



REPORT NO. 3684A

**DISSOLVED OXYGEN AND ECOSYSTEM  
METABOLISM MEASUREMENT IN NORTHLAND  
RIVERS**

**World-class science  
for a better future.**



# DISSOLVED OXYGEN AND ECOSYSTEM METABOLISM MEASUREMENT IN NORTHLAND RIVERS

ERIC GOODWIN, ROGER YOUNG

Prepared for Northland Regional Council  
Envirolink Project 2135-NLRC225

CAWTHRON INSTITUTE  
98 Halifax Street East, Nelson 7010 | Private Bag 2, Nelson 7042 | New Zealand  
Ph. +64 3 548 2319 | Fax. +64 3 546 9464  
[www.cawthron.org.nz](http://www.cawthron.org.nz)

REVIEWED BY:  
Paula Casanovas



APPROVED FOR RELEASE BY:  
Joanne Clapcott



---

ISSUE DATE: 25 May 2022

RECOMMENDED CITATION: Goodwin E, Young RG 2022. Dissolved oxygen and ecosystem metabolism measurement in Northland rivers. Prepared for Northland Regional Council. Cawthron Report No. 3684A. 36 p. plus appendices.

© COPYRIGHT: This publication may be reproduced in whole or in part without further permission of the Cawthron Institute or the Copyright Holder, which is the party that commissioned the report, provided that the author and the Copyright Holder are properly acknowledged.



## EXECUTIVE SUMMARY

New Zealand needs to maintain and improve the health of our aquatic ecosystems. A healthy ecosystem can be diagnosed by the rates of its ecosystem processes, which will fall into typical ranges. Ecosystem metabolism (the combination of primary production [or carbon fixing through photosynthesis] and ecosystem respiration) is an example of an ecological process indicator and was included as an attribute in the National Policy Statement-Freshwater Management 2020. Northland Regional Council (NRC) sought advice on measurement of stream ecosystem metabolism by undertaking continuous dissolved oxygen (DO) monitoring from Cawthron. This report, which is mainly funded by Envirolink and partly by NRC, provides details on the following:

- development of an R-code (for statistical computing) approach for estimating ecosystem metabolism from continuous dissolved oxygen (DO) records
- results of an analysis of ecosystem metabolism from 6 sites in Northland with continuous DO data, and
- recommendations to help guide future monitoring of DO including good data management practice and measuring ecosystem metabolism in Northland rivers.

As part of this advice, a 2-day workshop (27–28 April 2021) was held in Whangarei with relevant NRC staff to discuss the theoretical concept and techniques behind measuring stream ecosystem metabolism. Discussions covered continuous DO monitoring, data management, and lessons from the review of the raw DO data provided by NRC. The workshop included a field trip to some of NRC's DO monitoring sites and discussion about DO logger placement and monitoring approaches.

Stream metabolism is affected by variations in dissolved oxygen in river water that are controlled by photosynthesis by aquatic plants (increasing DO concentrations), respiration by all aquatic organisms (reducing DO concentrations), and exchange of DO through the water surface. Using continuous DO records, and a measure of mean upstream depth, it is possible to determine the rates of photosynthesis and respiration that are occurring.

We developed R code to extract, summarise and check continuous DO data from Northland database systems and then implemented a night-time regression method to estimate metabolism continuously for each site. Ecosystem metabolism (daily) was successfully estimated using formulae implemented in the R code for all 6 sites for up to 2 years. Annual DO minima were also calculated from the DO datasets. Minimum DO concentrations were at good or satisfactory levels at most of the sites, but were consistently in the C-band at the Mangere site and the minimum 7-day DO was below the national bottom line at Mangere during the 2020 drought (see table, next page).



Table. Minimum annual DO concentrations for the 6 Northland rivers from each of the hydrological years where continuous data were recorded. Colours represent the National Objectives Framework bands from the NPS-FM (2020): A-band (blue), B-band (green), C-band (orange) and D-band (red). Minimum 7-day values are from the whole year, while minimum 1-day values are from the summer period as defined in the NPS-FM (2020), which is 1 November–30 April.

River	Hydrological Year	Minimum 7-Day DO (mg·L <sup>-1</sup> )	Minimum 1-Day DO (mg·L <sup>-1</sup> )
Hatea	2020	7.25	6.95
Hatea	2021	7.7	7.46
Mangere	2019	6.03	5.46
Mangere	2020	4.82	4.02
Mangere	2021	5.07	4.76
Opouteke	2020	6.79	6.57
Opouteke	2021	7.08	6.81
Otaika	2020	6.46	6.17
Otaika	2021	6.66	6.03
Punakitere	2020	8.06	7.3
Punakitere	2021	7.37	6.73
Waitangi	2020	6.67	5.96
Waitangi	2021	6.79	6.31

Rates of gross primary production (GPP, photosynthesis) and respiration (R) were high at the Mangere site and potentially indicative of poor ecosystem health, although the location of the DO logger at this site influenced the reliability of the metabolism estimates, particularly during low flow periods. Rates of GPP and R at the other sites were indicative of excellent to satisfactory ecosystem health.

The National Environmental Monitoring Standard for DO measurement provides good guidance on DO measurement. In order to use the DO data for subsequent metabolism measurement, smart deployment of well-calibrated DO sensors (preferably optical sensors) in the right locations is critical. Some of the main recommendations we made are:

- Sensors/loggers should be deployed continuously to gather information on annual, seasonal and daily fluctuations in DO and ecosystem metabolism. If long-term continuous logging is difficult due to resource constraints, then consider collecting data during the summer/autumn period.
- A regular monthly monitoring site check and instrument clean is recommended, alongside recalibration of the DO sensors, if required.

- Sensors should be calibrated at a fixed altitude (e.g., NRC offices) and later raw data need to be corrected for differences in altitude and barometric pressure between actual site and calibration location.
- Sensor calibration should be done at a temperature and salinity similar to that which will be experienced during the subsequent deployment.
- Sudden step-changes in DO measurements associated with cleaning and maintenance are a signal that careful data processing is required to address calibration issues and sensor drift.
- An estimate of mean depth in the river reach upstream of the DO logger location is required to convert volumetric measurements to area-based measurements, so metabolism estimates can be compared among sites.

The code used to collate and analyse the data and create the plots presented in this report is provided in the Appendices. To avoid multiple versions of the code circulating and evolving independently, we intend to provide a centrally curated version of the code in the future that incorporates any updates. Therefore, if readers are wanting to develop systems for automated metabolism calculation, we recommend contacting the authors of this report first rather than just immediately applying the code provided in the Appendices to this report.





## TABLE OF CONTENTS

1. INTRODUCTION .....	1
2. STREAM METABOLISM ESTIMATES AND DEVELOPING R CODE .....	3
2.1. Processes affecting oxygen concentration in water.....	3
2.1.1. <i>Rate of photosynthesis</i> .....	4
2.1.2. <i>Rate of respiration</i> .....	5
2.1.3. <i>Rate of reaeration</i> .....	5
2.2. Estimating each component – the theory .....	6
2.3. Practical implementation of respiration and reaeration estimation.....	7
2.3.1. <i>Dark period identification</i> .....	7
2.3.2. <i>Temperature correction</i> .....	8
2.4. Estimation of gross primary production and net ecosystem metabolism .....	8
2.4.1. <i>Volumetric to areal conversion</i> .....	9
2.4.2. <i>Reliability and QC of metabolic estimates</i> .....	11
3. INPUT DATASETS .....	13
3.1. Hatea at Whareora Road.....	15
3.2. Mangere at Knights Road.....	16
3.3. Opouteke at Suspension Bridge .....	17
3.4. Otaika at Otaika Valley Road.....	18
3.5. Punakitere at Taheke .....	19
3.6. Waitangi at Waimate North Road .....	20
4. SUMMARISED DISSOLVED OXYGEN AND METABOLISM RESULTS.....	21
4.1. Dissolved oxygen .....	21
4.2. Ecosystem metabolism.....	22
5. RECOMMENDATIONS REGARDING DO MONITORING AND METABOLISM CALCULATION .....	27
5.1. National Environmental Monitoring Standard for dissolved oxygen.....	27
5.2. Monitoring considerations.....	27
5.2.1. <i>Sensor types</i> .....	27
5.2.2. <i>Deployment locations</i> .....	28
5.2.3. <i>Deployment duration</i> .....	28
5.2.4. <i>Measurement frequency</i> .....	28
5.3. Calibration and maintenance.....	28
5.4. Other data required .....	30
5.4.1. <i>Water temperature</i> .....	30
5.4.2. <i>Mean depth upstream</i> .....	30
5.4.3. <i>Light</i> .....	30
5.4.4. <i>Water quality</i> .....	31
5.5. Data processing.....	31
5.6. Metadata .....	32
5.7. Data storage and archiving.....	32
5.8. Curation of code for metabolism calculations .....	33
6. SUMMARY .....	34
7. ACKNOWLEDGEMENTS .....	36

