Reef Water Quality Protection Plan

The Great Barrier Reef catchments wetland monitoring program

Analysis methods

Australian Government

Queensland Government

Great Barrier Reef catchments wetland monitoring program

Analysis methods

Environmental Monitoring Assessment Science

September, 2017

Prepared by

Jan Tilden
Environmental Monitoring Assessment Sciences
Science Delivery Division
Department of Science, Information Technology and Innovation
PO Box 5078
Brisbane QLD 4001

Disclaimer

This document has been prepared with all due diligence and care, based on the best available information at the time of publication. The department holds no responsibility for any errors or omissions within this document. Any decisions made by other parties based on this document are solely the responsibility of those parties. Information contained in this document is from a number of sources and, as such, does not necessarily represent government or departmental policy.

If you need to access this document in a language other than English, please call the Translating and Interpreting Service (TIS National) on 131 450 and ask them to telephone Library Services on +61 7 3170 5725

Citation

Tilden, J. and Vandergragt, M. 2017 Great Barrier Reef Catchments Wetland Monitoring Program: analysis methods. Wetland Sciences, Queensland Department of Science, Information Technology and Innovation, Brisbane.

Acknowledgements

This report has been prepared by the Department of Science, Information Technology and Innovation. Acknowledgement is made of Dr Wayne Robinson, Institute for Land, Water and Society, Charles Sturt University for independent review of the report and David Moffatt and Dr Caitlin Johns, Queensland Department of Science, Information Technology and Innovation for their contribution to the review process.

Main cover image: Geoff Borschmann

September, 2017

Contents

List of tables

List of figures

1 Introduction

This is a technical report on the methods for analysing data gathered for the Great Barrier Reef catchments wetland monitoring program. A linear statistical model of individual wetland assessment scores is adopted to analyse the components of variance affecting the score distribution. Analysis focuses on methods to detect signals of status, change and trends in anthropogenic disturbance, given the highly variable natural condition of freshwater wetlands.

Table 1 presents a brief timeline and projection of the program up to 2019, including the development of the tool used for wetland assessments.

Table 1 Development timeline of the GBR catchments wetland monitoring program and the WFAT–M instrument

Year	Great Barrier Reef catchments wetland monitoring program	Wetland Field Assessment Tool for Monitoring (WFAT–M)
2013	Conceptual framework and conceptual modelling for GBR catchments wetland monitoring program	Testing the feasibility of using the WFAT–M in the GBR wetland monitoring context, redeveloping and refining the instrument
2014	Pilot study informs monitoring design including: sample size, allocation of sampling effort in space (sub-population and sampling method), allocation of sampling effort in time (panel design[1])	Pilot study validates and refines WFAT–M as an instrument for rapid assessment and long-term monitoring of pressure on wetland values and state of wetland values
2015	Publication of report *The Great Barrier Reef catchments wetland monitoring pilot study: assessment methods and monitoring design*	
2015	Sample year 1a[2], 17 wetlands assessed	Information gathered for further refinement of WFAT–M methods
2016 **(this report written)**	Sample year 1b, 23 wetlands assessed, analysis proposal drafted (this document), baseline data analysed based on N = 40 wetlands (sample year 1a + sample year 1b)	Information gathered for further refinement of WFAT–M methods, inter-operator reliability (repeatability) testing initiated
2017	Baseline data reported. Sample year 2, 40 wetlands assessed. Data analysed, consisting of panel 1, surveys 1 and 2 (20 wetlands assessed *twice*) and panels 2 and 3, survey 1 (40 wetlands assessed *once*). Total n of wetlands assessed *any* number of times = 60. Review program to determine if desired level of precision to report on status, change and trend will be achieved.	Further validation and calibration of the WFAT–M Review and refinement of WFAT–M. Improve the precision of the instrument by addressing sources of extraneous variance.
2018	Sample year 3, 40 wetlands assessed. This is the first opportunity to test for change, as 40 wetlands will now have been assessed more than once.	Version 2 of WFAT–M completed.
2019	Sample year 4, 40 wetlands assessed. 60 wetlands will now have been assessed more than once.	

[1] A 'panel design' for ecological monitoring is one in which a number of sub-samples, referred to as panels, each has a different repeat pattern through time.

[2] While the monitoring design aimed to assess 40 wetlands per year, assessment of the first 40 wetlands extended across two years due to limited capacity in the early stage of the project.

1.1 Monitoring design

The Great Barrier Reef catchments wetland monitoring pilot study: assessment methods and monitoring design (Tilden et al, 2015) was published as a technical report under the Reef Water Quality Protection Plan (Reef Plan). This pilot study report assessed the performance of the Wetland Field Assessment Tool for Monitoring (WFAT–M) in discriminating across the spectrum of wetland disturbance from pristine to highly disturbed and recommended refinements to the WFAT–M. The pilot study also made recommendations about the monitoring program design for the Great Barrier Reef (GBR) catchments wetland monitoring program.

These recommendations were accepted by the Reef Plan Independent Science Panel and adopted for the subsequent phase of the GBR catchments wetland monitoring program. That is, gathering baseline data on (a) pressure on wetland environmental values, particularly as related to land use and (b) the state of wetland environmental values. To gather those data, a random, spatially balanced sample of 100 natural freshwater floodplain wetlands from dense aggregations within the GBR catchments (the sub-population) was selected.

Justifications for the monitoring design, sample size, sampling methods, sub-population and sampling effort in time and space are in the 2015 technical report (Tilden et al, 2015).

1.2 Analysis methods

The current report focuses on methods for analysing the data gathered using the chosen monitoring design, an augmented serially alternating panel design with an eight year repeat pattern. The primary assessment questions to be answered through analysis of this data are:

- What is the current level of *pressure* on the defined wetland environmental values[3] of GBR natural freshwater floodplain wetlands in dense aggregations? (status)
- What is the current *state* of those wetland environmental values in GBR natural freshwater floodplain wetlands in dense aggregations? (status)
- What is the *impact* of pressure on wetland state? That is, are the wetland environmental values of the sub-population improving, deteriorating or remaining the same over time? (trend)

Here we address whether and how those questions will be answered, using data gathered with the WFAT–M and applying the analytical tools described here.

Matters to be considered include:

- What statistics and statistical tests will be used for reporting on status, change and trend?
- What is the power of those statistical tests to detect changes or trends in wetland status, given the chosen experimental design?
- What assumptions are being made about the ability of WFAT–M to assess status and detect change and trend in GBR wetlands and how will these assumptions be tested?

[3] Wetland environmental values assessed by the WFAT–M are defined in Section 2.1.

2 Background

The following discussion gives background information relevant to the various analyses planned for each year in the first assessment cycle.

The Reef Plan 2013 wetland target is that:

> There [will be] no net loss of extent, and *an improvement in the ecological processes and environmental values, of natural wetlands*.

Wetland extent is monitored under another program within the Paddock to Reef integrated monitoring, modelling and reporting program (Paddock to Reef Program, 2015). The GBR catchments wetland monitoring program addresses the second part of the target (italicised). Because it is not feasible to monitor all 5327 wetlands in the sub-population being considered (described in Tilden et al, 2015), the Reef Plan target cannot be tested by conducting a census. Inference must be made about the status and trend of wetland environmental values based on data gathered from a spatially balanced random sample of wetlands from this sub-population. Further, since the aim of the Reef Plan is to protect the Great Barrier Reef through best management, the baseline wetland status variable of interest is *the level of disturbance to wetlands under different land uses (anthropogenic disturbance).* The targeted improvement to ecological processes and environmental values will be achieved through improved land management. Consequently, the WFAT–M, a rapid assessment instrument to assess the level of land use-associated disturbance to natural freshwater wetlands, has been developed to gather data for the GBR wetland monitoring program.

2.1 Statistical assumptions about WFAT–M scores

WFAT–M scores for individual wetlands comprise two separate indices[4], **pressure** and **state**; the first represents land use **pressure** on natural wetland environmental values (WEVs) while the other scores the **state** of those values (equivalent to wetland condition). For each index, the following four WEVs are evaluated, with multiple indicators assessed per WEV:

- WEV 1 The biological health and diversity of the wetland's ecosystems
- WEV 2 The wetland's natural physical state and integrity
- WEV 3 The wetland's natural hydrological cycle
- WEV 4 The natural interaction of the wetland with other ecosystems, including other wetlands.

The process of generating scores for each wetland is outlined below and illustrated in Table 2:

- Individual WFAT–M indicators are assessed on ordinal scales with scores generally ranging from one to five.
- For each of the four WEVs, pseudo-interval aggregated scores, from 1 to 5, are produced.
- Overall (pseudo-interval) scores per wetland are generated by aggregating the WEV scores to yield single index values, one for pressure and one for state.
- For reporting purposes, all scores are converted to a scale from A to E.
- For indicator and WEV scores, five score classes are reported.
- For the overall pressure and state scores, a thirteen point scale (A, A–, B+, B, B–, C+….E+, E) is used.

Norman (2010) summarises work supporting the use of parametric statistics on data of this type.

[4] Index is defined as a compound measure that aggregates multiple indicators

In line with the above, the main data to be used for assessing status and trend of GBR wetland environmental values are the WFAT–M overall wetland pressure and state scores, along with the means and variances of these scores across the GBR catchments. To assess trend, scores from the same wetlands will be compared across multiple monitoring periods (notionally of one year) using analysis of variance (see section 3, especially 3.5.1 for more detail). Individual WEV scores and their means and variances will also be analysed. As well, indicative non-parametric descriptive statistics, such as medians and interquartile ranges, will be reported.

