Burdekin Water Quality Metric Development

A Healthy Waters Partnership Analysis

Adam Shand

September 30, 2024

Table of Contents

# Introduction

The Burdekin region is vast, with dispersed communities, resources, and data. The successful completion of the Burdekin Expansion project into the Burdekin region relies on an accurate understanding of the time and effort required to develop appropriate data analysis methods, obtain and management data sources, conduct the analysis, and report the results. Not least of which includes all types of freshwater and estuarine water quality.

This document is a case study into the estimated resources required to develop and maintain an effective water quality data analysis method in the entire Burdekin region. This case study uses two test sites:

* Barratta Creek at Northcote,
* and Burdekin River at Sellheim,

and works through each step required in a formal technical report. The case study also explores challenges and difficulties experienced during the process.

# Method

The method that was selected for use in the Burdekin Expansion water quality data analysis is based on the Wet Tropics Waterways Partnerships’ current water quality analysis. In contrast to the Healthy Water Partnership’s water quality analysis method, this water quality method utilises high flow and low flow water quality objectives. The decision to use this method was based on the high and low flow events that characterise the Burdekin region and the availability of high and low flow water quality objectives.

A general overview of the method selected is as follows:

1. Obtain water quality data (e.g. Total Nitrogen)
2. Obtain water quality objectives (both high and low flow)
3. Obtain watercourse flow data (hourly)
4. Obtain high/low flow cut off values (e.g. ‘X’ cumecs)
5. Clean datasets
6. Characterise the number of days per year that were high or low flow
7. Calculate water quality scaling factors based on all available historical data
8. Assign each water sample as high or low flow based on the closest hourly flow value
9. Perform Quality Control and Quality Assurance checks
10. Score each water quality sample against the appropriate water quality objective and scaling factor
11. Weight the high and low flow scores based on the number of high and low flow days
12. Calculate indicator category scores by averaging appropriate indicators
13. Calculate index scores by averaging appropriate indicator categories

## Obtaining Water Quality Data

Currently there is only one viable source of water quality data identified in the Burdekin region, the Catchment Loads Monitoring Program (CLMP). This is due to the requirement that the data is a) long-term, and b) regular (not event based). This requirement unfortunately eliminates many project-based water quality datasets that may be centered around flood monitoring or may only last for a few years.

CLMP data is accessed through “Tahbil”, an online publicly accessible data access portal. The two sites used for this project were both sourced from Tahbil:

* Barratta Creek at Northcote
* Burdekin River at Sellheim

Although the online portal does currently offer 32 total sample sites, not all of these sites are viable for the analysis. Screenshots of the Tahbil site are presented below in ([Figure 1](#fig-tahbil-1)) and ([Figure 2](#fig-tahbil-2)).

|  |
| --- |
| Figure 1: A screenshot of the Tahbil portal interface. Note the large number of options for location (left), and site (middle). This screenshot also demonstrates a severe lack of water quality data in the northern, western, and southern reaches of the Burdekin region. |

|  |
| --- |
| Figure 2: A zoomed in screenshot of the Tahbil portal interface. Note the large number of available sites, and extensive results. |

## Obtain Water Quality Objectives

Water quality objectives are sourced from the Environmental Protection (Water and Wetland Biodiversity) Policy 2019 documents. To find the specific values:

* head to the [EPP Portal](https://environment.desi.qld.gov.au/management/water/policy)
* Navigate to “Burdekin, Haughton and Don basins”
* For barratta objectives:
  + open “Haughton River Basin” (Schedule 1 - Document column)
  + find “Barratta Creek upper catchment fresh waters”
* For sellheim objectives:
  + open Burdekin River (upper) Sub-Basins” (Schedule 1 - Document column)
  + find “Burdekin River (above Dam) sub-catchment waters”, “MD”

Values found in these documents are then manually transcribed into an excel spreadsheet. Screenshots of the tables are presented below in ([Figure 3](#fig-epp-1)) and ([Figure 4](#fig-epp-2)).

