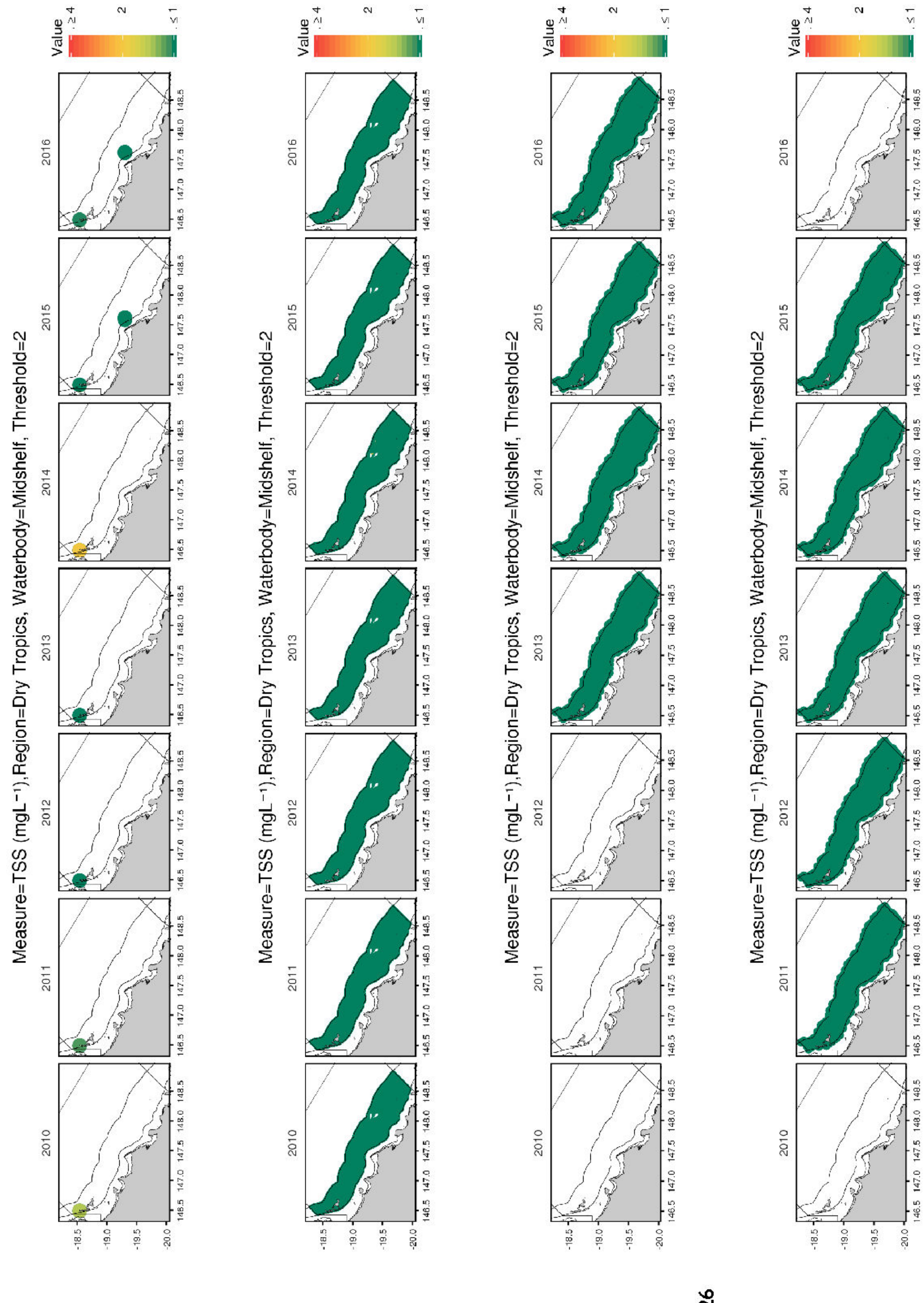
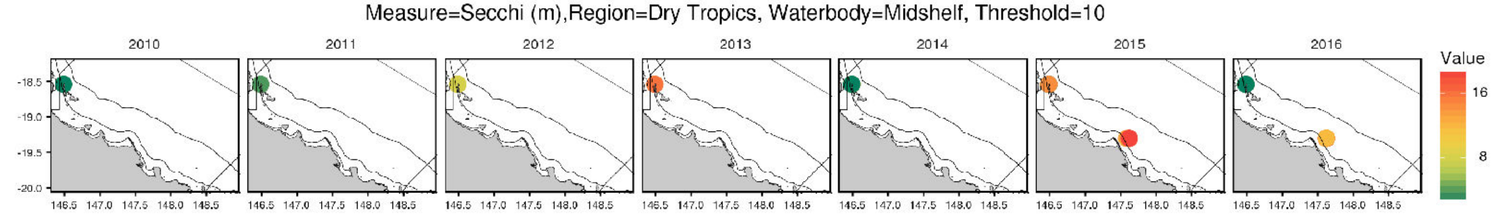
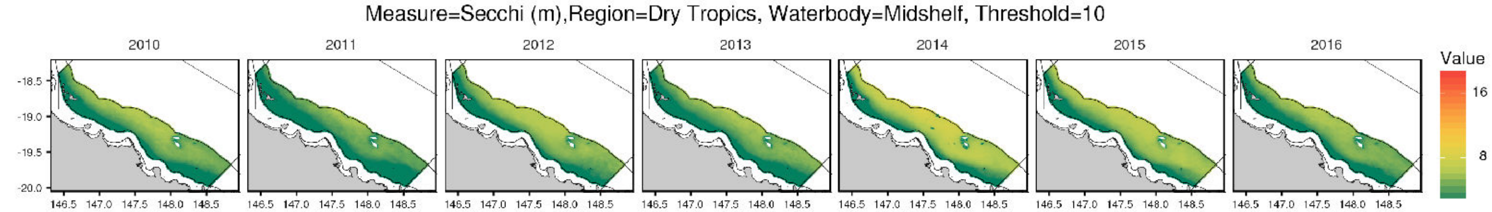