8. REFERENCES .....	36
9. APPENDICES .....	37
APPENDIX 1. R CODE .....	37
APPENDIX 2. RECORDING SHEET AND DEPTH MEASUREMENTS USED TO ESTIMATE MEAN REACH DEPTH UPSTREAM OF EACH DO LOGGER DEPLOYMENT SITE. ....	65
APPENDIX 3. TIME SERIES PLOTS OF DAILY METABOLISM RESULTS.....	66
A3.1 Hatea at Whareora Road.....	66
A3.2 Mangere at Knights Road.....	68
A3.3 Opouteke at Suspension Bridge.....	69
A3.4 Otaika at Otaika Valley Road.....	70
A3.5 Punakitere at Taheke .....	71
A3.6 Waitangi at Waimate North Road .....	72

## LIST OF FIGURES

Figure 1. Typical daily fluctuation in dissolved oxygen. ....	3
Figure 2. A regression of change in oxygen level against oxygen deficit for a night-time period. ....	6
Figure 3. Time series of river depth and derived metabolic properties reaeration, respiration, gross primary production (GPP) and net ecosystem metabolism (NEM). ....	10
Figure 4. Relationship between river stage (water level in mm) and reaeration coefficient k. ....	11
Figure 5. Locations of the monitoring sites where the continuous DO and water temperature data were collected. ....	14
Figure 6. Input data series for Hatea at Whareora Road. ....	15
Figure 7. Input data series for Mangere at Knights Road. ....	16
Figure 8. Input data series for Opouteke at Suspension Bridge. ....	17
Figure 9. Input data series for Otaika at Otaika Valley Road. ....	18
Figure 10. Input data series for Punakitere at Taheke. ....	19
Figure 11. Input data series for Waitangi at Waimate North Road. ....	20
Figure 12. Comparing distributions of metabolic values between rivers. ....	24
Figure 13. Potential banding system for ecosystem metabolism as discussed by the Freshwater Science and Technical Advisory Group 2019 that helped inform the National Policy Statement-Freshwater Management (2020). ....	26

## LIST OF TABLES

Table 1. Minimum annual DO concentrations for the 6 Northland rivers from each of the hydrological years where continuous data were recorded. ....	21
Table 2. Metabolism summary statistics for the 6 river sites in Northland. ....	23
Table 3. Median ecosystem metabolism measurements for the 6 Northland sites. ....	26

## 1. INTRODUCTION

Recent changes to legislation have emphasised the importance of ecosystem health as a value to be maintained and improved in New Zealand freshwater systems. A biophysical framework for ecosystem health was included in the National Policy Statement for Freshwater Management 2020 (NPS-FM 2020) and incorporates five components—water quality, water quantity, habitat, aquatic life and ecological processes. Ecosystem metabolism is an example of an ecological process indicator and was included as an attribute in the NPS-FM 2020. Continuous DO monitoring data are used to calculate ecosystem metabolism. Northland Regional Council (NRC) has been monitoring dissolved oxygen (DO) continuously at 6 sites for the last few years, recognising and anticipating the need for a broader assessment of river and stream ecosystem health in their State of the Environment monitoring programme. NRC sought advice on DO and metabolism measurement from Cawthron via an Envirolink project. The project goals were to:

1. Provide a workshop for NRC staff covering:
  - a. collection of continuous DO data, routine maintenance and quality coding following standard protocols
  - b. monitoring timing, data range and necessity of long-term monitoring
  - c. careful selection of monitoring sites and location (e.g., constraints regarding river/stream habitat types)
  - d. additional field measurements (e.g., upstream habitat types, continuous flow, continuous temperature) required for calculation of ecosystem metabolism or the interpretation of the metabolism parameters
  - e. methodology or user manual for calculating ecosystem metabolism (e.g., spreadsheet model, shiny app, or code for in-house database).
2. Prepare a report on ecosystem metabolism (together with a tool which could be used in-house) using the quality coded data collected and prepared by the council staff.

A 2-day workshop (27–28 April 2021) was held in Whangarei with NRC staff and covered metabolism measurement using continuous DO monitoring, DO data management, quality assurance/quality control processes and lessons from the review of the raw DO data provided to Cawthron by NRC. The workshop included a field trip to four DO monitoring sites and discussion about DO logger placement and monitoring approaches.

This report provides details on the development of an R-code based approach for estimating ecosystem metabolism from continuous DO records, the results of an analysis of ecosystem metabolism from the 6 sites with continuous DO data collected by NRC (with appropriate quality assurance/quality control processes applied to the

raw data by NRC as discussed at the workshop), and recommendations to help guide future monitoring of DO and ecosystem metabolism in Northland.

## 2. STREAM METABOLISM ESTIMATES AND DEVELOPING R CODE

One aspect of stream health is the level of biological activity in it. A healthy stream will support a moderate level of metabolic activity, made up of photosynthetic primary production during the day and ecosystem respiration (occurring day and night). Poor stream health may be reflected either by very low levels or very high levels of metabolic activity. The level of these metabolic activities is reflected in the magnitudes of fluctuation of DO in the water, which occur on a daily basis. Figure 1 shows oxygen rising to a maximum in the mid-afternoon and falling to a minimum in the early hours of the morning. Increases are driven by photosynthesis, and decreases are driven by respiration. However, oxygen fluctuations are not only due to metabolic activity, and the two activities described above need to be isolated to estimate their magnitudes separately. In this section of the report, we describe the rationale and steps required to calculate metabolism which have been implemented in R-code as provided in Appendix 1.

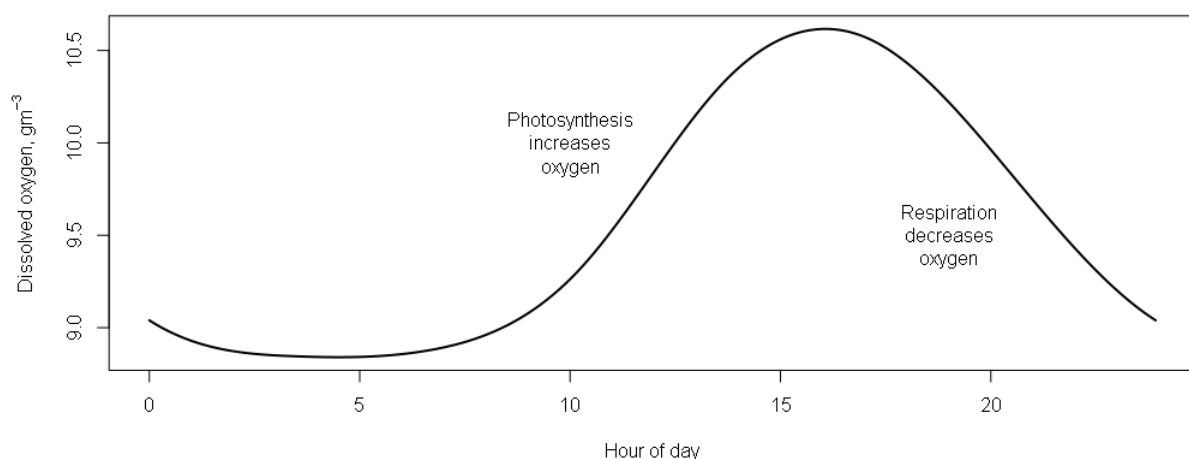


Figure 1. Typical daily fluctuation in dissolved oxygen.

This section describes a method to estimate the separate photosynthetic and respiratory metabolic activity levels from time series of recorded oxygen concentrations in stream water. Other methods are available but require additional data (e.g., time series of photosynthetically active radiation, or light levels).

### 2.1. Processes affecting oxygen concentration in water

Metabolism in stream ecosystems comprises photosynthesis during daylight hours, and respiration occurring at all times. Photosynthesis increases dissolved oxygen in the water and respiration depletes dissolved oxygen from the water. Alongside these

biological processes, there is physical exchange of oxygen gas between stream water and atmospheric air. This non-biological gas exchange can happen in either direction: oxygen can transfer into the water from the atmosphere or out of the water to the atmosphere. The gas transfer at the water surface is regulated by atmospheric pressure and temperature and acts to equilibrate dissolved oxygen with the atmospheric oxygen 'partial pressure'. That is, if the oxygen concentration in the water is below the 'saturation' concentration for the conditions (temperature, total pressure), oxygen will diffuse from the atmosphere into the water (reaeration). If oxygen concentration in water is higher than the saturation concentration for the conditions, oxygen will diffuse from water back to the atmosphere. The saturation concentration is the amount of oxygen water will hold at equilibrium, given the temperature and atmospheric oxygen partial pressure. As oxygen concentration rises and falls (as shown in Figure 1) it may pass above and below the saturation concentration. Sometimes it will be in surplus, sometimes in deficit, relative to saturation concentration.

$$\frac{dDO}{dT} = P + R + kD \quad \text{Equation 1}$$

In Equation 1 the term  $\frac{dDO}{dT}$  means the change in dissolved oxygen over time (units g/m<sup>3</sup>/day). Oxygen levels in the water may rise or fall over time, giving, respectively, positive or negative values of  $\frac{dDO}{dT}$ . P represents the contribution to change in oxygen made by photosynthesis. Because photosynthesis generates oxygen, P should be a positive contribution. R represents the contribution to change in oxygen made by respiration. Because respiration depletes oxygen, R is expected to be negative. The kD term represents the physical gas exchange with the atmosphere through the water surface. This component may be positive or negative, depending on the degree of oxygen deficit (D, units g/m<sup>3</sup>), which may be positive (water undersaturated in oxygen) or negative (water supersaturated in oxygen). k is the reaeration coefficient (units per day  $\equiv$  day<sup>-1</sup>), characteristic of the stream at a certain water level, positive and ideally constant over time at a given level of flow. Shallow turbulent streams have a high k, while deep, slow-flowing rivers will have a low k.