|  |
| --- |
| Figure 3: A screenshot of the Environmental Protection Policy water quality objectives for the Barratta Creek Catchment. Note the units shown for each indicator, and the instructions that ‘for single value WQO’s, medians of test data should be less than or equal to the WQO’. |

|  |
| --- |
| Figure 4: A screenshot of the Environmental Protection Policy water quality objectives for the Burdekin River above Dam sub-catchment. Note the units shown for each indicator, and the instructions that ‘for single value WQO’s, medians of test data should be less than or equal to the WQO’. |

## Obtain Watercourse Flow Values

Watercourse flow values are obtained from the Water Monitoring Information Portal (WMIP). Much like the CLMP data, the WMIP data is provided across a range of sites. However, some locations are not directly shared between the two programs. This presents challenges when matching the CLMP data to the best WMIP data option. Maps demonstrating the variation between the programs are presented below in ([Figure 5](#fig-map-1)) and ([Figure 6](#fig-map-2)).

|  |
| --- |
| Figure 5: The location of all Burdekin and Haughton Basin CLMP (red) and WMIP (green) monitoring locations. Note the concentration of CLMP sites in the eastern part of the basin, compared to the expansive network of WMIP sites. |

|  |
| --- |
| Figure 6: A zoomed in map of the same CLMP and WMIP sites. Note that even sites in close proximity may not share the exact same watercourse. |

## Obtain High Flow and Low Flow Cutoff Values

High flow and low flow cut off values are also obtained from the Environmental Protection (Water and Wetland Biodiversity) Policy 2019 documents. WQOs are provided under high and low flow conditions already, and the criteria defining what is considered high or low flow at the specific site is also defined.

In the case of the test sites, the values are as follows:

* Barratta Creek at Northcote: 3.6m3/s
* and Burdekin River at Sellheim: 175.9m3/s

Any flow reading above the value is classified as “high”, and anything below is classified as “low”.

## Clean Data

Once data is obtained it must then be cleaned, checked, and edited to be usable in the analysis. This includes selecting the columns of interest, removing rows of metadata embedded within the file, updating column names, changing the format of time values, adding contextual information, standarising units, and more. Before and after tables demonstrating some of these initial changes are presented below in ([Table 1](#tbl-before)) and ([Table 2](#tbl-after)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: The barratta flow values table before any data cleaning has been applied.   |  |  |  |  | | --- | --- | --- | --- | | **Time** | **X119101A** | **X3** | **X4** | | and | 140 |  |  | | Date |  |  |  | | Date and time |  |  |  | | 12:07:00 09/10/1974 | 0 | 9 |  | | 08:00:00 07/11/1974 | 0 | 9 |  | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: The barratta flow values table after data cleaning has been applied.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Site** | **DateTime** | **Flow** | **FinancialYear** | **Date** | | Barratta | 1974-10-09 12:07:00 | 0 | 1975 | 1974-10-09 | | Barratta | 1974-11-07 08:00:00 | 0 | 1975 | 1974-11-07 | | Barratta | 1974-12-04 08:00:00 | 0 | 1975 | 1974-12-04 | | Barratta | 1974-12-31 04:23:00 | 0 | 1975 | 1974-12-31 | | Barratta | 1974-12-31 05:22:00 | 0.059 | 1975 | 1974-12-31 | |

Data cleaning must also occur between datasets and within specific columns, due to the data being sourced from multiple providers.