Figure 24: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 TSS (2009–2016) for the Dry Tropics Midshelf Zone.

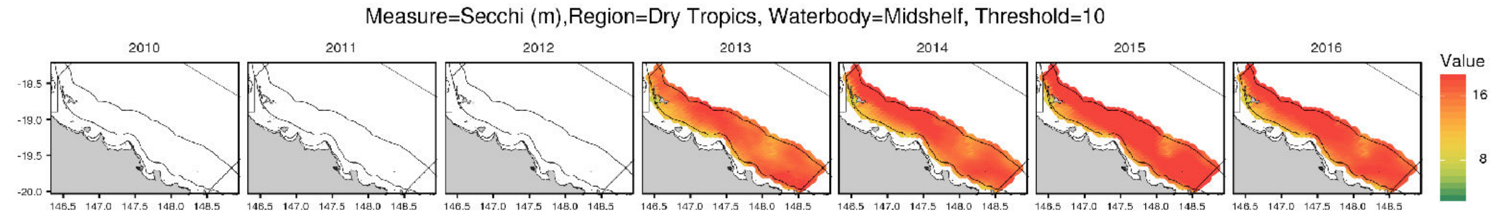
a) AIMS insitu



b) Satellite



c) Reefs



d) eReefs926

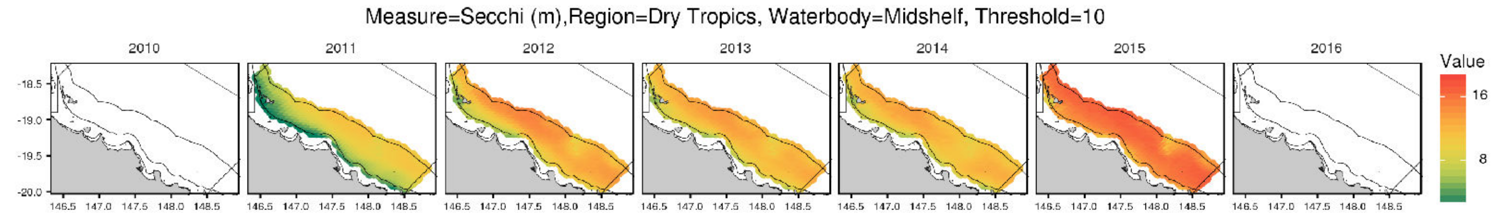


Figure 25: Spatial distribution of observed a) AIMS insitu, b) Satellite, c) eReefs and d) eReefs926 Secchi depth (2009–2016) for the Dry Tropics Midshelf Zone.