### **2.1.1. Rate of photosynthesis**

The rate of photosynthesis (reflected in oxygen generation) fluctuates over a 24-hour cycle. Its rate is determined by several factors including aquatic plant biomass, light levels, nutrient concentrations, carbon dioxide availability and temperature. Light fluctuates reasonably predictably over a daily cycle, while the other effects may be steadier over longer time periods, or less predictable. Daily fluctuation in the rate of photosynthesis largely follows the fluctuation in light levels. The rate of photosynthesis can be estimated from its effect on the increase in dissolved oxygen during the day,

which is also influenced by the effects of respiration and reaeration. Once the rates of respiration and reaeration are known, the rate of photosynthesis can be determined.

### **2.1.2. Rate of respiration**

The rate of respiration (reflected in oxygen depletion) is independent of light levels, but dependent on the active biomass, nutrient concentrations, oxygen availability and temperature at the site. Respiration is therefore steadier over time, and specifically it can be assumed to be occurring at the same rate during daytime as during the nights preceding and following. The respiration rate can be estimated from its effect on decreases in dissolved oxygen at night when it is confounded only by the effects of reaeration. Once the rate of reaeration is known, the rate of respiration can be determined.

### **2.1.3. Rate of reaeration**

When oxygen is being produced by photosynthesis during the day, the concentration in the water may become higher than the saturation concentration, becoming 'supersaturated' (a negative deficit). Under these conditions, oxygen will exchange from the water to the atmosphere at a rate dependent on the degree of supersaturation and on the characteristics of the stream. A turbulent stream with lots of 'white water' will transfer oxygen to the atmosphere faster than a calm stream with a smooth surface.

During the night when the depleting effect of respiration dominates, the oxygen partial pressure in the water may drop below the saturation concentration. There may be some oxygen 'deficit', and gas transfer will occur from the air into the water, at a rate dependent on the degree of oxygen deficit, and on the characteristics of the stream.

The characteristics of the stream that determine the reaeration coefficient ( $k$ ) will change with flow. Generally, the reaeration coefficient will increase as flow increases, due to increased velocity and turbulence of the water. However, increases in flow can also increase water depth which tends to decrease the reaeration coefficient.

At points in time when water oxygen concentration is at equilibrium with the atmosphere (100% of saturation concentration, or 0 deficit) there will be no physical gas exchange at the surface. The  $kD$  component of Equation 1 will equal 0. If this occurs at night, when the contribution of photosynthesis is also 0, then  $\frac{dDO}{dt}$  will reflect the rate of respiration alone.



## 2.2. Estimating each component – the theory

The approach to isolating the two aspects of metabolic activity relies on the fact that photosynthesis does not occur at night-time when it is dark, leaving just the effects of respiration and physical aeration influencing dissolved oxygen concentrations.

Because the rate of reaeration is linearly related to oxygen deficit, while respiration can be assumed to be constant if oxygen is not limiting, the reaeration coefficient,  $k$ , can be estimated from the gradient (slope) obtained by regressing change in  $O_2$  ( $dDO/dT$ ) against  $O_2$  deficit. Further, the rate of respiration can be estimated by looking at the rate of change in  $O_2$  at times when deficit is zero—the intercept of the regression line with the y axis (where  $x = 0$ ).

Figure 2 shows an example of the relationship that one might expect to see between oxygen deficit and rate of oxygen change, during night-time. When the deficit is zero, the only process causing a change to oxygen levels will be the respiration occurring during the dark. This is estimated by the intercept of the regression line with the  $x = 0$  line (as indicated by a dashed line in the graph). The reaeration coefficient describes how gas exchange increases with increasing deficit—the gradient of the regression line, assuming respiration is constant.

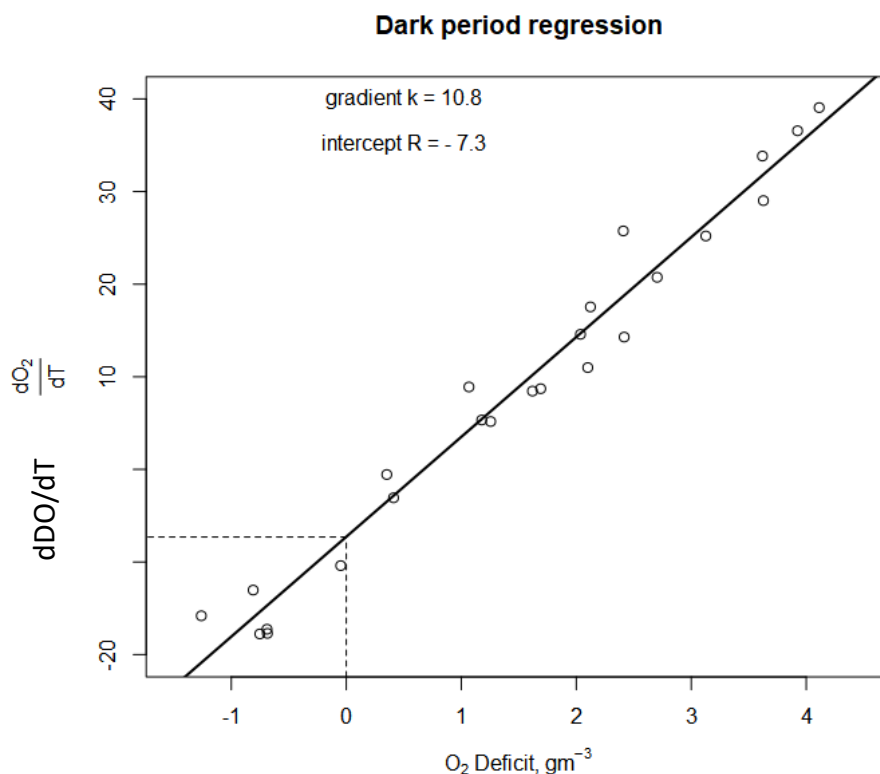


Figure 2. A regression of change in oxygen level against oxygen deficit for a night-time period. The gradient of the line denotes the reaeration coefficient (in this example  $10.8 \text{ day}^{-1}$ ) and the y-intercept denotes the respiration rate (in this example,  $-7.3 \text{ g}\cdot\text{m}^{-3}\cdot\text{day}^{-1}$ ). Respiration is shown as a negative because it involves oxygen uptake.

The regression of oxygen change against deficit during a time with no photosynthesis, provides an estimate of both the reaeration coefficient ( $k$ ) and the rate of respiration ( $R$ ). We assume that the rate of respiration and the reaeration coefficient are constant over time (at least over moderate time periods), so for the light periods preceding and following the dark period shown in the regression of Figure 2, the respiration and reaeration values are known.  $\frac{dDO}{dT}$  is known from measurement. From Equation 1 that leaves only  $P$ , the rate of photosynthesis, unknown. With just one unknown it is easy to rearrange and solve the equation, thus estimating photosynthesis as well.

$$P = \frac{dDO}{dT} - R - kD \quad \text{Equation 2}$$

## 2.3. Practical implementation of respiration and reaeration estimation

### 2.3.1. Dark period identification

The procedure described above relies on using data from a period when photosynthesis is not occurring. Predicted sunset and sunrise times could be obtained from meteorological sites or calculated algorithmically, but the site may be subject to geographic or riparian shading, such that photosynthesis ceases before sunset. It is the gradient soon after photosynthesis ceases that best characterises the rate of respiration occurring during the day. Therefore, missing this period by defining the dark period as starting at sunset can reduce accuracy of metabolism estimates.

It cannot be assumed that as soon as oxygen concentration starts to decrease from its maximum, that photosynthesis has ceased and that the decrease is due to respiration alone being active. It simply means that respiration and reaeration are outweighing any effects of photosynthesis. However, the point of maximum rate of decrease in oxygen can be taken as the start of the dark period, 'darkStart'.<sup>1</sup>

Complementarily, the end of the dark period can be taken as the point where the oxygen deficit is at its maximum. This will tend to be the oxygen minimum.