Table 2 WFAT–M scoring and reporting scales

Indicator scores	1			2			3			4			5
Reported indicator scores	A			B			C			D			E
Aggregated sub-index scores (4 x WEVs*)	≤1.8			>1.8 to ≤2.6			>2.6 to ≤3.4			>3.4 to ≤4.2			>4.2
Reported sub-index scores	A			B			C			D			E
Aggregated overall pressure and state scores	≤1.267	>1.267 to ≤1.534	>1.534 to ≤1.8	>1.8 to ≤2.067	>2.067 to ≤2.334	>2.334 to ≤2.6	>2.6 to ≤2.867	>2.867 to ≤3.134	>3.134 to ≤3.4	>3.4 to ≤3.667	>3.667 to ≤3.934	>3.934 to ≤4.2	>4.2
Reported overall scores	A	A–	B+	B	B–	C+	C	C–	D+	D	D–	E+	E

*Wetland environmental values

2.2 What's in a WFAT–M score?

A random, spatially balanced sample of 100 GBR natural[5] freshwater floodplain wetlands from dense aggregations has been chosen for monitoring the status and trend of environmental values of these wetlands across the whole GBR catchment. During the dry season of each year, 40 wetlands are assessed using an augmented, serially alternating panel design (see Table 6 and also Tilden et al, 2015). The average WFAT–M score of each sub-sample of 40 wetlands in each year is an estimate of the average WFAT–M score of the whole GBR sub-population of wetlands for that year. Hence it is an estimator of the status of sub-population wetlands, assuming the WFAT–M validly assesses wetland status and that the sampling design is representative of GBR freshwater, floodplain wetlands.

The distribution of the sample wetland scores around this mean in any year includes several important sources of variability that affect the sensitivity of the WFAT–M instrument and the associated monitoring design to assess *status* of wetland environmental values. Additional components of variance related to the natural seasonal and annual variability of wetlands influence the ability to detect *trend*.

[5] In this context, the word "natural" is a historical term, referring to wetlands that are naturally occurring, as opposed to fully constructed. A naturally occurring wetland that has been bunded is a natural wetland under this definition, whereas a farm dam constructed at a high point in the landscape (for gravity feed) with no surrounding natural wetland vegetation would not be considered natural.

Following Larsen et al (1995), Urquhart et al (1998) and others, a linear statistical model of individual wetland WFAT–M scores has been adopted to analyse the components of variance affecting in the distribution of WFAT–M scores:

$Y_{ij} = \mu + W_i + T_j + E_{ij}$ (equation 1)

Where

i refers to wetlands

j refers to periods

Y_{ij} represents the predicted WFAT–M score of wetland *i* and period *j*

μ is the overall mean of all WFAT–M scores across the wetland sample during the time interval of interest

W_i is the mean difference of a particular wetland from the overall mean

T_j is the difference between the GBR-wide mean wetland WFAT–M score during any year and the overall mean

E_{ij} is the residual term, including all other unaccounted for variance. This is further discussed below.

The total variance in WFAT–M scores for a random sample of wetlands may be characterised as *population* variance plus *extraneous* variance (after Larsen et al, 2001 and Larsen and Christie, 1991). Population variance is spatial variability, describing the measured differences among wetlands in a geographically defined population or sub-population of wetlands during the index period. In the case of the GBR wetland monitoring program, the index period is the Queensland dry season, notionally, from the beginning of April to the end of September. Extraneous variance comprises all components of variance that are not a part of population variance but that reduce the precision of estimates of status and trends. Various options may be available for estimating or partialling out the different components of extraneous variance to improve the precision of status estimation and trend detection for GBR wetlands.

2.2.1 Population variance

Freshwater floodplain wetlands across the Great Barrier Reef Catchments embody a considerable range of natural variability in geomorphology, hydrology, habitat types and climatic influences. This is in addition to variability in the level of anthropogenic disturbance to wetland environmental values. It is anthropogenic disturbance, amenable to management actions on the part of land holders, that the WFAT–M seeks to capture. In order to do so, the instrument has been designed to be 'blind' to natural population variability as far as possible. To the extent that natural population variability persists in the data, it represents noise in the attempt to detect an anthropogenic disturbance signal.

Despite efforts to remove natural variability, the WFAT–M score Y_{ij} predicted by equation 1 likely includes some element of that as well as variability associated with anthropogenic disturbance. Consequent assumptions are:

- that efforts to design out the natural population variance have resulted in an index that mainly captures anthropogenic disturbance, albeit with some unknown artefact of natural population variability, and

- to the extent that natural population variability persists in WFAT–M overall scores, it does not interact with anthropogenic disturbance. In other words, the likelihood of a particular freshwater floodplain wetland exhibiting a certain level of anthropogenic disturbance is independent of the natural characteristics of the wetland.

Section 3.2.3 describes a method to test for the existence of different types of natural spatial variability in the data, such as variability due to climatic zone and wetland habitat type and to determine whether any such variability is independent of the level of anthropogenic disturbance.

2.2.2 Extraneous variance

Extraneous variance comprises all variance components that are not part of population variance. Different sources of extraneous variability can be teased out as follows (again, based on Larsen et al, 2001):

Year variance

This relates to the amount by which WFAT–M scores for all wetlands in the sub-population are high or low in a particular year, due for example, to climatic factors such as rainfall or Southern Oscillation Index (SOI). In the absence of a trend in wetland condition, this variance component fluctuates around a central value. If there is a trend, it will fluctuate around the trend line. Such concordant year variance in the data will markedly decrease the power of statistical analyses to detect trend. Remedies for this are to wait until a sufficient number of monitoring periods have passed to restore power to an acceptable level, or partial out concordant year variance using a covariate, such as annual variability in rainfall across the whole GBR catchment or in the SOI. Increasing the sample size will have no impact on the effect of year variance on power (Urquhart et al, 1998, Urquhart and Kincaid, 1999).

Wetland x Year interaction

The condition of an individual wetland fluctuates from year to year around its own central value or trend line independently from all other wetlands. This component describes that part of wetland's year-to-year variation not already accounted for by the 'year' component. To estimate this interaction component it is necessary to revisit multiple wetlands several times during the same index period[6] and to do so for several years. In the GBR wetland monitoring design, which does not include revisits during the same index period, 'wetland x year' is confounded with index variance (see below) and hence, contributes to the residual component of the model in equation 1.

Figure 1 graphically represents the difference between year variance and wetland x year interaction.

[6] The 'index period' is the testing season; that part of the year within which wetlands are assessed in the GBR wetland monitoring program.

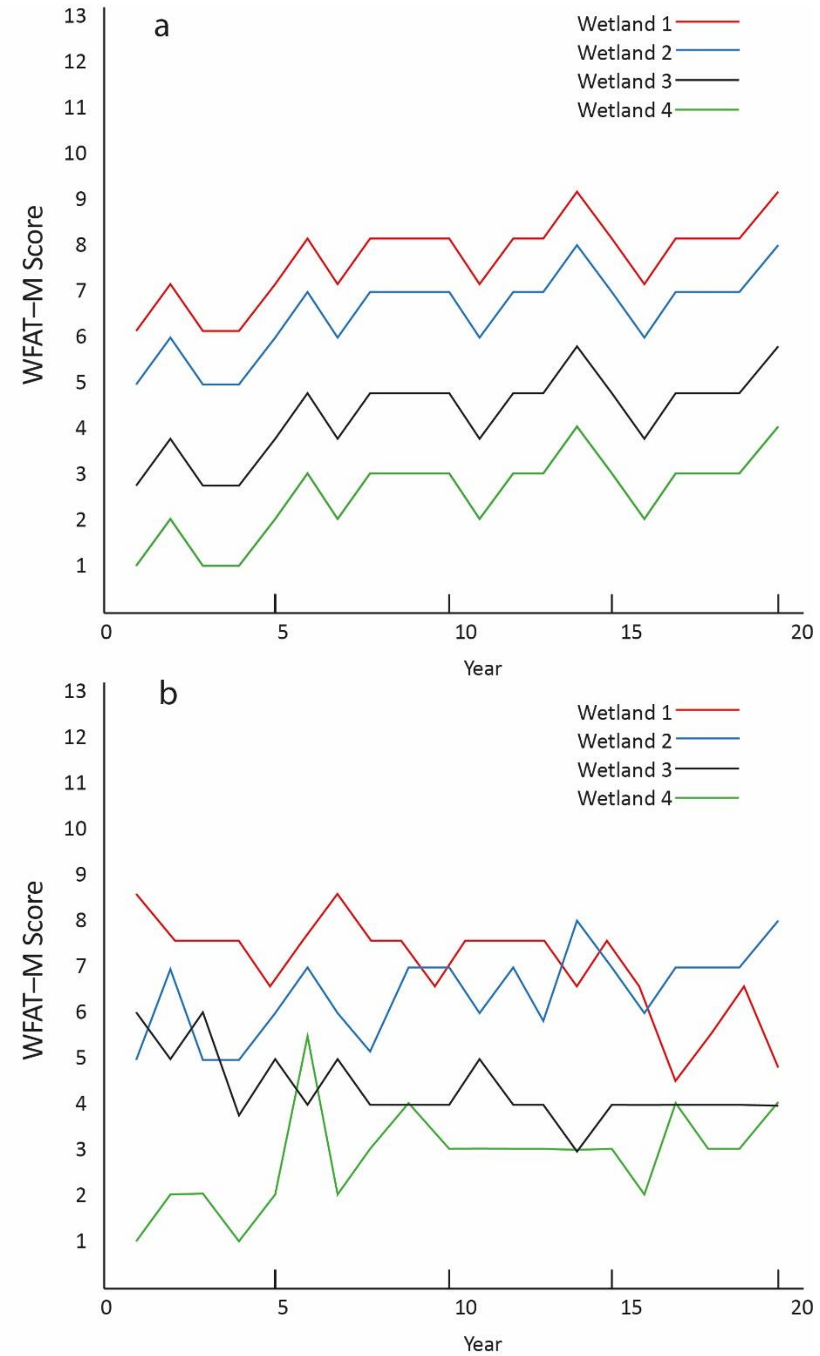

Figure 1: Diagrams a and b contrast two important components of variance. Diagram a illustrates concordant year-to-year variation among four wetlands as a result of annual variability across the whole GBR region, diagram b shows the independent year to year variation among four wetlands. It can also be characterised as the interaction between a particular wetland and its local conditions in any given year. Both are sources of extraneous variance. Diagram a shows 'year' variance and diagram b 'wetland x year interaction'. Interaction variance can be reduced by increasing the number of wetlands in the sample but concordant year variance cannot. Only the passage of time will increase power to detect a trend in the data with even a modest amount of concordant year variance. (After Larsen et al 2001).