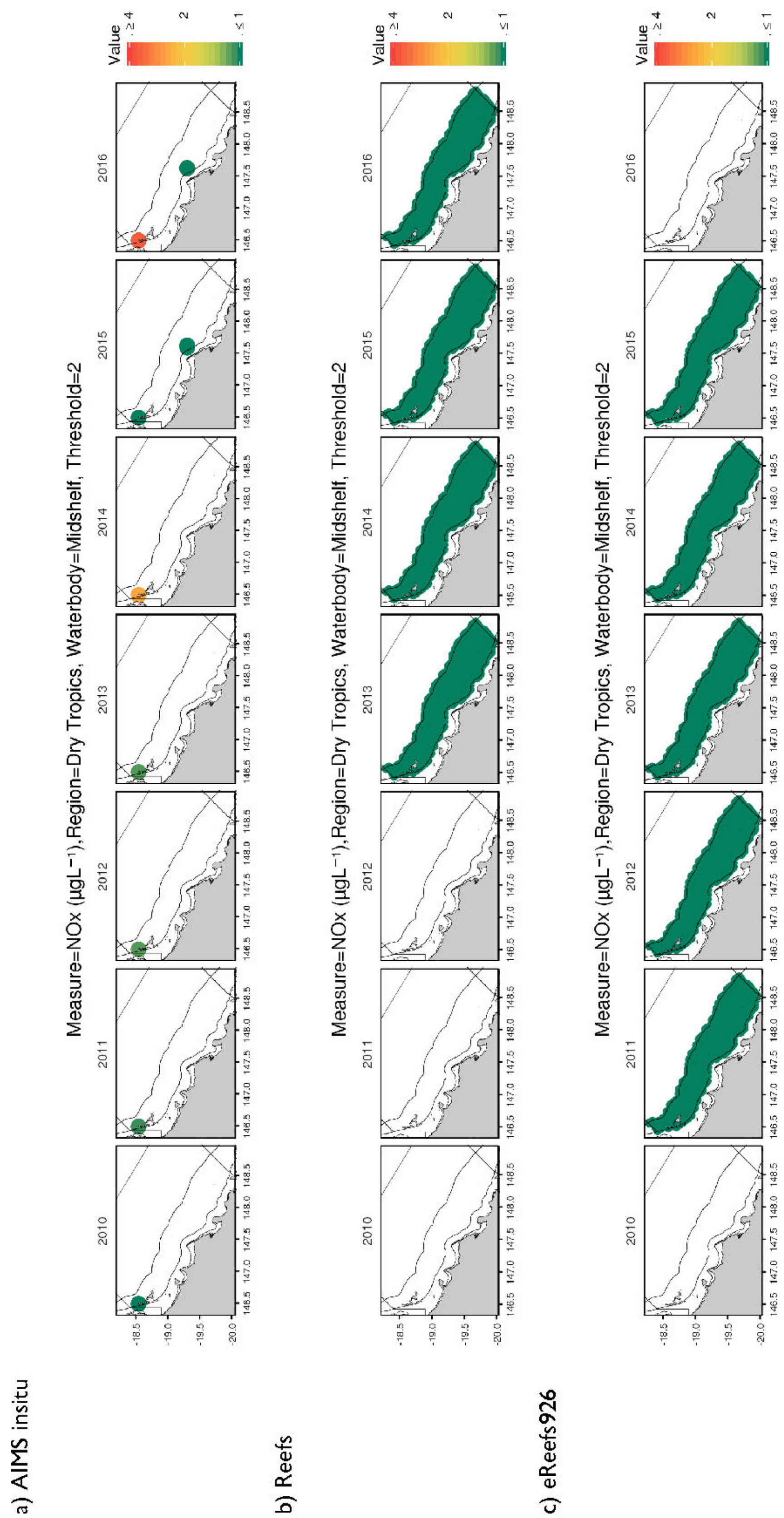


Figure 26: Spatial distribution of observed a) AIMS insitu, b) eReefs and c) eReefs926 NOx (2009–2016) for the Dry Tropics Midshelf Zone.

4.5 Comparison of data sources

Ensuring that the data underpinning the metric calculations are 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 these calculations. 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.

Based on substantial inconsistencies in the magnitude and variation of the observations between sources (AIMS insitu, Satellite and eReefs models), we recommend not to aggregating across the streams of data. Although it might be possible to normalize each source such that they do all have the same basic characteristics prior to aggregation[3], all the various approach to achieve normalization rely on the availability of independent estimates of either data reliability, accuracy or biases present in each source. Unfortunately, such information is not available.

Instead of aggregating the sources together, 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 insitu 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.

The 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 modelled and projected on to a 4km grid on a daily basis without any time series gaps between 2013 and 2016
- the eReefs926 data are modelled and projected on to 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 (and biased) spatio-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[4]. 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,

[3] indeed this is one of the functions of indexing metrics (see section **??**)

[4] Satellite data and eReefs models are of limited value in shallow water

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 27 and 28 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, 9, 10 & 11). Tables 9, 10 & 11 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[5] 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 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. 29 – 32 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

[5] 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.

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.