In almost all oxygen sensors there is 'noise', or small-magnitude random fluctuations, observed in continuous measurements of oxygen concentrations. This 'noise' in the dataset means that the absolute maximum in oxygen deficit may not occur exactly at dawn. Similarly, the highest rate of decrease in oxygen concentration between two subsequent measurements may not occur exactly at dusk. To help remove the effects of random instrument noise, and better isolate the overall maximum and minimum of these parameters, the time series can be smoothed before seeking extremes of

---

<sup>1</sup> Courier font refers to terms used in R code.

oxygen concentrations. We have used a five-point moving mean before subsequent analysis to smooth the data and thus reduce the effects of random instrument noise. Data falling between the `darkStart` and `darkStop` are coded as dark period data, to be used in the regression. Data between `lightStart` and `lightStop` are coded as light period data.

The process is then repeated for the next day represented in the dataset until the dark period start and stop times have been determined for the whole dataset. These start and stop points will vary over the seasons of a year, as well as exhibiting day to day variability due to cloud cover.

Once the timing of the dark period start and stop times are determined, the next step is to extract the data for each night consecutively, and construct a regression of  $dDO/dT$  against oxygen deficit so rates of respiration and the reaeration coefficient can be estimated, as shown in Figure 2.

### **2.3.2. Temperature correction**

Rates of respiration and the reaeration coefficient are both affected by temperature. We have assumed that  $R$  doubles with a 10 °C increase in temperature (Phinney & McIntire 1965) and the reaeration rate increase by 2.41% per degree Celsius (Kilpatrick et al. 1989). Where water temperature was available, it was used to refine the nightly average values for  $k$  and  $R$  from the night-time regression calculations into a time series with variation reflecting the fluctuations in temperature.

## **2.4. Estimation of gross primary production and net ecosystem metabolism**

Assuming that  $R$  and  $k$  from the adjacent night-time data are the same during the daytime period (or as mentioned above ideally corrected to reflect variations in water temperature over the 24-hour period), it is possible to calculate the rate of photosynthesis or gross primary production (GPP) using Equation 2 and  $dDO/dT$  data from the light period.

Once GPP and  $R$  data are calculated, it is then possible to calculate net ecosystem metabolism (NEM) as the sum of photosynthesis and respiration:  $NEM = P + R$ .

Results are yielded at the same temporal resolution as the oxygen data input (e.g., every 15 minutes), but are in units of mass per area per day (e.g.,  $gm^{-2}\cdot day^{-1}$ ). To estimate total oxygen change per day, the many values per day need to be integrated. With a common timestep between samples, this is simply a case of calculating the mean of each for each 24-hour time period. For instance, the total GPP for a day in  $gm^{-2}$  is the mean of 96 quarter-hourly estimates for that 24-hour period.

With an extended dataset it is possible to calculate daily metabolism values over the length of the dataset, enabling comparisons from day to day, season to season or year to year. An example of daily results for R, GPP and NEM at a site (Hatea River at Whareora) is shown in Figure 3.

#### ***2.4.1. Volumetric to areal conversion***

The regressions of change in oxygen concentration (in  $\text{g}\cdot\text{m}^{-3}\cdot\text{day}^{-1}$ ) against oxygen deficit (in  $\text{g}\cdot\text{m}^{-3}$ ) give reaeration coefficient estimates in units of 'per day' and give respiration rate estimates in units of grams of oxygen per cubic metre per day ( $\text{g}\cdot\text{m}^{-3}\cdot\text{day}^{-1}$ ). To better compare rates of metabolic activity between rivers of different scale, the volumetric estimates are made relative to unit riverbed area, by multiplying by the mean water depth upstream of the oxygen monitoring location. In most rivers, the bulk of metabolic activity occurs on and within the riverbed rather than within the water column itself. Rates of respiration, primary productivity and net metabolism are then expressed in grams of oxygen per square metre per day ( $\text{g}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ ), giving an indication of the relative metabolic activity of the riverbed.

Recommendations on how to estimate mean reach depth are given in Section 5.4.2.

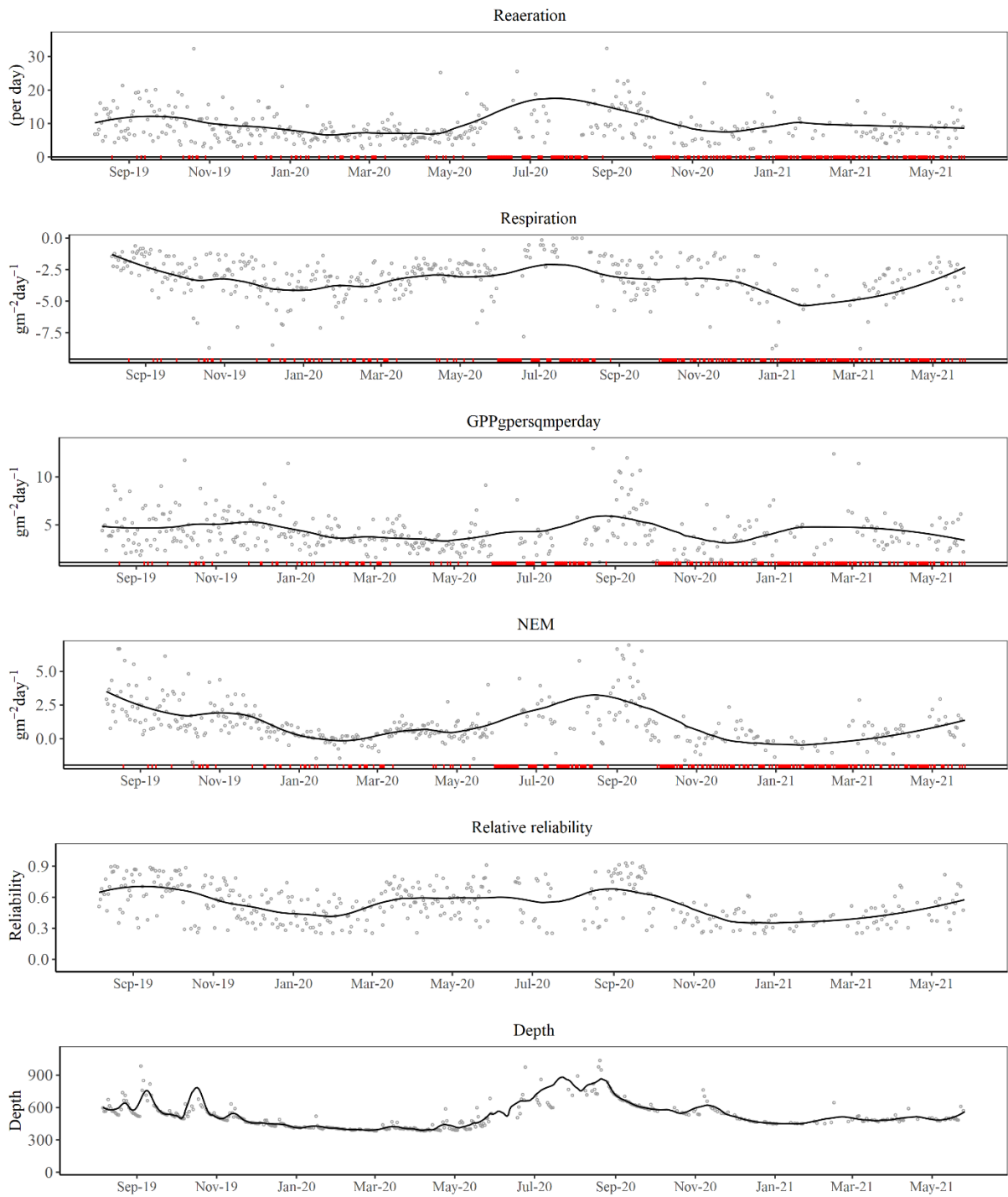


Figure 3. Time series of river depth and derived metabolic properties reaeration, respiration, gross primary production (GPP) and net ecosystem metabolism (NEM). Red points on x axes indicate where estimates were removed due to spurious sign or low reliability.

The stage (water level) is expected to affect  $k$  (reaeration) directly, and while the correlation shown in Figure 4 is poor, there does seem to be a tendency for higher reaeration coefficient at higher stage.