* Flow column names are: Site, DateTime, Flow, FinancialYear, Date
* Metadata column names are: Site, FlowType, FlowCutSign, FlowCutValue, FlowUnit, Indicator, IndicatorUnit, Objective
* Water quality column names are: Region, Basin, SubBasin, SiteName, Latitude, Longitude, DateTime, Date, FinancialYear, Analyte, Value, Unit
* Metadata indicator names are: Amm N, Oxid N, Total N, FRP, TSS
* Water quality indicator names are: Ammonium nitrogen as N, Filterable Reactive phosphorus as P, Oxidised nitrogen as N, Total nitrogen as N, Total suspended solids

As well as for the specific units for each indicator, as presented in [Table 3](#tbl-unit-1).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Each indicator and its associated units as recorded in the water quality dataset sourced from the Tahbil database, and the metadata dataset as transcribed from the EPP documents.   |  |  |  |  | | --- | --- | --- | --- | | **Metadata Indicator** | **Metadata Unit** | **Water Quality Indicator** | **Water Quality Unit** | | Amm N | ug/L | Amm N | mg/L | | Oxid N | ug/L | FRP | mg/L | | Total N | ug/L | Oxid N | mg/L | | TSS | mg/L | TSS | mg/L | | FRP | ug/L | Total N | mg/L | |

## Characterise Daily Flow

According to the Wet Tropics Waterways methodology, daily flow is calculated by comparing mean daily discharge at each monitoring site and applying the mean daily base-flow (MDBF) cut off value (Orr et al. 2014), or an estimated cut off value. It is generally expected that the majority of days are classified as “low flow” ([Table 4](#tbl-flow-count-1)) ([Figure 7](#fig-flow-daily-1)), however within any “low flow” day, the hourly flow could be classified as “high flow” as long as the daily mean was less than the cut off value ([Figure 8](#fig-flow-daily-2)). The importance of this distinction is highlighted under [Scaling Factors](@sec-scaling-factors).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: The number of high flow and low flow days recorded at each of the testing sites. It is expected that the majority of days are “low flow”.   |  |  |  | | --- | --- | --- | | **Site** | **FlowType** | **n** | | Barratta | High | 1822 | | Barratta | Low | 12779 | | Sellheim | High | 2156 | | Sellheim | Low | 16188 | |

|  |
| --- |
| Figure 7: A plot of high flow (red) and low flow (blue) days recorded at each of the testing sites. Note that some values |

|  |
| --- |
| Figure 8: A plot of a single day of flow values. Note that within a day flow could be both high flow (red) and low flow (blue) days recorded at each of the testing sites. Note that some values |

## Calculate Scaling Factors

Water quality data are scored by comparing the median of the measured values of the indicator against the WQO for the indicator. A scaling factor is then used to standardise the result of this comparison. As per the Wet Tropics Waterways methodology, scaling factors are calculated as follows:

*“The historical GBR CLMP data were pooled from all basins (seven sites). The data were separated into high-flow and base-flow periods using an approximation method, where any ‘eventflow’ data (indicated by consecutive samples within a single day or over consecutive days) represented samples taken above the event-flow threshold, and that conversely, any discrete ‘ambient’ samples (approximately monthly) were taken below the event-flow threshold (and therefore represented base-flow). The 90th percentile was set as the SF and was calculated for each data set (Table 22).”*

Thus, the specific timing of the water quality sample relatively to the nearest hour of flow data is of great significance. As shown in [Figure 9](#fig-flow-daily-3) should the sample on January 19th occurred earlier it may have been taken during “high flow” conditions.

|  |
| --- |
| Figure 9: A plot of a single day of flow values. Note that within a day flow could be both high flow (red) and low flow (blue) days recorded at each of the testing sites. Note that some values |

Once calculated the 90th percentile can then be added to the metadata table ([Table 5](#tbl-metdata-1)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 5: A exert (Barratta only) of the metadata required to conduct the water quality analysis.   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | **Site** | **FlowType** | **FlowValue** | **FlowUnit** | **Indicator** | **IndicatorUnit** | **Objective** | **SF** | | Barratta | Low | 3.6 | m3/s | Amm N | ug/L | 18 | 87.2 | | Barratta | Low | 3.6 | m3/s | Oxid N | ug/L | 85 | 1.58e+03 | | Barratta | Low | 3.6 | m3/s | Total N | ug/L | 590 | 2.88e+03 | | Barratta | Low | 3.6 | m3/s | FRP | ug/L | 25 | 86 | | Barratta | Low | 3.6 | m3/s | TSS | mg/L | 10 | 43.2 | | Barratta | High | 3.6 | m3/s | Amm N | ug/L | 10 | 79.8 | | Barratta | High | 3.6 | m3/s | Oxid N | ug/L | 45 | 2.25e+03 | | Barratta | High | 3.6 | m3/s | Total N | ug/L | 760 | 4.29e+03 | | Barratta | High | 3.6 | m3/s | FRP | ug/L | 40 | 109 | | Barratta | High | 3.6 | m3/s | TSS | mg/L | 40 | 239 | |