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

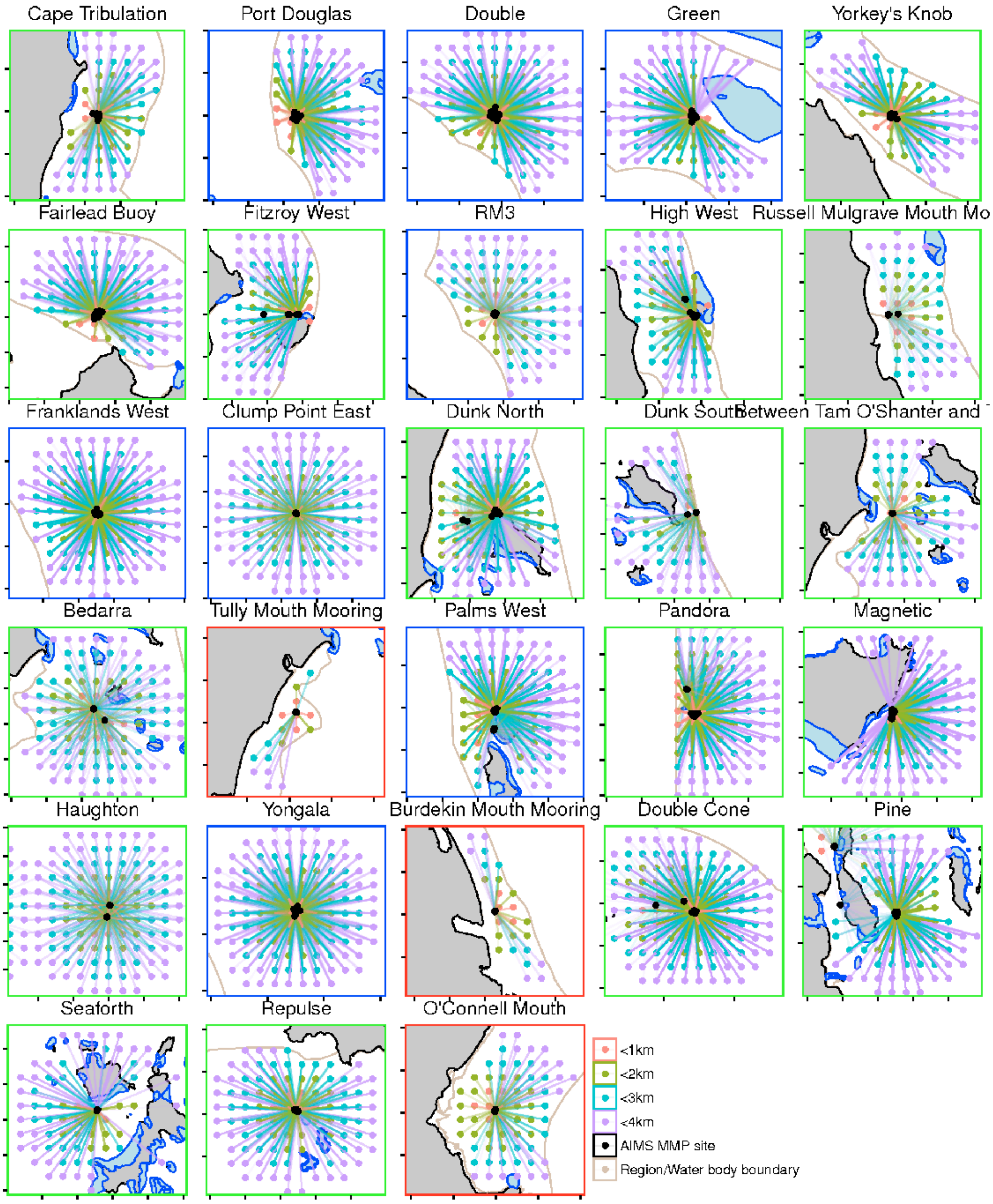


Figure 27: Location of Satellite cells within 5km of AIMS niskin samples. Panel borders represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