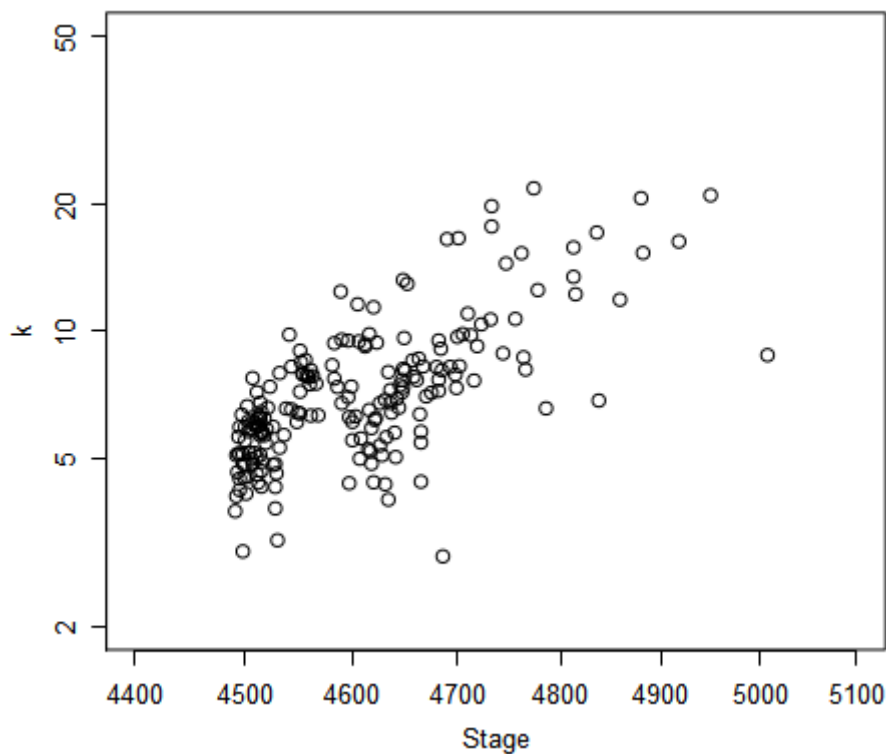


Figure 4. Relationship between river stage (water level in mm) and reaeration coefficient k.

#### ***2.4.2. Reliability and QC of metabolic estimates***

While we use the regression method to estimate the gradient and intercept that represent k and R, it should be noted that the standard requirements and assumptions for a regression are not strictly met by time-series datasets. Specifically, a regression requires that each data point be independent of the others, which requires specific treatment for time series data that is not made here. While the point estimates of k and R are robust to these violations of regression assumptions, the reported intervals around these estimates will be smaller than they should be, and the model performance (i.e.  $R^2$ , the measure of fit) will be exaggerated. Nonetheless, these caveats apply to all data subsets equally, and thus the regression performance diagnostics are comparable within similar metabolic studies. The  $R^2$  value from the regression should be considered as an indication of 'relative reliability', rather than a true  $R^2$  value as would be calculated from independent datapoints. In general,  $R^2$  values above 0.8 are indicative of very robust metabolism estimates, while  $R^2$  values less than 0.4 indicate poor reliability and any associated metabolism estimates should be treated with caution. In the following analyses, we decided that any metabolism estimates calculated with a relative reliability below 0.25 should be removed and replaced with an NA.

The relative reliability of metabolic estimates can vary dramatically throughout a dataset, with periods of poor reliability often being seen during periods of rainfall and high flows (when DO concentrations of river water reflect reaeration processes occurring in raindrops rather than metabolic activity within the river). Poor reliability of metabolism estimates may also occur during extremely low flows if water in the stream channel is no longer well mixed and stratification occurs in the water column. Relatively poor reliability of metabolic estimates may also occur at times in streams with very low rates of production and respiration where the metabolic signal is of similar size to random variations in dissolved oxygen or sensor noise. It is important to note that low reliability of metabolism estimates does not necessarily mean that there is a problem with the accuracy of the DO measurements themselves, but rather that processes other than just photosynthesis, respiration and reaeration are influencing DO.

In addition, it is not possible for the reaeration coefficient to be negative, the respiration rate to be positive or the GPP rate to be negative. Therefore, any estimates that are 'out-of-range' like this must be spurious. Because the various parameters estimated in the metabolism calculations are linked, any out-of-range estimate for one parameter casts doubt on the accuracy of the others. So, for days where any estimate was out of feasible range, all parameters were set to NA, and not included in subsequent analyses.



### 3. INPUT DATASETS

Continuous DO and water temperature data collected from 6 river monitoring sites in Northland have been used for this project. The locations of these sites are shown in Figure 5.

Mean water depth data were measured upstream of each data logger using a reach length of about 20 times the stream width. Mean depth was measured from 50 depth data points collected along 5 transects (i.e., 10 points at each transect) regularly spaced through the reach to cover the range of variability in channel form (Appendix 2). Hydro stage data at each monitoring site were used to develop a relationship between stage and site mean depth.

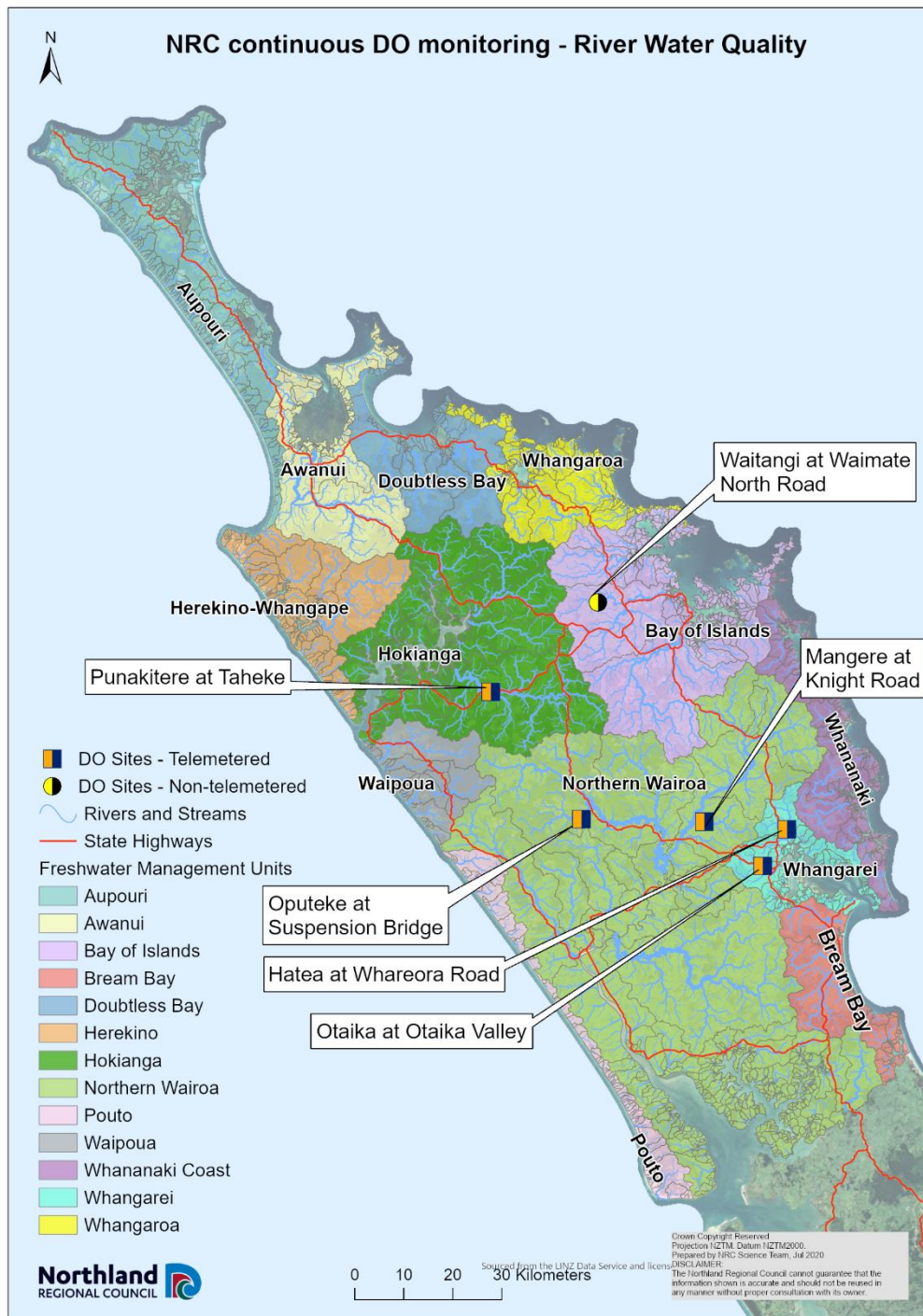


Figure 5. Locations of the monitoring sites where the continuous DO and water temperature data were collected.

### 3.1. Hatea at Whareora Road

DO and water temperature data have been collected continuously at the Hatea at Whareora Road site since mid-2019. DO concentrations at this site followed a typical annual pattern with highest concentrations during winter when water temperatures were lowest, and lowest concentrations in summer/autumn (Figure 6). Daily fluctuations in DO concentrations were typically highest in summer, reflecting increased biological activity during this period (Figure 6). Sharp reductions in the size of daily DO fluctuations were evident at times, particularly in winter 2020, and appeared to coincide with periods of particular high flow (Figure 6; a 100-year flood occurred at this site in July 2020). These flushes probably resulted in removal of aquatic plant biomass from the system, subsequently reducing biological activity and daily DO fluctuations.