Index variance

Index variance is defined as 'the aggregate variation seen by repeated applications of a sampling protocol for the indicator of interest during the index period' (Larsen et al, 1995). In the case of the GBR wetland monitoring program, the index period is the Queensland dry season, notionally, from the beginning of April to the end of September. Index variance is a component that may be amenable to reduction by improvements to the methodological design of the WFAT–M instrument. Index variance includes:

- Measurement variance – the component of variation introduced anywhere in the sequence of events from the point of data collection to the final location of the validated and verified

data point in the database. This is sometimes called error variance. It is reduced by improving data handling methods.

- Inter-operator or team to team variance – from differences introduced by different sampling teams using the same protocol. Variance associated with inter-operator reliability can be reduced by training and 'calibrating' operators to interpret empirical phenomena (such as percentage vegetation cover) consistently and by removing as much ambiguity as possible from descriptions of standard operating procedures so all operators apply them the same way.
- Sampling effort variance – this is variance arises from the different sampling effort applied by the same operator at any given time in any individual wetland. For example, effort may vary with the time of day, the weather and other factors affecting individual effort.
- Local spatial variance – this arises from applying the same protocol in different parts of a wetland on the same assessment occasion. Variation from this source is reduced in trend assessments by locating traverse and sampling sites as close as possible to the same locations in repeat assessments. For status assessment, local spatial variance is introduced because all operators would not choose the same initial set of sites to conduct assessments. To some extent this source of variance is controlled by having clear instructions about how to choose sampling sites. Otherwise, this source of variance (along with all other components of index variance) is controlled by increasing the sample size.
- Temporal variance and trends – within the index period a wetland may vary in consistent ways, for example there may be a tendency for wetlands in dry places to achieve better scores in the early part of the index period when moisture levels are still relatively high. Keeping seasonal/monthly timing of repeat assessments consistent helps reduce variance from this source.
- Remaining index variance – all other unaccounted for sources of indicator noise during the index period.

Together, wetland x year interaction variance and index variance contribute to the residual component of equation 1. Their effect on the precision of status or estimation of trend detection can be reduced by increasing the number of wetlands in the sub-sample.

3 Analyses of WFAT–M data

Table 3 sets out planned analyses for each period in the first cycle of the monitoring program. Periods are notionally of 1 year duration, with 40 wetlands (two panels of 20) being assessed per year; however, the first period was extended over two years due to insufficient resources being available in 2015 to assess 40 wetlands. The baseline data collection for the monitoring program comprises two periods, the 2015–16 period and the 2017 period for a total of 60 wetlands assessed.

The table is a summary of the planned analyses. It is followed by more detailed descriptions of the planned analyses. Sections are organised by assessment period.

Table 3 Great Barrier Reef wetland monitoring planned analysis for first assessment cycle

Period	Data gathered	Analysis suggested	Aim of analysis
2015–2016	20 panel 1 20 panel 2 N=40	1. Descriptive statistics, parametric and non-parametric e.g. mean and variance of WFAT-M scores for pressure and state, medians, interquartile ranges (see 3.1.1 for more detail)	Describe pressure on wetland values and state of wetland values at time 1. Compare variance in pressure and state scores with 2014 pilot data used in power analysis to determine sample size.
		2. Analysis of individual indicators (see section 3.1.2 for more detail)	Evaluate indicators. Decide whether to reject indicators or calibrate them. Distribution of indicator scores helps determine how they will be analysed (e.g. should scores be transformed?)
		3. Analysis of contribution of various indicators to WFAT–M pressure and state index scores, including sensitivity analysis, exploratory factor analysis (see Section 3.1.3)	Assesses performance of indicators in the context of the instrument as a whole: what is the consequence of dropping out indicators one by one? Which indicators covary? Which are orthogonal? What does each indicator contribute to explaining the variance in WFAT-M scores? Is there redundancy among the indicators that would suggest removing some indicators? Do sensible underlying factors emerge from the data? Which indicators would be most useful to include in a model to predict WFAT-M score, based on variables that can be assessed on the desktop?
		4. Test for non-response bias (Section 3.1.4)	To establish whether a suspected non-response bias persists in the full period one dataset (40 wetlands).
		5. Calculate post-stratification weights to apply to regions in subsequent analysis (see Section 3.1.5 and Appendix 1)	Correct for failed assumption made in modifying the GRTS selection rules, namely that, with respect to region, wetlands would be non-responsive completely at random. The rules were modified because of the impracticality of replacing non-responsive wetlands in one region with the next wetland in a *different* region.
		6. Calculate non-response weights and combine with post-stratification weights to yield a set of weights which can be applied to WFAT–M data analysis (see Section 3.1.5 & Appendix 1).	Acknowledge and/or correct for potential non-response bias in the data.
		7. Test correlation between WFAT–M pressure and state scores for 40 wetlands and landscape hazard scores for the wetlands' surrounding ACA sub-catchments.	To validate the WFAT–M with available landscape level (Level 1) data.

Period	Data gathered	Analysis suggested	Aim of analysis
2017 Note, this completes the baseline assessment. The total number of wetlands in the baseline sample = 60.	20 panel 1 repeated 20 panel 3 N=40	1. Descriptive statistics, parametric and non-parametric (Section 3.2.1)	Describe pressure on wetland values and state of wetland values at time 1. Compare variance in pressure and state scores with 2014 pilot data used in power analysis to determine sample size and comment on any difference.
		2. Begin to examine components of variance. For panel 1 data, measured twice, is there any sign of concordant year variance across wetlands? (Section 3.2.2)	Concordant year variance could greatly reduce the power of the WFAT-M to detect trends in wetland condition. If there is any sign of this in the data we will identify potential covariates that may allow us to factor it out once sufficient data have been gathered.
		3. For panel 1 and 3 data (assessed in the same year), look at spatial variability. (Section 3.2.3)	Natural site to site differences, will contribute noise and reduce power to detect trend. If this seems likely, we will identify any covariates, such as tropical vs subtropical, density class or broad habitat type that will allow us to factor out some of the spatial variance. The WFAT–M aims to detect anthropogenic disturbance, not natural spatial variability among wetlands.
	Additional assessments to test inter-operator reliability.	4. Intra-class correlation using a mixed factorial design (Section 3.2.4)	To test the reliability of the WFAT-M. To estimate a component of the residual variance. To assess the performance of individual indicators. To explore ways of reducing this source of variance in the data. To look for any bias arising as a result of systematic differences among assessors.
2018	20 panel 1 20 panel 2 N=40 wetlands assessed at least twice	By now a method of handling non-response bias should have been decided and there should be some idea of the effects of different components of variance on the precision of the WFAT-M. Also panel 1 wetlands have now been tested 3 times.	
		1. Recalculate power statistics. (Section 3.3.1)	Establish precision of estimates of wetland pressure and state across the GBR – how good is our data?
		2. Test for change between period 1 and period 2. (Section 3.3.2)	This is the first opportunity to report on the ecological component of the Reef Plan 2013 wetland target. Looks for change in GBR wetlands between time 1 and time 2.
		3. Examine panel 1 data over three times for evidence of concordant year variance. (Section 3.3.3)	To detect concordant year variance in the data.
	Additional assessments for seasonal variability. Randomly choose wetlands from panels 1 and 2 to assess multiple times during the index period.	4. Estimate variance attributable to interaction between year and wetland. (Section 3.3.4)	This is a source of noise in the data and power to detect change would increase if it could be factored out.

Period	Data gathered	Analysis suggested	Aim of analysis
2019	20 Panel 1 20 Panel 3 N=60 wetlands assessed at least twice	Test for change with greater precision	This is the first full repeat of the baseline sample. The aim is to improve the precision of change detection.
2020	20 Panel 1 20 Panel 4 (new) 20 wetlands assessed 5 times	1. Test for trend if power is high enough to warrant it.	Report on wetland target. Looks for trend in GBR wetlands across time 1, time 2, time 3 and time 4.
		2. Test for difference between panel 1 and panel 4.	Look for impact of repeated assessment on panel 1. Are panel 1 and panel 4 still part of the same population?
2021	20 Panel 1 20 Panel 5		
2022	20 Panel 1 20 Panel 4		
2023 (end cycle)	20 Panel 1 20 Panel 5		

3.1 2015–2016

By the end of 2016, 40 wetlands had been assessed. This data will be reported as 2015–16 baseline data. The 2017 assessments of panel three will provide an additional 20 wetlands for a *total* baseline sample of 60 wetlands.

Forty wetlands was estimated to be the minimum sized sample needed to detect change between two periods with a precision of ±1 point on the WFAT–M scale (power = 80%, α = .05); however, it is essential to have a baseline sample of more than 40 wetlands to allow for the inevitable attrition of wetlands from the sample and also to permit the use of multivariate techniques (see for example section 3.1.3).

3.1.1 Summary statistics

Descriptive statistics, parametric and non-parametric will be calculated. Mean and variance of WFAT–M scores for pressure and state will characterise the baseline level of anthropogenic disturbance to GBR natural freshwater floodplain wetlands, along with the mean and variance per wetland environmental value (WEV) for pressure and state.

Non-parametric measures of central tendency and dispersion around the mean will also be reported.

These statistics will be compared with values from the 2014 pilot study to see whether predictions (such as lower variance in the randomly drawn 2015–2016 baseline sample) are upheld.

The summary of wetland baseline status will be used to produce the first report card from the GBR wetland monitoring program.

3.1.2 Analysis of individual indicators

The distribution of scores for each indicator will be plotted and summary statistics calculated to see whether indicators capture the spectrum of wetland disturbance in a meaningful way. Indicators

that do not discriminate will be rescaled, if appropriate, or rejected from the WFAT–M index if not. The possibility of transforming indicator scores as part of calibrating the WFAT–M will be explored, although at this stage there is no identified independent data or instrument to which the WFAT–M index can be calibrated. See Section 4 for further discussion of this challenge.

3.1.3 Indicator contribution to pressure and state indices

Spearman's rank correlations will be calculated to examine the influence of individual indicators on overall WFAT–M scores of pressure and state and on WEV scores, as well as the contribution of each of the WEV scores to the overall WFAT–M scores. The influence of individual indicators will be further examined by dropping indicator scores one at a time from WFAT–M score calculations to see how this affects the overall scores and the WEV (sub-index) scores. The influence of WEVs on the overall WFAT–M scores will be examined in a similar manner. Applicable methods are described in Robinson and Lee's (2009) sensitivity analysis of the AQUABAMM index.