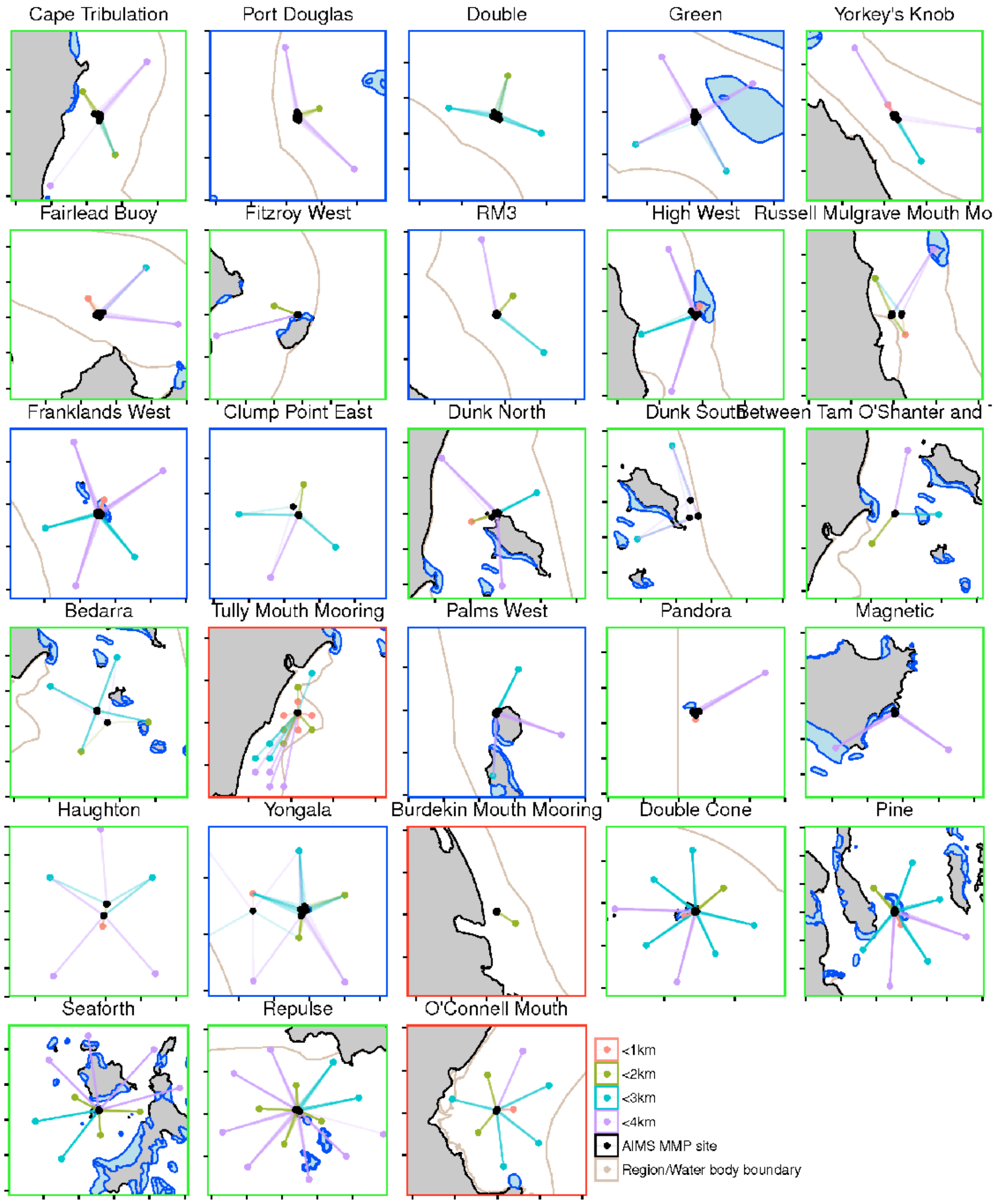


Figure 28: Location of eReefs cells within 5km of AIMS niskin samples. Panel borders represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