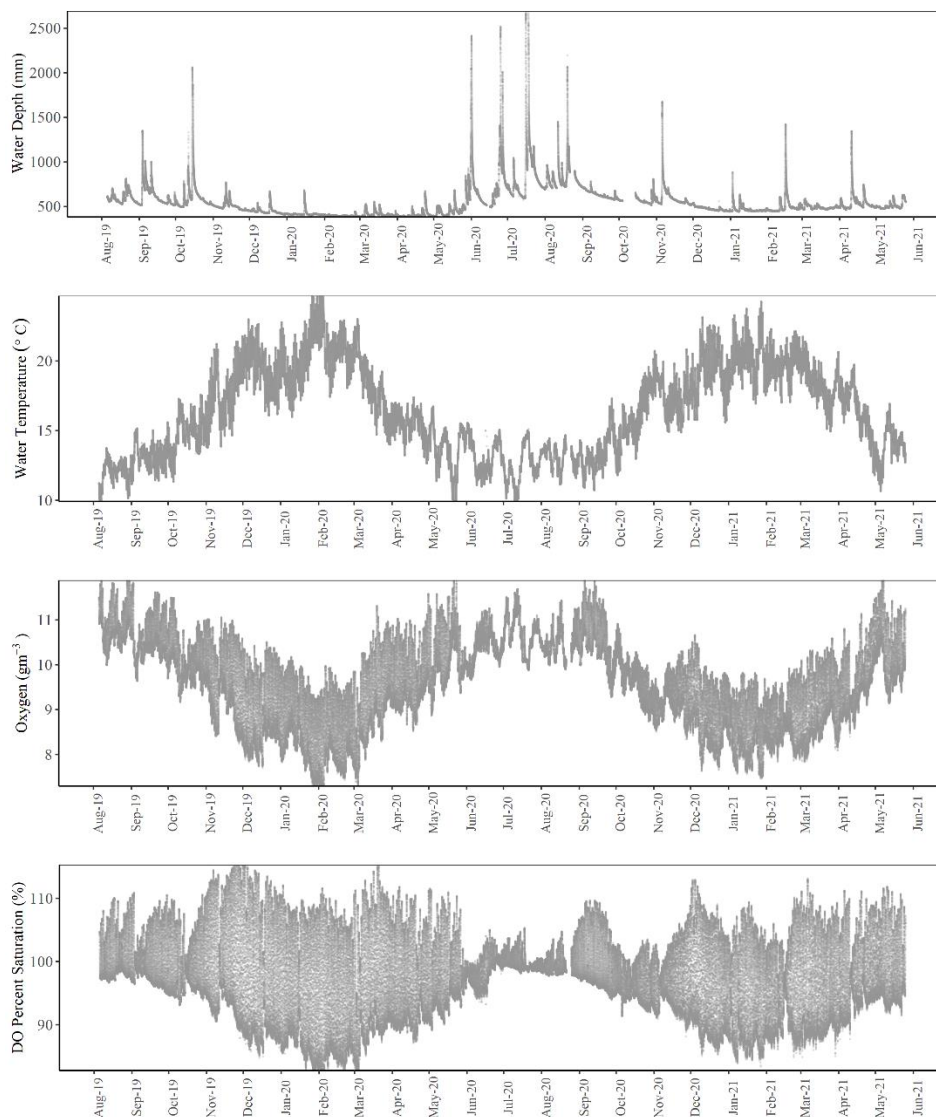


Figure 6. Input data series for Hatea at Whareora Road.

### 3.2. Mangere at Knights Road

DO concentrations at the Mangere at Knights Rd site were highly variable throughout the period of record. Highest daily average DO concentrations, and lowest daily DO fluctuations, were typically seen during winter, while spring, summer and autumn were typically characterised by large daily DO fluctuations (Figure 7). Water temperatures followed the expected pattern of low temperatures in winter and highest temperatures in summer (Figure 7).

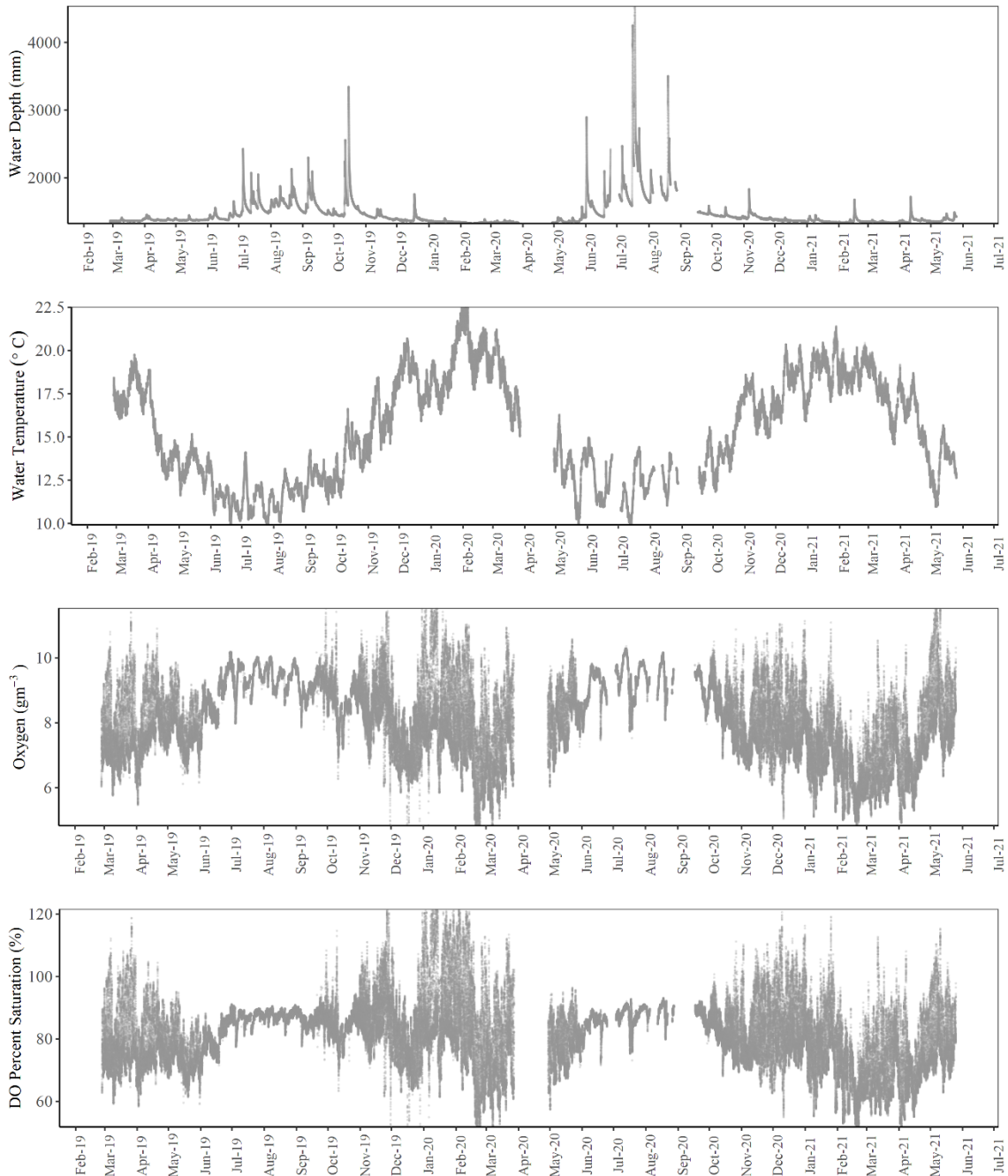


Figure 7. Input data series for Mangere at Knights Road.

### 3.3. Opouteke at Suspension Bridge

DO and water temperature measurements followed a consistent pattern over the two years of monitoring at the Opouteke at Suspension Bridge site. Average daily DO concentrations were highest in winter when water temperatures were coldest, while summer and autumn was characterised by lower average daily DO, but larger daily DO fluctuations, reflecting increased biological activity/metabolism (Figure 8).