## Assign Water Quality Samples as High or Low Flow

Due to the method used to calculate the scaling factors, the water quality samples have already been assigned the appropriate high or low flow conditions. No further work is needed to complete this step.

## QA/QC Checks

Once data has been provided its appropriate context and metadata, it must undergo stringent Quality Control and Quality Assurance checks.

|  |
| --- |
| Note |
| For the purposes of this demonstration the QA/QC steps will be skipped. These steps are identical to those already implemented in the current Healthy Waters Partnership’s water quality analysis method. |

## Score Water Quality Data

The scoring methodology of water quality data is as follows:

1. Split all water quality data into high flow or low flow based on the nearest hourly flow reading
2. Calculate mean daily values for both high and low flow groups (sometimes several samples are taken per site per day)
3. Pool mean daily values from all sites within a watercourse together
4. For each watercourse, calculate monthly medians, per indicator per high and low flow.
5. For each watercourse, calculate annual medians from monthly medians, per indicator per flow.
6. For each watercourse, use the annual medians, WQOs, 80th (or 20th for Low DO) percentiles, and SFs to calculate the standardised condition score (0-100) per indicator per flow ([Figure 10](#fig-wq-score-1), [Figure 11](#fig-wq-score-2))
7. Weight the score for high flow and the score for low flow according to the ratio of high flow and low flow days
8. Sum the weighted scores together
9. For each watercourse, calculate indicator category scores by averaging indicator scores. E.g.:
   * (TP + DIN)/2 = Nutrients
   * (Turbidity + DO)/2 = Physical Chemical Properties
10. For each indicator and indicator category, group watercourse scores by sub basin and average to calculate sub basin indicator and indicator category scores.
11. For each indicator and indicator category, group watercourse scores by basin and average to calculate basin indicator and indicator category scores.
12. For each basin, average indicator category scores to calculate the water quality index score.

|  |
| --- |
| Figure 10: The rules, formulas and scoring range for indicators in the freshwater basins and estuarines for the report card. |

|  |
| --- |
| Figure 11: The middle point represents the annual median, the top whisker the 80th percentile and the bottom whisker the 20th percentile. If the median complies with WQOs, the score will be within either the “Good” or “Very Good” ranges, if the median does not comply it will be in the “Moderate”, “Poor” or “Very Poor” ranges. Medians that do not meet the WQOs are scaled between the WQOs using a scaling factor (SF) nominally defined as the 90th (or 10th) percentile of the historic water quality data. |