Correlations between individual indicators and with the aggregated pressure and state scores per wetland and per WEV will show whether there is redundancy in each of the indices, pressure and state. Consideration will be given to reducing any redundancy.

The performance of the four sub-indices will also be examined. Are indicators correlated with their own sub-indices or with other sub-indices? Or are they not correlated with any of the sub-indices? What are the strengths of these correlations? Following the pilot study, indicators were rearranged within a slightly changed set of sub-indices and some new indicators were added. The performance of these changes will be evaluated using the correlation analysis.

As well, Exploratory Factor Analysis (EFA) will be attempted to examine the underlying structure of the wetland data. The constraint on performing such a multivariate analysis is the small sample size. Various rules of thumb have been suggested for the minimum ratio of sample size to number of independent variables for sound use of multivariate techniques, ranging from ratios of 200:1 to 3:1 (see for example the Handbook on Constructing Composite Indicators, OECD 2008, p 66). De Winter et al (2009) explore sample size limits for EFA and find that minimum sample size for factor recovery depends on factor loading (the higher the better), the number of factors extracted (smaller is better) and the number of independent variables being assessed (smaller is better). The WFAT–M indices for State and Pressure currently have 12 and 17 indicators (independent variables) respectively. Based on de Winter et al (2009) findings, for factor loadings greater than 0.6, a sample of 40 would allow two factors to be recovered, while 60 would give acceptable results to three factors. Four factors can be recovered with factor loadings of greater than 0.8 and sample sizes as low as 23. De Winter et al (2009) also found that EFA methods were relatively robust to unequal loadings within and between factors. Of interest is whether factors extracted reflect the sub-index structure of the WFAT–M.

If EFA is not reportable after 2015–16 data gathering, it will be done at the end of 2016–17 when 60 wetlands will have been assessed.

3.1.4 Test for non-response bias

When using an oversample to replace non-responsive wetlands[7] in a GRTS spatially balanced random sample, the standard procedure is to replace the non-responding wetland with the next wetland in the GRTS draw. We drew wetlands, with equal probability, from across the whole GBR

[7] Non-responsive wetlands are wetlands that cannot be assessed for some reason, such as landholder refusal or inaccessibility.

catchment, without stratifying by region. Because the draw is catchment-wide, the next wetland in the GRTS list was often from a different region to a non-responding wetland. This proved to be logistically unwieldy. Due to the contractual arrangements of the project, the process of contacting landholders for permission to access wetlands was organised on a Natural Resource Management Region basis, therefore replacement of non-responding wetlands also had to be organised regionally.

To allow this, we varied the GRTS rules by sorting the sample into NRM regions and replacing non-responders with wetlands from the same region. This process would have maintained a spatially balanced random sample of wetlands across the whole GBR catchment, had the non-responsive wetlands been missing completely at random (MCAR). In the interest of being able to use our sample to make inferences about the target wetland sub-population across the whole GBR, it is necessary to test this proposition – that wetlands *are* missing completely at random.

Early experience of seeking permission to assess wetlands in the 2015–16 baseline sample strongly suggested that non-responsive wetlands are likely not to be MCAR i.e. that there would be a non-response bias in the resulting sample and that this would be related to the intensity of land use. When wetlands were surrounded[8] by conservation lands, access was always granted to assessors, while wetlands in areas of intensive land use such as cane production or mining were often non-responsive because access was refused. Moderate levels of land use intensity (e.g. extensive grazing) seemed to be associated with non-response levels about midway between those of conservation lands and intensive land use. After seeking access to 54 wetlands, the null hypothesis that wetlands were missing completely at random was tested using Fisher's Exact Test and rejected ($p < .0001$). Figure 2 shows the proportion of wetlands to which owners allowed access in each of 3 land use intensity categories.

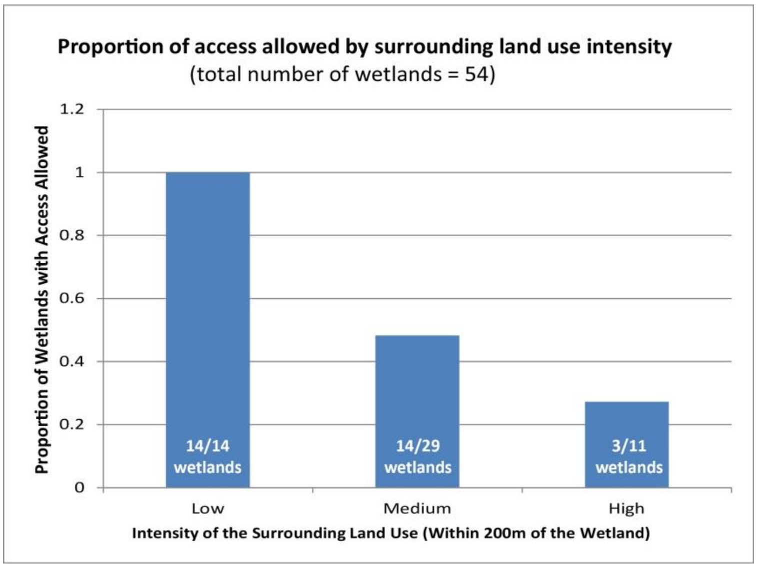

Figure 2: Proportion of wetlands accessed by land use intensity.

The null hypothesis of wetlands missing completely at random will be tested again after the first period of baseline data collection (2015–16).

3.1.5 Dealing with non-response bias – step one: post stratification weighting by region

Assuming a similar level and type of non-response bias emerges in the full set of period 1 baseline data, two actions are needed. First, because non-response is (likely) correlated with region (due to

[8] Defined by land use within the 200 m buffer zone of the wetland

broad differences in land-use intensity between regions), the MCAR assumption that allowed us to vary the GRTS method for replacing non-responding wetlands does not hold. There is no longer an equal probability of any wetland in the sub-population ending up in the sample. Rather, a landholder of any given wetland will be proportionally more likely to be approached for permission to assess that wetland if the wetland is in a region of high land-use intensity, such as the Wet Tropics and less likely to be approached if the wetland is in a place such as Cape York which has a low level of intensive land use. To correct for this, post-stratification weighting of wetland scores by region will be applied (Johnson, 2008; Kolenikov, 2016).

3.1.6 Dealing with non-response bias – step 2: acknowledging and/or correcting for bias

The average level of disturbance of GBR wetlands (status) will be misrepresented by a sample with a non-response bias, although the bias will not affect the ability to detect trend. Provided enough wetlands have been assessed, we will still be able to say whether the amount of disturbance to wetlands has increased, decreased or remained the same.

There are various ways to deal with the non-response bias. The preferred method or methods from among these are still under exploration. They are:

1. **Estimate the bias**. Try to estimate the bias and report a correction. Two ways of doing this are as follows:
    a. Using desktop variables as predictors, create a model based on wetland desktop and field assessment data from wetlands whose landholders allowed access. Predict the scores of the non-responding wetlands using this desktop predictor. Use the difference in means between the two groups (i.e. responding vs non-responding) to estimate the bias.
    b. Make an extra effort to get some of the owners of non-responding wetlands to respond, then use those wetland assessments to predict the bias. (This is not our favoured approach for statistical as well as logistic reasons.)
2. **Weighting**: given the estimated non-response bias, which is related to land use intensity (in the 200 m buffer), recalculate the probability of each wetland ending up in the sample and use *weighting* to determine the contribution of each wetland to the GBR-wide statistics. Wetlands less likely to end up in the sample carry more weight (or to put it another way, in the relationship between the sample and the population, they 'stand for' more wetlands.) These weights will be combined with those applied for regional corrections (see 3.1.5). To get the correct weights it will be necessary to estimate any interaction between region and land use intensity in determining the likelihood of non-response. This interaction factor will be eliminated from the calculated weight. The proposed method of calculating and combining post-stratification and non-response weights is given in appendix 1.
3. **Imputation**: If a model can be developed that accounts for enough of the variance in the data, use this model (based on desktop variables), to impute scores to those wetlands not 'missing completely at random' (i.e. when permission is refused) and include these wetlands in the data analysis.
4. **Report the bias**: Without trying to estimate or correct for the bias, admit that it exists and include in reports comments such as 'caution is needed in interpreting these data as they are likely to underestimate the degree of disturbance of wetlands. This is because permission to visit wetlands was related to surrounding land use, which is likely to be correlated with the level of disturbance.' This warning should be prominent to avoid having people use the data in ways that are not appropriate.

In the interests of transparency, results will be reported both with and without any correction made to deal with non-response.

3.1.7 Validation of WFAT–M pressure and state indices with landscape hazard data

Because no recent GBR-wide gold standard level 3 data[9] are available to validate the WFAT–M, we propose to use multiple lines of evidence for validation. This method follows the approach used by Stein et al (2009) to validate the California Rapid Assessment Method (CRAM) for wetlands, a rapid assessment method comparable to the WFAT–M. Included in the lines of evidence explored by Stein et al (2009) was the relationship between CRAM scores and the Landscape Development Index (LDI), a level 1 landscape measure of human disturbance. Correlations between CRAM overall scores and LDI were used as a measure of CRAM responsiveness to condition along a gradient of stress.

A comparable level 1 dataset for validating the WFAT–M is the GBR-wide land-use hazard to wetland data compiled in 2013 (DSITI, 2015). In this study, catchments across the GBR were assessed for their level of land-use hazard to wetlands. Aquatic Conservation Assessment (ACA) sub-catchment spatial units (see e.g. Clayton et al, 2006) were used in these hazard assessments with each of these sub-catchments receiving an overall land-use hazard score on an ordinal scale of one to six.

In the 2014 GBR wetland monitoring pilot study (reported in Tilden et al, 2015) a preliminary validity check of the WFAT–M pressure and state scores was performed using the Level 1 (Desktop) landscape hazard data. For the wetlands assessed in the pilot study, the landscape hazard scores for their surrounding sub-catchments were found to be highly correlated (Spearman's Rho) with both pressure on wetland environmental values ($\rho = 0.64$, $p < 0.01$) and state of wetland environmental values ($\rho = 0.64$, $p < 0.01$) as measured by the WFAT–M.