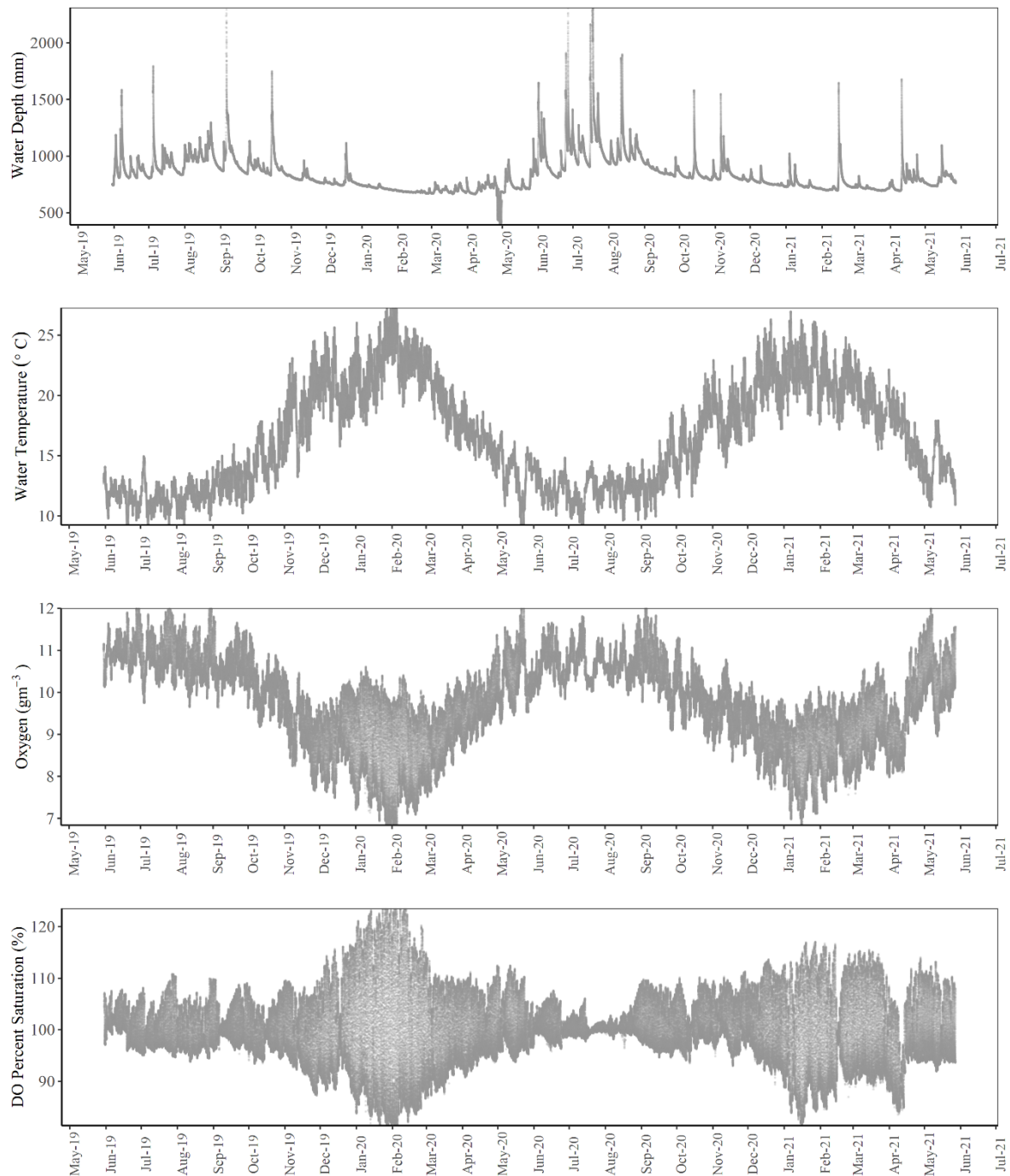


Figure 8. Input data series for Opouteke at Suspension Bridge.



### 3.4. Otaika at Otaika Valley Road

DO concentrations at the Otaika at Otaika Valley Road site also followed the typical pattern of largest daily fluctuations in summer and autumn reflecting increased amounts of biological activity/metabolism, and higher daily average DO (but smaller daily variations) in winter (Figure 9).

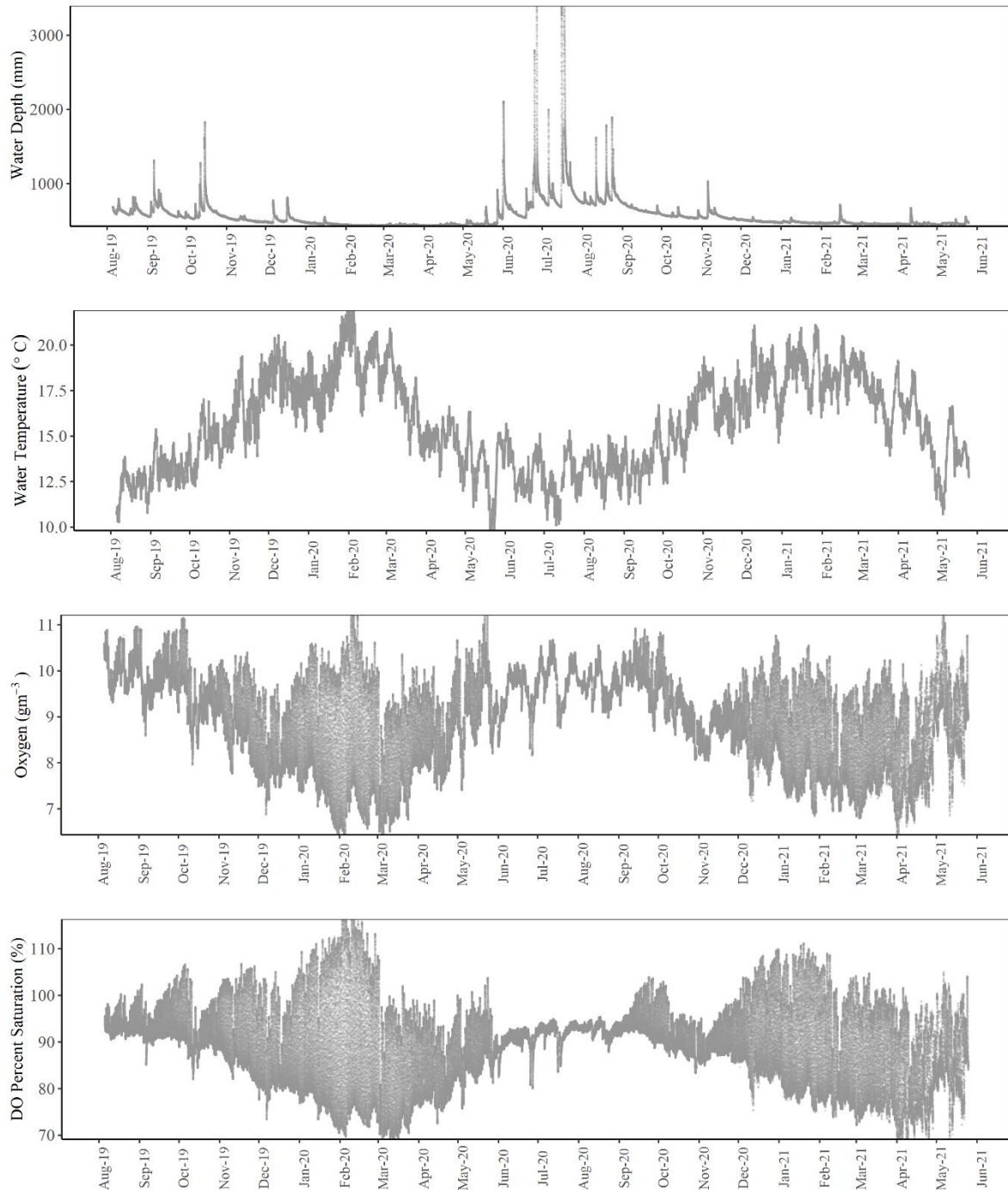


Figure 9. Input data series for Otaika at Otaika Valley Road.

### 3.5. Punakitere at Taheke

DO concentrations at the Punakitere at Taheke site also fluctuated seasonally and in response to elevated flows. The high flows in winter 2020 coincided with a reduction in daily DO fluctuations, perhaps due to flushing of aquatic plant biomass. This led to a sustained period of repeated high flow events and relatively high average daily DO (Figure 10).