Below are some examples throughout the water quality data analysis method.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 6: A exert (Barratta only) of the monthly medians calculated based on high flow and low flow data.   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Site** | **DateTime** | **Indicator** | **IndicatorValue** | **FlowValue** | **FlowType** | **MonthlyMedian** | | Barratta | 2023-02-02 23:27:00 | Amm N | 65.5 | 13.8 | High | 19 | | Barratta | 2023-02-04 00:56:40 | Amm N | 34.3 | 60.8 | High | 19 | | Barratta | 2023-02-05 08:59:00 | Amm N | 57 | 189 | High | 19 | | Barratta | 2023-02-06 02:45:00 | Amm N | 17.5 | 291 | High | 19 | | Barratta | 2023-02-06 22:25:00 | Amm N | 3 | 358 | High | 19 | | Barratta | 2023-02-07 18:37:00 | Amm N | 16 | 148 | High | 19 | | Barratta | 2023-02-08 17:32:30 | Amm N | 14.5 | 75 | High | 19 | | Barratta | 2023-02-10 01:50:00 | Amm N | 12 | 26.4 | High | 19 | | Barratta | 2023-02-16 21:55:00 | Amm N | 24 | 2.37 | Low | 31.5 | | Barratta | 2023-02-26 20:15:00 | Amm N | 39 | 1.43 | Low | 31.5 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 7: A exert (Barratta only) of the annual medians calculated based on high flow and low flow data.   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Site** | **Month** | **FlowType** | **MonthlyMedian** | **AnnualMedian** | | Barratta | 7 | Low | 38.5 | 32.2 | | Barratta | 8 | Low | 32.5 | 32.2 | | Barratta | 9 | Low | 58 | 32.2 | | Barratta | 10 | Low | 64.5 | 32.2 | | Barratta | 11 | Low | 47 | 32.2 | | Barratta | 12 | Low | 32.2 | 32.2 | | Barratta | 1 | Low | 31 | 32.2 | | Barratta | 2 | Low | 31.5 | 32.2 | | Barratta | 3 | Low | 24 | 32.2 | | Barratta | 4 | Low | 23 | 32.2 | | Barratta | 5 | Low | 22 | 32.2 | | Barratta | 6 | Low | 26.5 | 32.2 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 8: A exert of the water quality scores calculated based on high flow and low flow data.   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Site** | **FlowType** | **FinancialYear** | **Indicator** | **Score** | **Objective** | **ScalingFactor** | | Barratta | High | 2023 | Amm N | 48.6 | 10 | 79.8 | | Barratta | Low | 2023 | Amm N | 48.4 | 18 | 87.2 | | Sellheim | Low | 2023 | Amm N | 90 | 10 | 13.4 | | Sellheim | High | 2023 | Amm N | 90 | 10 | 20 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 9: A exert of the flow ratio data based on financial year.   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | **Site** | **FinancialYear** | **FlowType** | **Count** | **Total** | **Ratio** | | Barratta | 2023 | High | 40 | 365 | 0.11 | | Barratta | 2023 | Low | 325 | 365 | 0.89 | | Sellheim | 2023 | High | 56 | 365 | 0.15 | | Sellheim | 2023 | Low | 309 | 365 | 0.85 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 10: A exert of the weighted water quality scores calculated based on high flow and low flow data.   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | **Site** | **FlowType** | **FinancialYear** | **Indicator** | **Score** | **Ratio** | **WeightedScore** | **FinalScore** | | Barratta | High | 2023 | Amm N | 48.6 | 0.11 | 5.34 | 48.4 | | Barratta | Low | 2023 | Amm N | 48.4 | 0.89 | 43 | 48.4 | | Sellheim | Low | 2023 | Amm N | 90 | 0.85 | 76.5 | 90 | | Sellheim | High | 2023 | Amm N | 90 | 0.15 | 13.5 | 90 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 11: A exert of the weighted water quality scores with colours assigned based on their score.   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | **Site** | **FinancialYear** | **Amm N** | **FRP** | **Oxid N** | **TSS** | **Total N** | | Barratta | 2021 | 49 | 32.1 | 59 | 53.5 | 54.2 | | Barratta | 2022 | 49.3 | 35.3 | 59.3 | 51.6 | 56.9 | | Barratta | 2023 | 48.4 | 38.8 | 58.6 | 52.8 | 53.6 | | Sellheim | 2021 | 90 | 81.1 | 62.6 | 58.4 | 48.8 | | Sellheim | 2022 | 90 | 78 | 65.6 | 60.6 | 52.7 | | Sellheim | 2023 | 90 | 81.9 | 65.2 | 59.2 | 53.2 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 12: A exert of the weighted water quality scores with colours assigned based on their score.   |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | **Region** | **Basin** | **SubBasin** | **Site** | **FinancialYear** | **AmmN** | **Frp** | **OxidN** | **Tss** | **TotalN** | **Nutrients** | **Phys\_Chem** | **Overall\_WQ** | | Burdekin | Haughton | Barratta Creek | Barratta | 2021 | 49 | 32.1 | 59 | 53.5 | 54.2 | 48.6 | 53.5 | 51.1 | | Burdekin | Haughton | Barratta Creek | Barratta | 2022 | 49.3 | 35.3 | 59.3 | 51.6 | 56.9 | 50.2 | 51.6 | 50.9 | | Burdekin | Haughton | Barratta Creek | Barratta | 2023 | 48.4 | 38.8 | 58.6 | 52.8 | 53.6 | 49.8 | 52.8 | 51.3 | | Burdekin | Burdekin | Lower Burdekin River | Sellheim | 2021 | 90 | 81.1 | 62.6 | 58.4 | 48.8 | 70.6 | 58.4 | 64.5 | | Burdekin | Burdekin | Lower Burdekin River | Sellheim | 2022 | 90 | 78 | 65.6 | 60.6 | 52.7 | 71.6 | 60.6 | 66.1 | | Burdekin | Burdekin | Lower Burdekin River | Sellheim | 2023 | 90 | 81.9 | 65.2 | 59.2 | 53.2 | 72.6 | 59.2 | 65.9 | |