The WFAT–M is based on Driver–Pressure–State–Impact–Response (DPSIR) conceptual modelling (Figure 3). Pressure classes and expert panel judgements from the land-use hazard assessment were used to operationalise the conceptual link between driver and pressure in the development of WFAT–M pressure indicators.

In that sense, the finding of a correlation between landscape hazard and WFAT–M overall *pressure* scores would indicate that the operationalisation had been successful. However, this finding would be trivial with regard to the validity of the WFAT–M because the relationship between land-use hazard scores and WFAT–M overall pressure scores has, in effect, been put in.

This is not the case for the overall state scores. Based on the WFAT–M DPSIR conceptual modelling and assuming that land-use pressures on wetlands are largely unmitigated, a correlation between land-use hazard score and the overall *state* of wetland environmental values would be evidence that the WFAT–M state index is valid for establishing the condition of wetlands along a disturbance gradient. This is because, unlike the WFAT–M pressure index, the development of indicators in the state index was independent of the landscape hazard assessment. The correlation between land-use hazard and WFAT–M overall state score therefore *validates* this conceptual link, rather than *establishing* it (as in the case of the driver–pressure link).

While more work is needed to validate that the WFAT–M really is measuring what it purports to measure, this is one line of evidence supporting that claim.

---

[9] In a three-tiered approach to monitoring and assessment, Level 1 consists of habitat inventories and landscape-scale assessment, Level 2 consists of rapid assessment, and Level 3 consists of intensive assessment (Kentula, 2007).

The analysis of the relationship between landscape hazard to wetlands and WFAT–M pressure and state scores will be repeated for the sample of 40 wetlands assessed in 2015–16 to determine whether the relationship found in the pilot study persists.

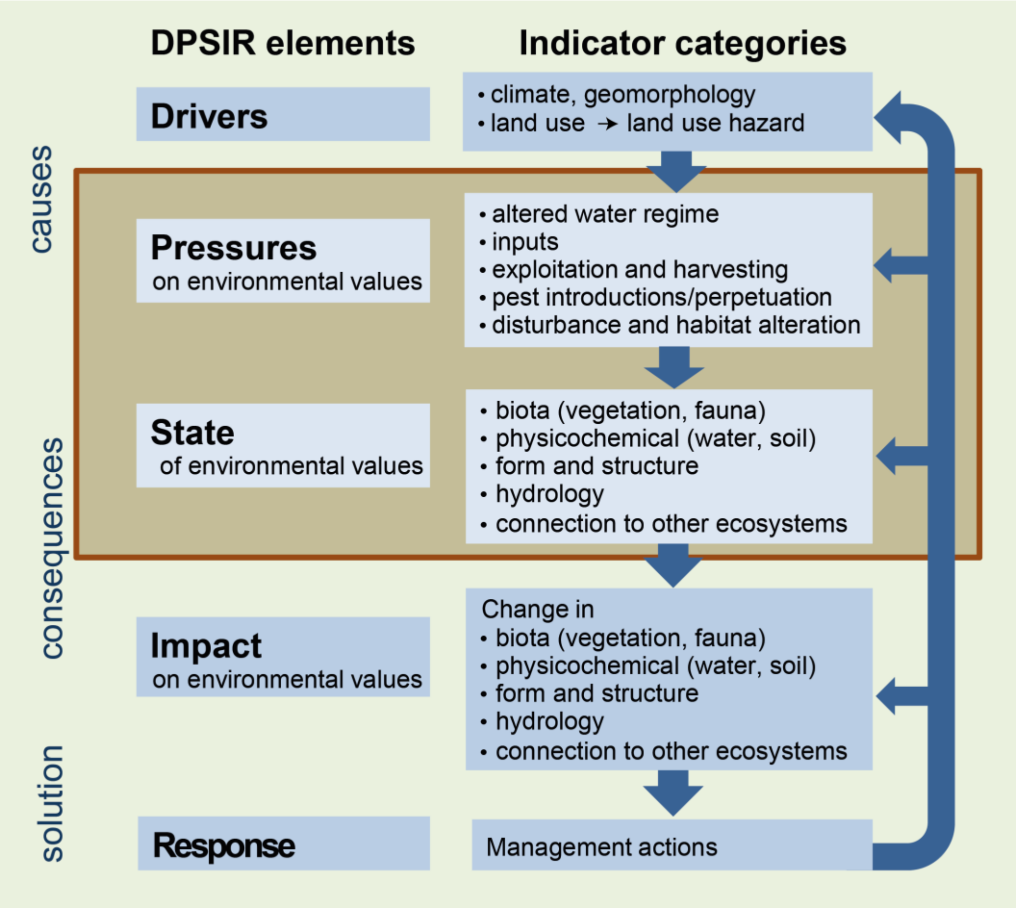

Figure 3 DPSIR conceptual framework linking land-use hazard and WFAT–M scores for pressure on environmental values and state of environmental values

3.2 2016–2017

By the end of 2017, three panels of wetlands for a total of 60 wetlands will have been assessed. Panels 2 and 3 will have been assessed once and panel 1 twice. This completes the baseline assessment period. During this period, we also aim to use a number of wetlands to test the inter-operator reliability of the WFAT–M.

3.2.1 Summary statistics

Statistics reported in Section 3.1.1 will be recalculated using the total baseline sample and compared with those obtained in the 2015–16 preliminary baseline period.

3.2.2 Look for any indication of concordant year variance

The twenty wetlands in panel 1 will have been assessed twice by the end of 2017. This provides the first opportunity to examine the data for any sign of concordant year variance (see Figure 1), an important source of extraneous variance listed in 2.2.2.

While analyses associated with partitioning variance will be conducted, as much as possible, using the R based *spsurvey* package employed for the GRTS draw, the statistical logic behind the planned analyses is spelled out here.

Matrices of WFAT–M scores provide the data for estimating components of variation, once sufficient data have been gathered. For pressure on wetland values, each wetland is characterised by a single index score (overall WFAT–M pressure score) each year and likewise for state of wetland values (overall WFAT–M state score). These scores are equivalent to the term Y_{ij} in equation 1, which represents the WFAT–M score of wetland *i* and time *j*. So the two dimensions of the data matrix are wetlands x years. Scores on the response variable (pressure or state) fill the matrix cells, resulting in two data matrices, one for pressure and one for state.

In terms of the model given in equation 1, these observations on 20 wetlands assessed in two consecutive monitoring periods can be expressed as in Table 4.

Table 4 WFAT–M scores for panel 1 assessments in two consecutive periods, expressed in terms of a linear model to predict scores

	Year	
Wetland	1	2
1	$Y_{1,1} = \mu + W_1 + T_1 + E_{1,1}$	$Y_{1,2} = \mu + W_1 + T_2 + E_{1,2}$
2	$Y_{2,1} = \mu + W_2 + T_1 + E_{2,1}$	$Y_{2,2} = \mu + W_2 + T_2 + E_{2,2}$
…		
i	$Y_{i1} = \mu + W_i + T_1 + E_{i1}$	$Y_{i2} = \mu + W_i + T_2 + E_{i2}$
…		
20	$Y_{20,1} = \mu + W_{20} + T_1 + E_{20,1}$	$Y_{20,2} = \mu + W_{20} + T_2 + E_{20,2}$

Analysis of variance can be applied to such a matrix to estimate the magnitude of each component of variance – year, wetland and remainder, the latter representing the combined variance of the wetland x year interaction and the index variance (Larsen et al, 2001).

With just two periods of data from 40 wetlands to analyse, it will not be possible to confirm any modest year effect in the data; however, as even a small year effect will adversely affect the power of the design to detect trend, early exploratory analysis for such an effect is a worthwhile exercise.

Generally, analyses of data in early years will focus on detecting changes between years (treating years as a discrete factor) rather than seeking to detect trend.

3.2.3 Spatial variability (components of W_i)

In Section 2.2.1, assumptions were made that the WFAT–M is blind to most natural spatial variability in wetlands, with just a small artefact remaining, and that any such artefact is independent of anthropogenic factors. The following investigation of the effects of spatial variables provides an opportunity to test these assumptions as well as potentially identifying explanatory effects that could be included in models to reduce the overall error variance and add power to analyses.

Together, panels 1 and 3 comprise 40 wetlands in a spatially balanced random sample from the same sub-population of GBR wetlands. These will be post-sampling stratified on one or more

spatial variables and data analysed for differences among levels of these variables. For example, if wetlands are stratified according to their broad habitat type (tree swamp, grass sedge herb swamp etc) analysis of variance can be used to see whether there are differences in the mean and variance of WFAT–M scores related to habitat type. The same exercise can be applied to other spatial variables such as tropical vs subtropical climate or density class). If the chosen spatial variables are irrelevant to WFAT–M score, as hypothesised, there should be no significant differences in mean or variance between levels of any given factor. The existence of such differences is a sign that natural spatial variability persists as a significant component of W, the population variance of wetlands.

To determine whether different types of wetlands vary in their propensity to be disturbed by human activity, an independent measure of anthropogenic disturbance will be crossed with each of the spatial variables. For example, wetland habitat type, broad climate zone or density class will be crossed with land use hazard or land use intensity classes. Interactions between spatial variables and anthropogenic disturbance factors would be a sign that different kinds of wetlands are more likely to be disturbed by human activities than others. This could be an issue if the state of wetland values or pressure on wetland values, as measured by the WFAT–M, is being compared across regions, basins or other areal attributes.

3.2.4 Test for inter-operator reliability

One component of index variance that may be amenable to reduction by improvements to the methodological design of the WFAT–M instrument is inter-operator or inter-team variance. Because this component of variance could be high and also because reducing it involves changes to the WFAT–M instrument, which in turn could lead to issues about the comparability of data, it is important to measure or estimate this variance as soon as logistically possible.

We propose to complete this in 2017 by having each of the five field-trained members of the WFAT–M team assess an additional set of wetlands (number to be determined).