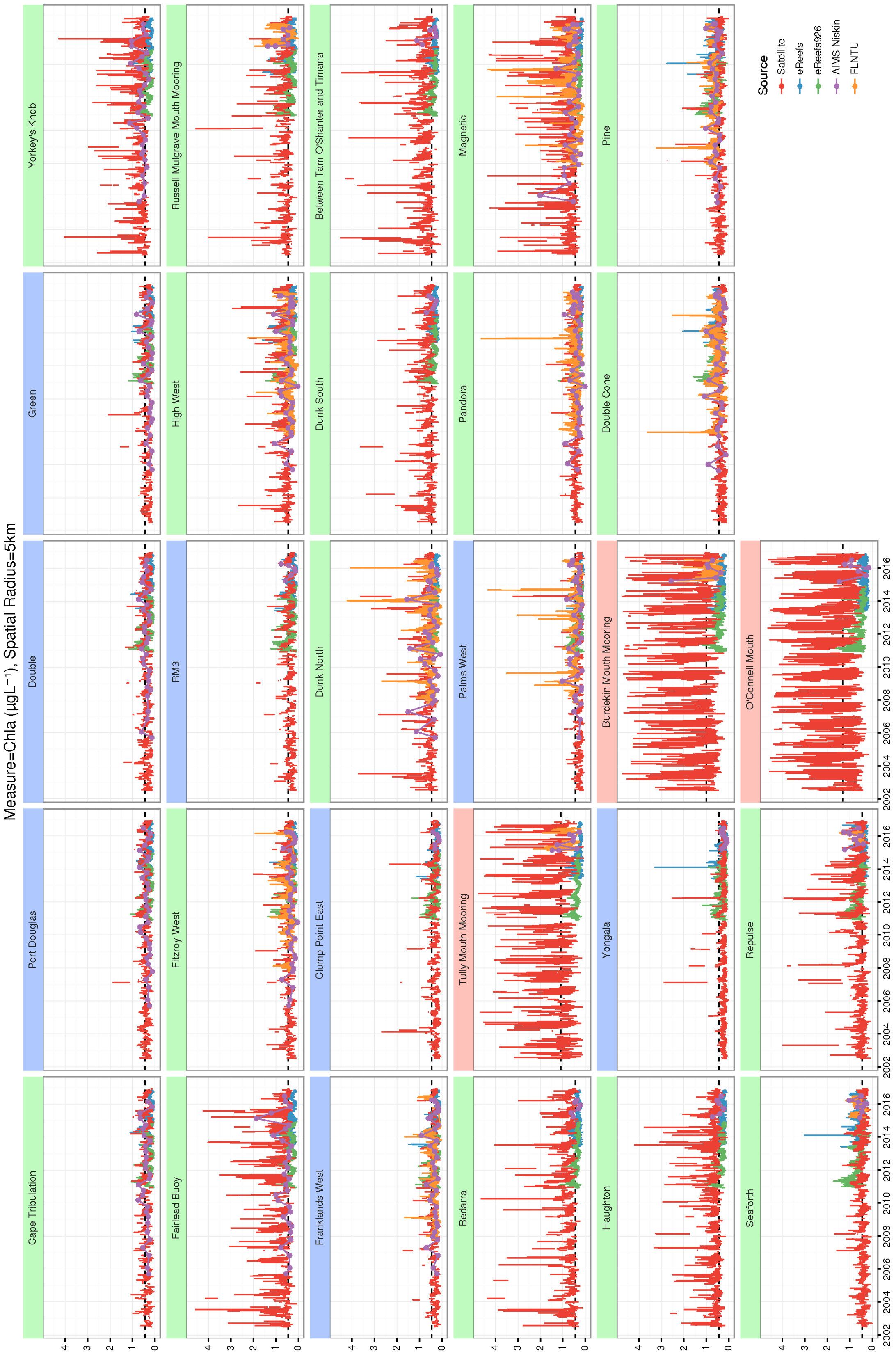


Figure 29: Temporal patterns in Chlorophyll-a within 5km of each AIMS MMP sampling site for eReefs, Satellite and AIMS insitu and FLNTU logger sources. Horizontal dashed line represents the guideline value. Title backgrounds represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