# Discussion

There are several key findings to discuss or address as a result of this exploratory analysis, these may be additional features yet to be included, issues identified with the analysis so far, or the level of effort required to implement changes. Each topic is addressed below.

## Timing

One of the mains goals outlined by this report was to determine an estimated effort/time required to develop and implement a fully functioning water quality analysis methodology that is tailored specifically to the Burdekin Region. This methodology would be robust, self-sufficient, and based on practices proven by similar methods implemented in surrounding regions such as Townsville, Mackay, and the Wet Tropics. A detailed log of the time and effort spent to develop each aspect of this report has been documented below, along with further extrapolation of the time required to develop the completed water quality analysis.

### Data

*Hours taken: 10h*  
*Additional hours for a complete analysis: 10h*

All required data used in this report had to be identified and accessed. This includes the water quality data (CLMP), flow data (WMIP), water quality objectives (EPP), high flow/low flow cut off values (EPP), and scaling factors (90th percentiles).

* Several hours were spent reviewing each data source and written documentation regarding access to the data.
* Limited time was spent actually “reading in” the data to be used in the analysis method.
* Additional time would be required to identify the appropriate flow data site (WMIP) for each water quality site (CLMP).
  + This is a manual process that scales linearly with the number of sites used in the analysis.
  + No time efficiency benefits can be gained due to “scripts” or “automation”.
  + Time is saved due to being familiar with the process as a result of this report.

### Translate Existing Methods

*Hours taken: 20h*  
*Additional hours for a complete analysis: 10h*

Some aspects of the current Healthy Waters Partnership Method could be directly translated and implemented into this report. This included methods such as the scoring function, data “read in” protocols, and dataframe management such as cleaning.

* The major of time spent on this aspect was directed at updating existing processes or adding minor changes. This is due to improvements in author understanding, and small discrepancies in existing and new data. Examples include:
  + Introducing forced, standardised column naming (rather than manually adhering to rules)
  + Changing indicator names (e.g. previously analysed PP, now analyzing Ammonium)
  + Writing code specifically tailored to this report (behind-the-scenes edits such as code folding)
* Additional time would be required to expand the method to the entire Burdekin Region.
  + If additional indicators, units, or sites are added the appropriate error checks needs to be conducted.
  + This component benefits from scripts/automation and would take significantly less time to expand.

### Write New Methods

*Hours taken: 30h*  
*Additional hours for a complete analysis: 10h*

Some aspects of the proposed Burdekin water quality analysis method have not been implemented in the current Healthy Waters Partnership report. This includes steps such as separating water quality samples into high flow and low flow observations, scoring new indicators, and aggregating high flow and low flow scores.