All operators would assess all wetlands in this additional set independently, to give a fully crossed design that can account for differences between operators, differences by wetland across all operators and interactions between wetlands and operators. The output of this analysis is a measure of correlation varying from zero – only random agreement among operators – to one – complete agreement. Negative correlations, that is, measures of disagreement, are also possible. Hallgren (2012) offers cut-off points for poor (<0.39), fair (0.40–0.59), good (0.60–0.74) and excellent (0.75–1) correlation. Significance as well as strength of correlation will be assessed.

When used to measure a component of index variance, intra-class correlation helps (a) to determine the power of the GBR wetland monitoring design to estimate wetland status and trend or, (b) to determine how many wetland assessments would needed to achieve a desired power (0.8 when α is set at 0.05).

3.3 2018

By the end of 2018, 40 wetlands will have been assessed twice. Conservative assumptions were applied to pilot data to determine that 40 wetlands was the minimum number needed to detect a change of ± 1 point on the 13 point WFAT–M scale after two periods (power = 0.8, α = 0.05). For example, it was assumed that the variance of a randomly selected sample of wetlands would be lower than that of the purposive pilot study sample designed to capture the full range of anthropogenic disturbance to natural freshwater wetlands. So using the variance of the *pilot sample* to determine the sample size should overestimate the minimum number needed.

Assumptions (not conservative) were also made about the WFAT–M in determining this minimum sample size, namely, that it is a valid and reliable instrument for measuring the level of anthropogenic disturbance to natural freshwater floodplain wetlands across the GBR catchment.

In 2016, indicators in the WFAT–M will undergo sensitivity and redundancy analysis and in 2016–17, inter-operator reliability testing will be carried out on the WFAT–M. As a result, the format of the WFAT–M will have changed slightly by 2018. As the aims of refining the WFAT–M are to establish validity, increase reliability and reduce error variance, it is hoped that the updated instrument will score wetland pressure and state more precisely, increasing the statistical power to detect change using a sample of 40 wetlands.

3.3.1 Recalculate power to detect change

By the end of 2018, decisions will have been made about how to handle the issue of non-response bias. This is relevant here because one consequence of the putative non-response bias is that it reduces the variance of the baseline data. While the variance in the randomly selected sample of monitoring program wetlands is predicted to be lower than that of the purposive sample used in the pilot study, this is not the reason for that prediction. The existence of a non-response bias will lead to an underestimate of the true underlying variance of the sub-population and this must be corrected or accounted for in determining the minimum number of wetlands to be assessed.

As well, some components of the variance in WFAT–M scores:

$$Y_{ij} = \mu + W_i + T_j + E_{ij}$$

will have been identified by the end of 2018 allowing some of the variance to be reduced or partialled out and improving the power of the monitoring design to detect change.

3.3.2 Test for change between time 1 and time 2

A paired-sample t test can be used to determine whether there is any change in GBR wetland status (of pressure or state) between the 2015/16 assessments of panels 1 and 2 and the 2018 assessments of panels 1 and 2, that is, all the wetlands that were assessed in both years. With α = 0.05 and assuming that 40 wetlands gives a power = 0.8, acceptance of the null hypothesis that there has been no change greater than ± 1 point on the WFAT–M scale entails a 20% chance of making a type II error, that is, claiming that there has been no change in average wetland status when there really has been a change.

3.3.3 Look for concordant year variance

Analysis of variance can be applied to the matrix in Table 5 to estimate the magnitude of each component of variance – year, wetland and remainder, the latter being the combined variance of the wetland x year interaction and the index variance (Larsen et al, 2001).

Table 5 WFAT–M scores for panel 1 assessments in three consecutive periods

	Year		
Wetland	1	2	3
1	$Y_{1,1} = \mu + W_1 + T_1 + E_{1,1}$	$Y_{1,2} = \mu + W_1 + T_2 + E_{1,2}$	$Y_{1,3} = \mu + W_1 + T_3 + E_{1,3}$
2	$Y_{2,1} = \mu + W_2 + T_1 + E_{2,1}$	$Y_{2,2} = \mu + W_2 + T_2 + E_{2,2}$	$Y_{2,3} = \mu + W_2 + T_3 + E_{2,3}$
…			
i	$Y_{i1} = \mu + W_i + T_1 + E_{i1}$	$Y_{i2} = \mu + W_i + T_2 + E_{i2}$	$Y_{i3} = \mu + W_i + T_3 + E_{i3}$
…			
20	$Y_{20,1} = \mu + W_{20} + T_1 + E_{20,1}$	$Y_{20,2} = \mu + W_{20} + T_2 + E_{20,2}$	$Y_{20,3} = \mu + W_{20} + T_3 + E_{20,3}$

Variance components are estimated using the following set of equations:

$S^2_t = S^2_w + S^2_y + S^2_i$

$S^2_i = MS_i$

$S^2_w = MS_s / N_y$

$S^2_y = MS_y / N_w$

Where S^2 is an estimate of variance, MS is mean square from the ANOVA summary, N is the number of wetlands or years and the subscripts are: t = total, w = wetland, y = year and i = index.

3.3.4 Estimate variance attributable to interaction between year and wetland.

Resources permitting, a sub-sample of wetlands could be tested twice or three times within one year to estimate extraneous variance caused by wetland x year interaction. Without this test, wetland x year (interaction) variance is an unknown component of index variance.

Unless year variance (as described in the previous section) can be determined to be very low it will not be worth making this investigation of the interaction between year and wetland. This is because the remedy for high interaction variance is to test more wetlands, while the remedy for year variance (concordant) is time – with each repeat of the testing cycle, the power to discern trend apart from concordant year variance increases. If concordant year variance is a significant factor, adding extra wetlands to the sample will not increase the power by a sufficient amount to justify the extra cost and effort (Larsen and Christie, 1995).

3.4 2019

By the end of 2019 60 wetlands will have been assessed at least twice and it will be possible to test for change from baseline wetland status with greater precision than in 2018.

3.5 2020

In 2020 a new panel of wetlands (panel 4) is introduced (refer to Table 6). Also 20 wetlands will have been assessed 5 times.

3.5.1 Detecting trend

Depending on estimated total variance in WFAT–M scores and hence power to detect trend, it may be possible to detect a trend in wetland status after 20 wetlands have been assessed 5 times. The method for doing this is explained by Urquhart et al (1998). Given:

$Y_{ij} = \mu + W_i + T_j + E_{ij}$ (equation 1)

Trend can be estimated by regressing WFAT–M scores onto a line whose parameters are β_0, the intercept and β_1, the slope of the trend across years, as follows:

$Y_{ij} = \beta_0 + \beta_1 j + R_{ij}$ (equation 2)

Where Y_{ij} is the predicted score for wetland *i* and time *j* and R_{ij} is the residual component.

The test for trend is then a test of the null hypothesis, that the expected WFAT–M score is equal to β_0 against the alternative that the expected score is equal to $\beta_0 + \beta_1 j$.

With a panel design, the overall estimate of trend is the weighted average of estimates from each set of wetlands having the same temporal visit pattern.

3.5.2 Testing for management effects

After repeated assessments, panel 1 wetlands may no longer be representative of the same sub-population of wetlands as a fresh panel (4) that has never been assessed. For example, the interest shown by DSITI Wetland Science in the 20 wetlands of panel 1 may have resulted in increased levels of management of the wetlands by landholders. Panel 4 can be compared with panel 1 to test for such effects.

4 Ability of WFAT–M to assess status and detect trend in GBR wetlands

The ability of the WFAT–M to assess status and detect change or trend in GBR wetlands rests on some assumptions about the validity and reliability of the WFAT–M that have yet to be fully tested.

4.1 Assessing status

The ability of a rapid assessment method such as the WFAT–M to assess the objective status of a population of wetlands depends on the existence of some gold-standard method of measuring wetland status across the population of interest. Gold-standard would include detailed studies of wetland condition in terms of its known correlates. For existing data to be ideal for validating the WFAT–M, it would be recently gathered from across the whole GBR catchment. No such methods or data exist. The best that can be hoped is that multiple lines of evidence in the form of WFAT–M correlates from more restricted datasets can be found (or generated) to give confidence that the WFAT–M does measure what it purports to measure – the pressure on and state of wetland environmental values.[10]

While attempts are underway to establish the construct validity[11] of the WFAT–M as an instrument and to anchor the scale of measurement to independent assessments of wetland status, there are some statistical properties that any instrument assessing the pressure on and state of wetland environmental values should have in order to be an index capable of supporting assessment of wetland status.

4.1.1 Range

A valid index of wetland condition or pressure on wetlands would be expected to be able to discriminate across the full spectrum of wetland condition from pristine to totally disturbed and across the full spectrum of anthropogenic pressures bearing on wetlands. The 2014 pilot study provides evidence that the WFAT–M meets this condition (see Tilden et al, 2015).

4.1.2 Distribution function

Expert opinion suggests that, while all wetlands would once have been pristine (i.e. not at all impacted by human land management), very few remain in such a state. Therefore a useful index of wetland pressure and condition would be expected to have a distribution function with some sort of a tail towards the least disturbed end of the range of scores. On the other hand, relatively few wetlands in the GBR catchments have been so altered that they no longer retain some wetland environmental values, so a tail towards the more disturbed end of the scoring range is also to be expected.

In the absence of any contradictory conceptual or empirical evidence, it is also reasonable to expect some sort of central tendency in the data, rather than, say, a bimodal distribution. The underlying distribution of the wetland population could be normal, log normal or could follow some other pattern of distribution with a central tendency. In the absence of any empirical information about this underlying distribution, an acceptable distribution in the baseline WFAT–M data would

[10] See Stein et al, 2009 for discussion of 'gold-standard' validation of wetlands and the use of multiple lines of evidence to validate a wetland rapid assessment method comparable to the WFAT–M.

[11] A test has construct validity if it really does measure what it purports to measure.

be one with a central tendency. Analysis of the data would be facilitated by a distribution function with known properties (such as normal or log normal).

Distributions of WFAT–M wetland data collected during the baseline monitoring period will be examined to see whether they conform to these expectations – discriminating across the full range of wetland anthropogenic disturbance and having a distribution with two tails and a central tendency. The possibility will be explored of calibrating the index to exhibit desirable statistical properties. Such calibration will be attempted only if validation of the WFAT–M with multiple lines of evidence confirms that any proposed manipulation of the statistical characteristics of the index is well supported conceptually and by independent evidence.