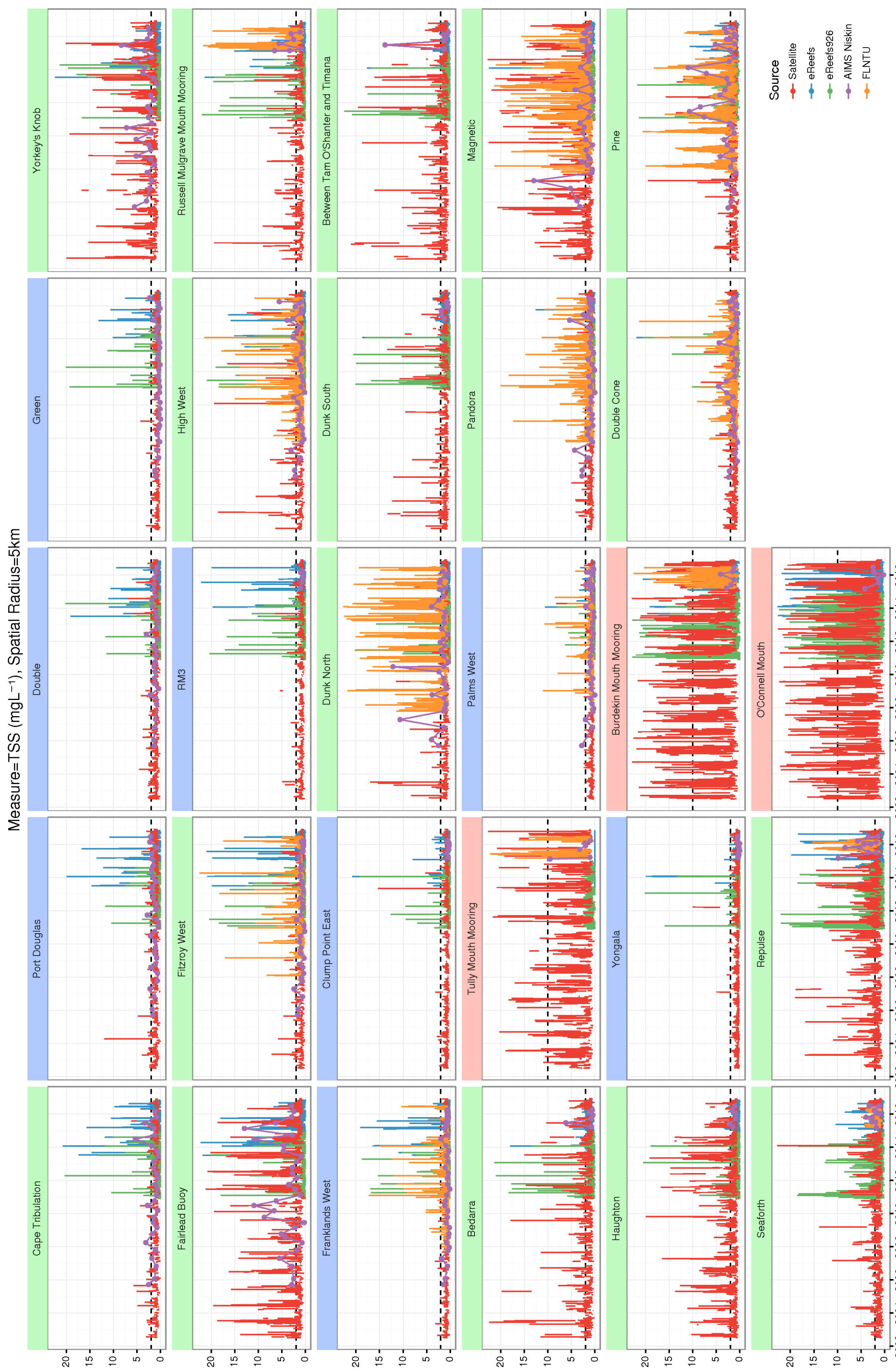


Figure 30: Temporal patterns in TSS within 5km of each AIMS MMP sampling site for eReefs, Satellite and AIMS insitu and FLNTU logger sources. Horizontal dashed line represents the guideline value. Title backgrounds represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