* Although several of these steps have been implemented by other reporting groups (e.g. Wet Tropics), these implementations are either manual (Wet Tropics) or in-sufficiently detailed to be directly applicable to this analysis.
* The majority of time spent on this aspect involved creating logical and documented methods for utilising high flow and low flow water quality observations, as well as creating informative material to educate new readers.
* Additional time would be required to expand the method to the entire Burdekin Region, however this component benefits from scripts/automation and would take significantly less time to expand.

### Develop Novel Reporting Methods

*Hours taken: 30h*  
*Additional hours for a complete analysis: TBA*

The objective of creating this Burdekin water quality analysis exploratory report identified a broader objective of developing a novel reporting method. Generally speaking, when conducting any form of data analysis (such as a water quality analysis), the process takes places within the coding program “R”. Once the analysis (R script) has been completed, all outputs (tables, figures, maps, results) are saved to file. These outputs are then copied from their file location and manually pasted into a final reporting document where introductions, discussions, findings are written (usually a word document).

Unlike the standard workflow described above, the nature of this exploratory report was two fold:

1. Write a “normal” data analysis method, who’s outputs would be included into a report
2. Write an analysis of the actual method, exploring time taken, and explaining processes

To accomplish both of these objectives the data analysis process (occurring with the R script) and the report writing process (previously occurring in Word), have been combined into one document that runs within R, and produces one technical report in the Word file type. Creating the analysis and report in this way saves time when discussing methodological choices, reduces complications regarding constantly updating results as methods are improved, limits the possibility of human error due to copy paste, introduces standardised and reproducible reporting formats, and satisfies the needs of both the data analyst (usually an R specialist), and the end user (most often only familiar with Word).

* The amount of work required for this aspect of the report is highly variable:
  + should the finished product follow standard procedures (the copy paste method), no additional time would be required - however significant time would be lost due to the slow and error prone nature of the copy paste method.
  + should this novel appropriate be implemented (or partially implemented), additional time may exceed 40h - however significant time would be saved once completed.

### Other Aspects

*Hours taken: 0h*  
*Additional hours for a complete analysis: 60h*

There are several steps of a completed Burdekin water quality analysis that were not addressed within this exploratory report. This include steps such as quality assurance and quality control checks, the exploratory data analysis stage, spatial aggregation of sites through to basins, and the visualisation and mapping of results.

Each of these steps are not unique to the Burdekin water quality analysis method, nor are they a cornerstone of completing an exploratory analysis. Thus each of these steps has been left un-addressed. The introduction of each of these steps would require a translation of the existing processes already implemented in the Healthy Waters Partnership’s analysis. An estimate of the time spent can be inferred from the “Translate Existing Methods” section above, and is thus roughly:

## Missing Features

There are a number of missing features yet to be included in the Burdekin water quality analysis. Each of this features, the reason for not including them, and the time required to include them, are covered under “Other Aspects” above.

## Methodology Errors

Throughout the course of the Burdekin water quality analysis exploration several methodology “questions” have been identified, these include:

* Spatial aggregation. The Burdekin region and its basins are extremely large, however the number of samples used to represent each of these basins is very small, or zero. It is unclear as to how the spatial aggregation of these limited samples is to be done, and how the basins score was calculated is to be explained.
* Water quality aggregation. When reporting, “water quality” is provided as a index composed of the “nutrients” and “physical-chemical properties” indicator categories. Each of these indicator categories is traditionally comprised of multiple indicators such as Ammonium, FRP, and TP. If an indicator category only consists of one indicator, it cannot be calculated. In the Burdekin region, the “physical-chemical properties” indicator category is only comprised of TSS and thus should not be calculated - and neither should “water quality”. It is still unclear how this aggregation is to be addressed.
* Reconciling scaling factors. The method used to calculated scaling factors for this report was to use the 90th percentile of all available historical data. This is in contrast to the method currently used by the Healthy Waters Partnership, which is to use values set by expert opinion. Although it is not incorrect to use two different methods, it is unclear as to which method is objectively better. Further, it is unclear if the scaling factor should be updated each year to include the latest available information - which would cause the shifting baseline problem.