4.2 Detecting change and trend

Addressing the Reef Plan (2013) wetland target is the primary aim of the GBR wetland monitoring program. That target focuses on improving the ecological condition of natural wetlands in the GBR catchment, so it is fundamental to the program that the WFAT–M and the GBR wetland monitoring design be able to detect change, and ultimately trend, in both the state of wetland environmental values (wetland condition), and the pressure on those values.

A power analysis was conducted on the results of pilot data to determine the minimum number of wetlands to be assessed in order to make a change between periods of ± 1 point on the WFAT–M scale ecologically meaningful. To put it another way, if there is ‘no change’ of that order in pressure or state we want to be able to say that this probably *does* reflect an absence of change (power = 0.8, $\alpha = 0.05$) and not the insensitivity of the instrument. Forty was determined to be the minimum number of wetlands required.

The chosen program monitoring design balanced competing needs – to work within the (then) available resource budget and to minimise the amount of time needed to test whether a sample of 40 wetlands did indeed yield the power to detect change with the desired level of sensitivity. A panel design with a very short initial repeat cycle was chosen because power to detect change, and ultimately trend, is greatly increased at the point when the same wetlands are revisited. This design is reproduced in table 6.

Table 6 Panel design for the WFAT–M monitoring program

Panel	Period									
	1 (2015–16)	**2** (2017)	**3** (2018)	**4** (2019)	**5** (2020)	**6** (2021)	**7** (2022)	**8** (2023)	**9** (2024)	**10** (2025)
1	20	20	20	20	20	20	20	20	20	20
2	20		20						20	
3		20		20						20
4					20		20			
5						20		20		
Year total	40	40	40	40	40	40	40	40	40	40
Total sample	40	60	60	60	80	100	100	100	100	100

The design has two repeat cycles, one nested within the other. The first short cycle finishes at the end of period four and the second at the end of period eight when the long cycle also finishes. It is highly likely that there will be no change in the overall condition of GBR wetlands from the baseline established in periods one and two to the end of the first short cycle; however, at this point (at the latest[12]) we will be able to determine whether the desired precision has been achieved.

The key statistic required is the variance of the *difference* in wetland WFAT–M scores between the baseline (periods one and two) and the first test for change (periods three and four). Exploratory tests on baseline data collected in 2015 (N=18) yield an observed variance of mean pressure on wetland environmental values equal to 4.84. This will almost certainly decrease when more wetlands are assessed (increasing the denominator of the variance formula). And the variance of the *difference* between the two times should be lower still, given the same wetlands are being measured at time 1 and time 2, therefore their scores will be highly correlated. To achieve the desired level of precision for the study, the variance of the *differences* in wetland WFAT–M scores after one cycle should be less than 4.85.

The decision about how to treat the non-response bias in the sample will also affect variance. If the method of imputation is used, there will be a small increase in variance, which can be calculated. On the other hand, the effect of non-response and post-stratification weighting on variance depends on how strongly the adjustment variables (in this case land use intensity and region) are associated with the outcome of the test (in this case, WFAT–M pressure score). High association between adjustment variables and outcome leads to a reduction in variance (Little and Vartivarian, 2004, Thomaidou et al, 2013). Pressure on wetland environmental values is likely to be highly associated with both land use intensity and with region, an association which can be tested.

Further, in generating pilot study 'time two' scenarios, based on the extent to which wetlands could possibly change over one period, the critical variance for estimating power to detect change – the variance of the difference between two times – was always less than the variance at time 1. These scenarios yielded simulated 'time 2' data based on assumptions about how wetlands could change

[12] A preliminary test of precision will be performed at the end of 2018 when 40 wetlands have been tested twice. This is the first opportunity to test the precision of the WFAT–M to detect change.

and were likely to change. If these assumptions are supported, the desired sensitivity should be achieved with the pressure assessments.

On the other hand, preliminary tests on 2015 (N=18) data for *state* of wetland environmental values have variance of 6.5. Again this will decrease for N=40 and should decrease again when difference between two times is considered. State of wetland values is also likely to be highly associated with land use intensity and region, though less so than pressure on wetland values (based on findings in the 2014 pilot study reported in Tilden et al, 2015). At this stage it is not possible to say that the variance of the difference between two times will be less than 4.85, however that is likely to be the case.

Power to detect a *trend* in the state of or pressure on wetland environmental values will increase with each repeat assessment of a panel of wetlands. Hence, if change between times 1 and 2 can be detected with a sensitivity of ± 1 point (power = 0.8, $\alpha = 0.05$) this sensitivity should increase as more repeat assessments are added to the data set. In other words, the panel design set out in Table 6 should yield sufficient power to make ecologically meaningful statements about changes and trends in the status of GBR natural freshwater floodplain wetlands in dense aggregations over time.

4.2.1 The effect of data updates

A factor so far unaccounted for, that will affect the sensitivity of the WFAT–M to detect change, is the cycle of updates for the various datasets used by the WFAT–M indicators. This is especially important for the pressure index, which is based mainly on desktop indicators. Data are not updated yearly, concordant with the WFAT–M assessment cycle, but on different time scales as set out in Table 7. Also, in the case of the land use dataset (Land use mapping – Queensland current) which is an important input to several WFAT–M pressure indicators, the data is not updated synchronously across the whole GBR catchment. As a result, although change may have occurred in any number of individual wetlands between two times, this will not be detected by the pressure index until after data updates have taken place.

Table 7 Data update schedule for datasets used in the WFAT–M

Dataset	Indicators	Version Used for Period 1 (2015-16)	Current Version Available	Planned Updates	Update Frequency
Wetlands	All	Source: Qld Wetland Data V3	Queensland Wetland Data Version 4.0 – Wetland areas 2013 (QDEHP, 2015)*	Version 5 (2017 extent) to be completed and released early-mid 2018	4 years
Landuse	P1, P3, P4, P16	Oct 2015 version (Cape York - 2013; Other GBR catchments - 2009)	Dec 2015 - updated with Wet Tropics 2015 data (QDSITI, 2016)**	Burdekin, Mackay-Whitsunday & Burnett-Mary probably early 2017 (subject to resourcing)	Goal is every 5 years
FPC	P2, P20, S14	Queensland 2011 Foliage Projective Cover Wooded Vegetation	Queensland 2014 Foliage Projective Cover Wooded Vegetation available on request (QDSITI, 2016)	2015 version of the FPC set to be prepared by April 2016	Annual
Current RE	P2, P20, P21, S14	2013 Remnant Regional Ecosystems of Queensland, Version 9.0 (April 2015)	2013 Remnant Regional Ecosystems of Queensland, Version 9.0 (April 2015)	v10 remnant RE 2015 mapping planned to be released in the second half of 2016	Normally every 2 years
Preclear RE	P21	Pre-clearing Regional Ecosystems of Queensland, Version 9.0 (April 2015) (QDSITI, 2016)	Pre-clearing Regional Ecosystems of Queensland, Version 9.0 (April 2015) (QDSITI, 2016)	v10 preclear RE 2015 mapping planned to be released in the second half of 2016	Normally every 2 years

Dataset	Indicators	Version Used for Period 1 (2015-16)	Current Version Available	Planned Updates	Update Frequency
Specified wetland Res	P21	2013 Remnant Regional Ecosystems of Queensland, Version 9.0 (April 2015)	2013 Remnant Regional Ecosystems of Queensland, Version 9.0 (April 2015)	v10 RE 2015 mapping planned to be released in the second half of 2016	Normally every 2 years
Aerial Imagery	P2, P5, P17, P18, S14	Various (date is noted in wetland assessment proformas)			
Sub-catchments	P10	30-50km2 Subcatchments 2014 (QDNRM, 2016)†	2015 version available on request	mid 2016	Normally only at end of reef funding cycle. Next is 2018. Note: exception in 2016 due to critical updates .
Sediment Supply	P10	Fine sediment generation data 2014 (QDNRM, 2016)	2015 version available on request	mid 2016	Normally only at end of reef funding cycle. Next is 2018. Note: exception in 2016 due to critical updates .
Prescribed Flooplains (SWMA)	P15	Source data: Qld floodplain assessment overlay (QFAO) 2013; Queensland Telemetry Gauging Station Network Locations 2014; Wet Tropics Water Resource planning area DNRM 2013; Burdekin River Basin and Haughton River Basin Hydrologic model catchments DNRM 2001; Fitzroy Basin Planning area map DNRM 2011; Mary Basin Water Resource Plan DNRW 2006 (QDSITI, 2014)	Source data: Fitzroy Basin water resource plan was amended on 11 September 2015	QFAO has no schedule for update of the whole dataset - as a local authority does detail flood modelling in an area, this will be included and the remaining data arranged to fit the new data; Water resource plans - none currently, see: https://www.dnrm.qld.gov.au/water/catchments-planning/catchments	None planned
Constructed Storages Hydromodifier	P17	Queensland Wetland Data Version 3	Queensland Wetland Data Version 4.0 – Wetland areas 2013	Version 5 (2017 extent) to be completed and released early-mid 2018	4 years
Connectivity Scores	S13	Source data: Australian Connectivity Index V1 2014, Queensland Wetland Data Version 3 (Australian Department of the Environment, 2014)	Connectivity Index: same; Wetland Data V4	None currently planned	None planned
Plant Pest Lists	P7, P8	March 2014		From July 2016 the Biosecurity Act 2014 will apply, this redefines categories for 'declared' plants and probably adds/removes some listings.	Irregular

*Queensland Department of Environment and Heritage
**Queensland Department of Science, Information Technology and Innovation
†Queensland Department of Natural Resources and Mines

Methods are needed to check indicator scores, either through ground-truthing or using desktop imagery. Otherwise, irregular and non-concordant data updates will increase the variance of the

difference between times and reduce the power of the WFAT–M, especially its pressure index, to detect change or trend. This is an additional source of extraneous variance in WFAT–M scores that would be included in the error or index term in equation 1.

Those desktop indicators that are assessed across a wetland’s 200 m buffer have ground-truthing in their standard operating procedures. Others, such as indicators based on the land use database, can be checked using aerial imagery or other datasets. Ways to reduce this source of variability in the WFAT–M data are still under investigation.