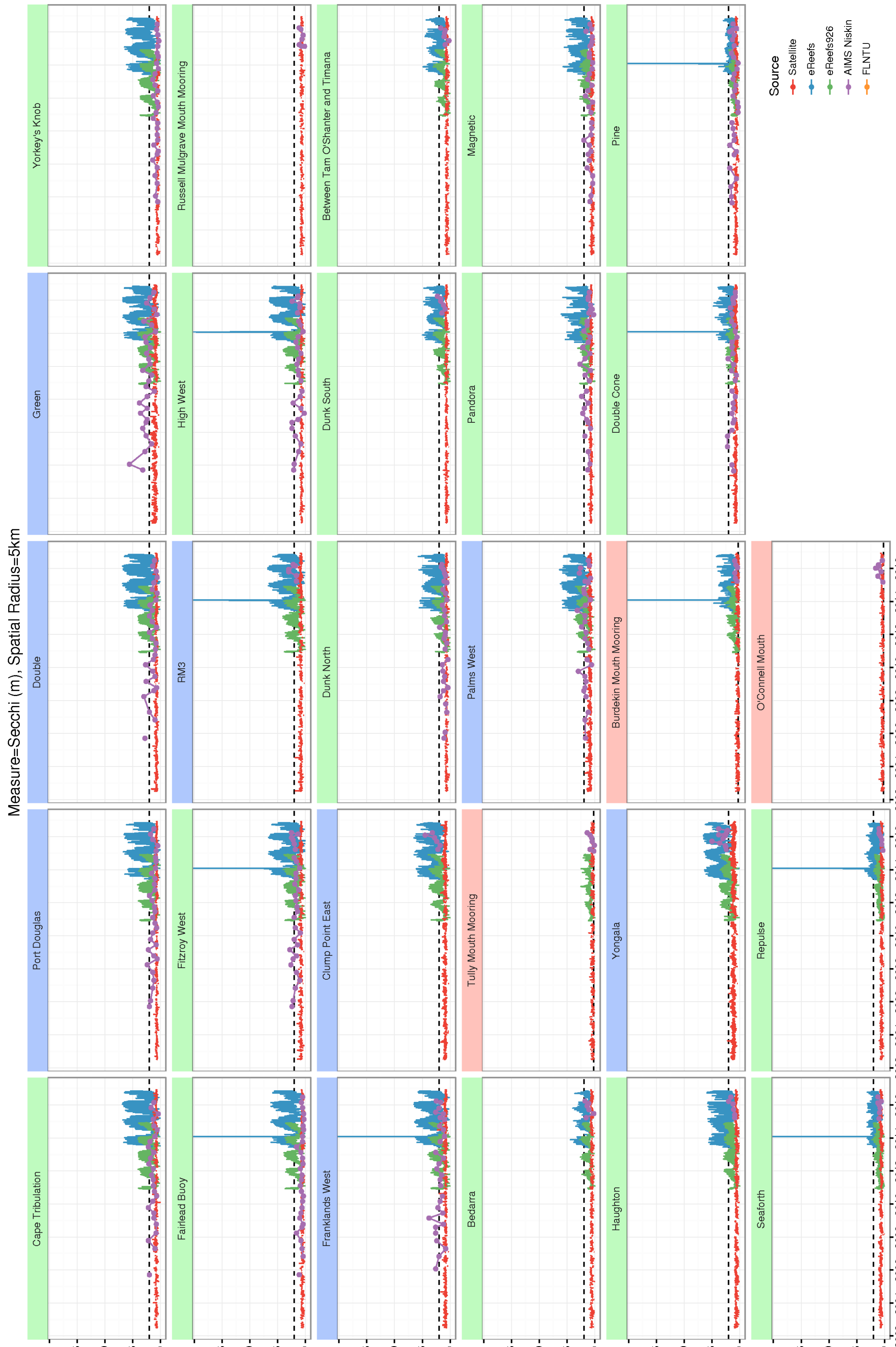


Figure 31: Temporal patterns in Secchi Depth within 5km of each AIMS MMP sampling site for eReefs, Satellite and AIMS insitu and FLNTU logger sources. Horizontal dashed line represents the guideline value. Title backgrounds represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).

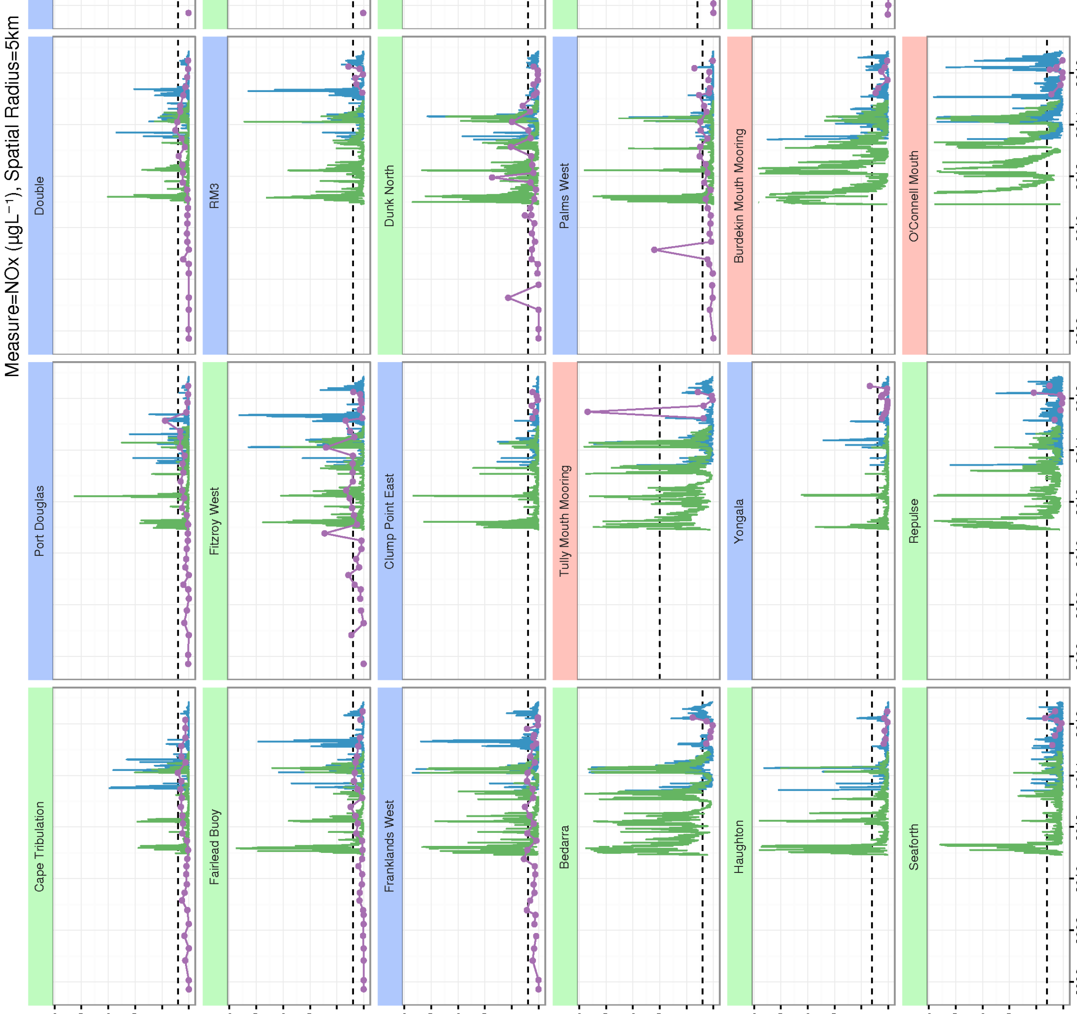


Figure 32: Temporal patterns in NOx within 5km of each AIMS MMP sampling site for eReefs, Satellite and AIMS insitu and FLNTU logger sources. Horizontal dashed line represents the guideline value. Title backgrounds represent water bodies (Red: Enclosed Coastal, Green: Open Coastal, Blue: Midshelf).