5 References

Australian Department of the Environment 2014 Australian Connectivity Index Version 1 2014, retrieved from http://www.environment.gov.au/fed/catalog/search/resource/details.page?uuid=%7B57f44b3f-ef00-4828-92b3-807353d4b4b2%7D

Clayton, P.D, Fielder, D.P., Howell, S. and Hill, C.J. 2006 *Aquatic Biodiversity Assessment and Mapping Method (AquaBAMM): a conservation values assessment tool for wetlands with trial application in the Burnett River catchment*, Environmental Protection Agency, Brisbane.

De Winter, J.C.F, Dodou, D. and Weirunga, P.A. 2009 Exploratory factor analysis with small sample sizes, *Multivariate Behavioural Research*, 44, 147–181.

DSITI 2015 *A landscape hazard assessment for wetlands in the GBR catchment*, Department of Science, Information Technology and Innovation Queensland Government, Brisbane.

Hallgren, K.A. 2012 Computing inter-rater reliability for observational data: an overview and tutorial, *Tutorials for Quantitative Methods in Psychology*, 8, 23–34.

Johnson, D.R. 2008 *Using weights in the analysis of survey data* Population Research Institute, PennState.

Kentula, M. E. 2007 Monitoring wetlands at the watershed scale. *Wetlands* 27, 412–15.

Kolenikov, S. 2016 Is it a post-stratification or a non-response adjustment? *Survey Practice* 9: http://www.surveypractice.org/index.php/SurveyPractice/article/view/315/html_61 Accessed 12-09-2016.

Larsen, D.P. and Christie, S.J. 1991 A framework for evaluating the sensitivity of the EMAP design, Chapter 4 in *EMAP surface waters 1991 pilot report*, United States Environmental Protection Agency. Office of Research and Development.

Larsen, D.P., Urquhart, N.S. and Kugler, D.L. 1995 Regional scale trend monitoring of indicators of trophic conditions in lakes *Water Resources Bulletin*, 31, 117–140.

Larsen, D.P., Kincaid, T.M., Jacobs, S.E. and Urquhart, N.S. 2001 Designs for evaluating local and regional scale trends, Bioscience, 51, 1069–1078.

Little, R. and Vartivarian, S.L. 2004 Does weighting for nonresponse increase the variance of survey means? *The University of Michigan Department of Biostatistics Working Paper Series*, 35, University of Michigan School of Public Health.

Norman, G. 2010 Likert scales, levels of measurement and the “laws” of statistics *Advances in Health Science Education* 15, 625–632.

OECD 2008 *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing.

Paddock to Reef Program 2015 Great Barrier Reef Report Card 2014: results 63–77. http://www.reefplan.qld.gov.au/measuring-success/report-cards/2014/ Accessed 17-03-2017.

Queensland Department of Environment and Heritage Protection 2015 *Queensland Wetland Data Version 4.0 – Wetland Areas of Queensland*, Queensland Herbarium, DSITI.

Queensland Department of Natural Resources and Mines 2016 *30–50 km^2 Subcatchments 2014,* Paddock to Reef Catchment Modelling Team, DNRM.

Queensland Department of Natural Resources and Mines 2016 *Fine Sediment Generation Data 2016,* Paddock to Reef Catchment Modelling Team, DNRM.

Queensland Department of Science, Information Technology and Innovation 2016 *Land use mapping Queensland*, Queensland Land Use Mapping Program, DSITI.

Queensland Department of Science, Information Technology and Innovation 2016 *Queensland 2014 Foliage Projective Cover Wooded Vegetation*. Remote Sensing Centre, DSITI.

Queensland Department of Science, Information Technology and Innovation 2016 *2015 Remnant Regional Ecosystems of Queensland, Version 10.0*. Queensland Herbarium, DSITI.

Queensland Department of Science, Information Technology and Innovation 2016 *Biodiversity Status of Pre-clearing Regional Ecosystems of Queensland, Version 10.0*. Queensland Herbarium, DSITI.

Queensland Department of Science, Information Technology and Innovation 2014 *Specified Surface Water Management Areas and Regulated River Reaches in Great Barrier Reef Catchments*. Wetland Science, DSITI.

Robinson, W. and Lee, J. 2009 *AQUABAMM: sensitivity analysis using the Burnett River catchment results*, University of the Sunshine Coast.

Stein, E.D., Fetscher, A.E., Clark, R.P., Wiskind, A., Grenier, J.L., Sutula, M., Collins, J.N. and Grosso, C. 2009 Validation of a wetland rapid assessment method: use of EPA's level 1-2-3 framework for method testing and refinement *Wetlands*, 29, 648–665.

Tilden, J., Borschmann, G., Walsh, C., Mayger, B. and Vandergragt, M. 2015 *Great Barrier Reef catchments wetland monitoring pilot study: assessment methods and monitoring design*, Department of Science Information Technology and Innovation, Brisbane, Queensland.

Thomaidou, F., Valavanioti, E. and Vassileiadis, M. 2013 Weighting of responses in the consumer survey: effects on variance and tracking performance of the Consumer Confidence Indicator, Task Force on the Quality of BCS Data, Foundation for Economic and Industrial Research, Athens.

Urquhart, N.S., Paulsen, S.G. and Larsen, D.P. 1998 Monitoring for policy-relevant trends over time, *Ecological Applications*, 8, 246–257.

Urquhart, N.S. 1999 and Kincaid, T.M. 1999 Designs for detecting trend from repeated surveys of ecological resources, Journal of Agricultural, Biological and Environmental Statistics, 4, 404–414.

6 Appendix A Calculating and combining post-stratification and non-response weights

A probability sample is defined by each sample member having a known probability of being chosen. To simplify the analysis, it was decided to give all wetlands in the GBR-wide wetland monitoring population an equal probability of being included.

The GRTS rule for replacing a non-responding sample member from an oversample is to go to the next one in oversample in order to maintain the spatially balanced random sample; however this proved to be awkward because of the regional basis of organising wetland permissions and visits.

To circumvent this problem, the wetland science team reasoned as follows:

(1) For the time being, maintain the order in which the wetlands were originally listed and divide the GRTS generated list of wetlands into 5 panels of 20, with the remainder constituting the oversample. Each panel is then, in its own right, a spatially balanced random sample with each wetland in the sub-population having an equal probability of inclusion.
(2) For each panel, count the number of wetlands in each region. These become the numbers that we aim to reproduce in that panel in the final sample.
(3) Then sort the original list by region.
(4) From each region, select the numbers (n) of wetlands needed to make up panels in the original regional configuration, going down the regional list in order.
(5) When wetlands are non-responsive, replace them with the next wetland *from the same region*.
(6) This method should preserve the original equal probability, spatially balanced, random nature of the sample *provided that wetlands are non-responsive completely at random*.

However, wetlands were not non-responsive completely at random. In some regions, landholders were less likely to give a ‘yes’ response than in others. An indicative set of numbers to illustrate this issue would be something like this: say the landholders of 20 wetlands had to be approached to get 20 wetlands for assessment in Cape York. Then, in the Fitzroy, the landholders of 40 wetlands would need to be approached to get 20 wetlands for the sample and in the Wet Tropics, 80 to get 20. Wetlands in Cape York were four times as likely as those in the Wet Tropics to have owners who say ‘yes’ to assessments, therefore, proportionally fewer wetlands had their owners approached.

Ignoring the non-response bias for now, owners of a wetland in the wet tropics are twice as likely to be approached as owners of a wetland in the Fitzroy, who are twice as likely to be approached as those in Cape York. So *if there was no non-response bias*, the probability of inclusion of any wetland in the Wet Tropics would be twice the probability of inclusion of one in the Fitzroy and so on. (The next step will deal with the effect of the non-response bias.)

To correct for the fact that we now have different probabilities of inclusion by region, post-stratification weights must be added to each wetland, by region, according to the likelihood that a wetland would have ended up in the sample *if there had been no non-response bias.* The aim of doing this is to restore, in the sample of wetlands whose owners were approached for an assessment, the characteristics of the *population*.

The more likely the wetland will end up in the sample, the lower its weight will be to compensate. For example, because we are approaching more wetland owners, proportionally speaking, in the

Wet Tropics, wetlands from here will get lower weights to compensate for this distortion away from equal probabilities. Table 1 shows a worked example of the method of calculating the post-stratification weights for region.

Table 1 Post-stratification weights for regions

Region	Population proportion	Sample proportion (of wetlands whose owners were approached)	Weight = pop prop/sample prop
CY	.24	.14	**1.71**
WT	.1	.2	**.5**
BDT	.17	.18	**.94**
MW	.02	.08	**.25**
F	.30	.22	**1.36**
BM	.17	.18	**.94**
Total	1	1	

Then something must be done to correct for the non-response bias.

There are several options, as outlined above in Section 3.1.6 but suppose we decide to use the method of weights. We now apply non-response weights to the sample by land use intensity (assuming this is the key factor). The aim with this second set of weights, is to restore, in the *responding sample*, the characteristics of the *approached sample*. [13] Table 2 shows a worked example.

Table 2 Non-response weights for land use intensity

Land use intensity	Approached sample proportion	Responding sample proportion	Weight = approached/responding
High	.6	.1	**6**
Medium	.3	.7	**.43**
Low	.1	.2	**.5**
Total	1	1	

Since we can only apply one weight to the data, we need a method to combine these. The most straightforward way to do this is to cross-tabulate the two variables, in this case region and land use intensity. The expected proportions in each cell are the proportions of wetlands in the regional populations that fall into the different land use intensity categories in the approached sample. The

[13] See Kolenikov (2016) for an explanation of the different types of weights and Johnson (2008) for the method of calculating weights.

observed proportions are those in the responding sample. We calculate weights using these proportions: expected/observed and use these in the data analysis.

The resulting table combining the post-stratification region weights and the land use intensity non-response weights reproduces the interaction between these two variables, allowing it to be taken into account in the resulting weights.

References

Johnson, D.R. 2008 *Using weights in the analysis of survey data* Population Research Institute, PennState

Kolenikov, S. 2016 Is it a post-stratification or a non-response adjustment? *Survey Practice* 9: http://www.surveypractice.org/index.php/SurveyPractice/article/view/315/html_61