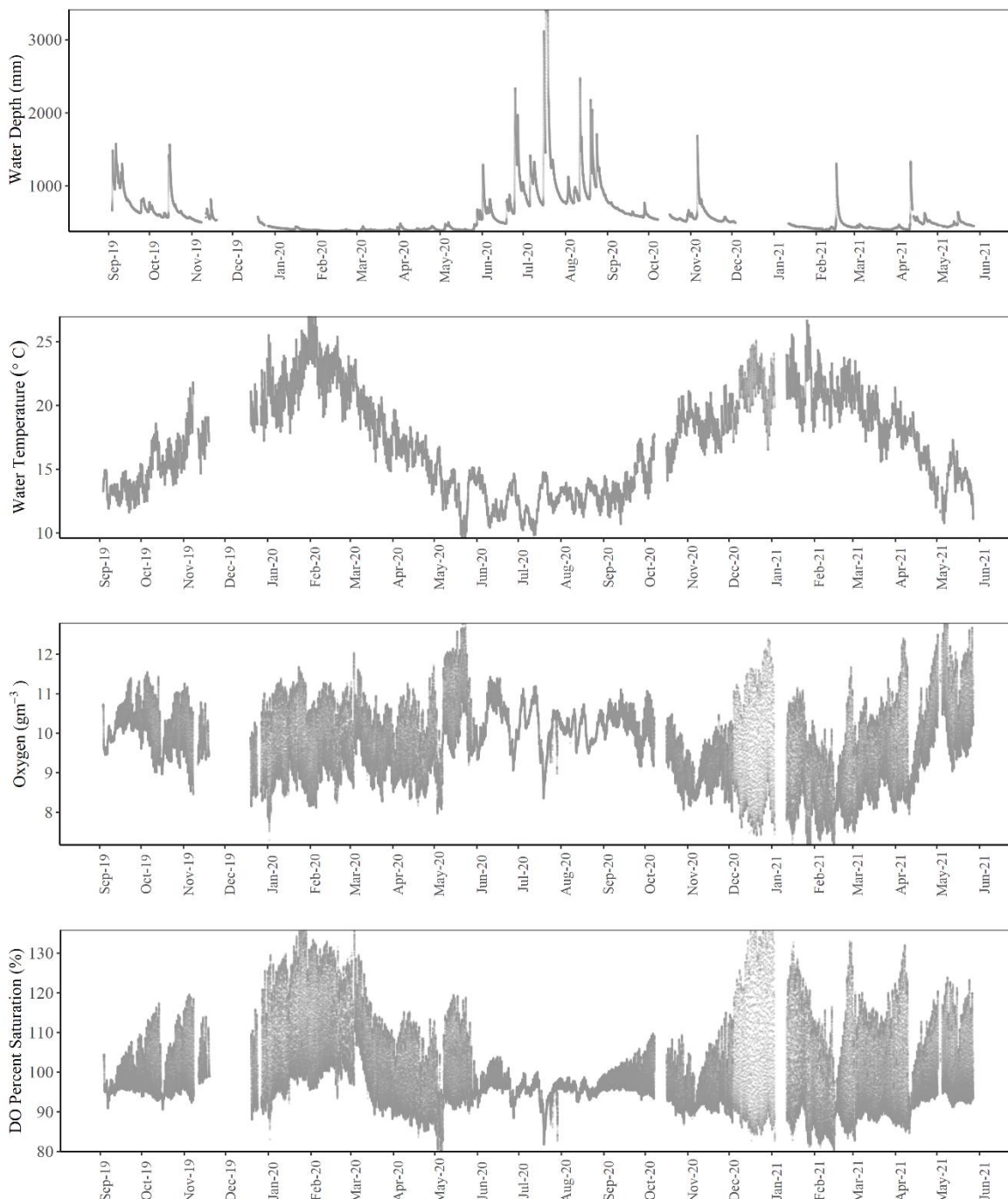


Figure 10. Input data series for Punakitere at Taheke.



### 3.6. Waitangi at Waimate North Road

The dataset from the Waitangi at Waimate North Road was shorter than that from the other sites and also had a substantial gap from June to November 2020. Daily fluctuations in DO were consistently high throughout much of the recording period, with the exception of early winter 2020. In that period, a decline in daily DO fluctuations coincided with several high flow events, which may have flushed aquatic plant biomass from the system, subsequently reducing biological activity/metabolism (Figure 11).

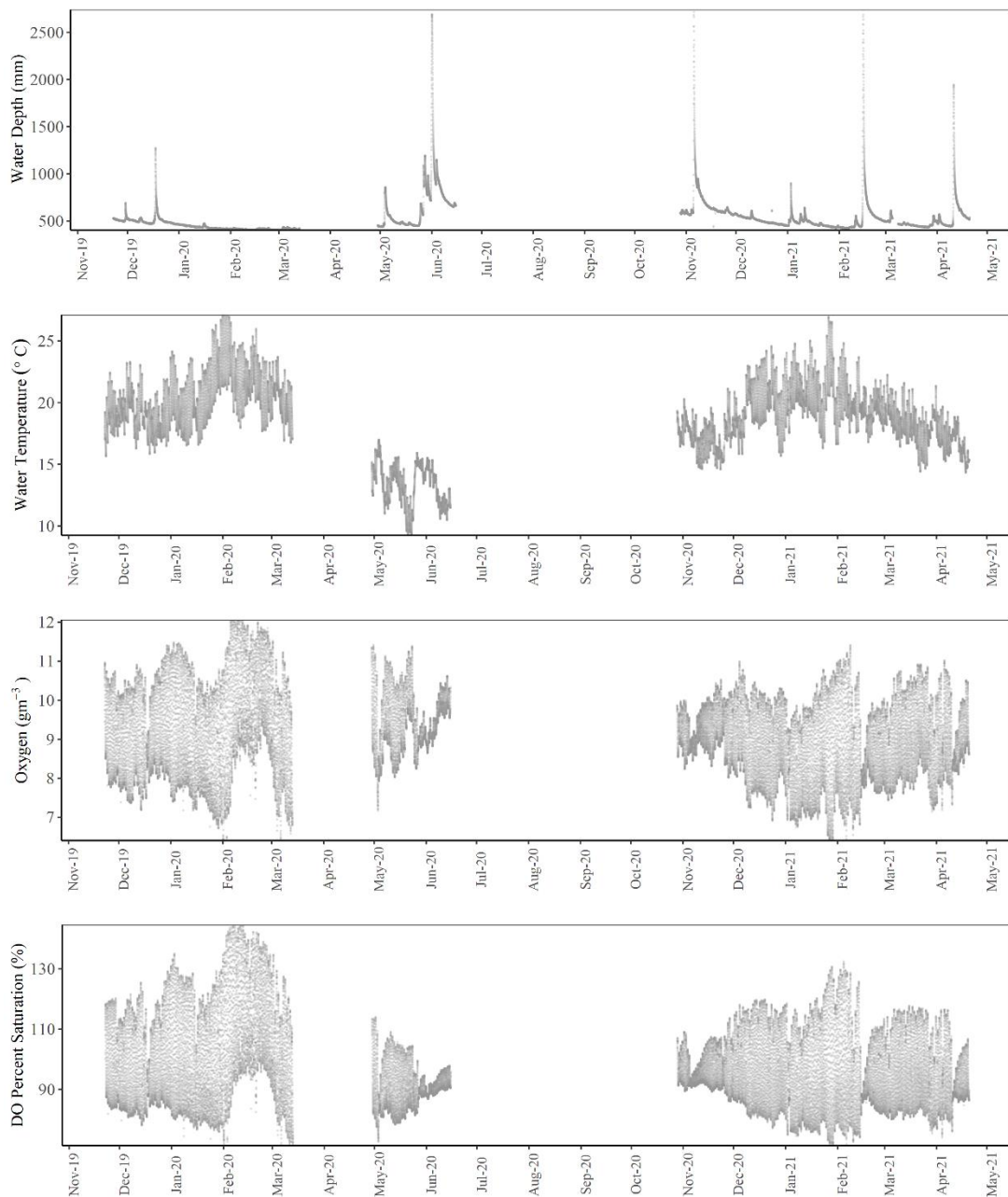


Figure 11. Input data series for Waitangi at Waimate North Road.

## 4. SUMMARISED DISSOLVED OXYGEN AND METABOLISM RESULTS

### 4.1. Dissolved oxygen

Minimum DO concentrations are listed as key attributes reflecting aspects of ecosystem health in the NPS-FM (2020) alongside ecosystem metabolism. We summarised the 7-day and 1-day minima for each of the 6 sites where continuous DO data have been collected by NRC (Table 1). There is some uncertainty in the two DO tables in the NPS-FM, with Table 7 of the NPS-FM specifying that statistics should be calculated from just the summer period while Table 17 of the NPS-FM does not specify a period for the calculation of statistics—inferring that minimum DO should be calculated for the whole year. In Table 1 below, 7-day values are calculated from the whole year, while 1-day values are from just the 1 November to 30 April (summer) period.

Table 1. Minimum annual DO concentrations for the 6 Northland rivers from each of the hydrological years where continuous data were recorded. Colours represent the National Objectives Framework bands from the NPS-FM (2020): A-band (blue), B-band (green), C-band (orange) and D-band (red). Minimum 7-day values are from the whole year, while minimum 1-day values are from the summer period as defined in the NPS-FM (2020) – 1 November – 30 April.

River	Hydrological Year	Minimum 7-Day DO (mg·L <sup>-1</sup> )	Minimum 1-Day DO (mg·L <sup>-1</sup> )
Hatea	2020	7.25	6.95
Hatea	2021	7.7	7.46
Mangere	2019	6.03	5.46
Mangere	2020	4.82	4.02
Mangere	2021	5.07	4.76
Opouteke	2020	6.79	6.57
Opouteke	2021	7.08	6.81
Otaika	2020	6.46	6.17
Otaika	2021	6.66	6.03
Punakitere	2020	8.06	7.3
Punakitere	2021	7.37	6.73
Waitangi	2020	6.67	5.96
Waitangi	2021	6.79	6.31

Minimum DO concentrations in the Hatea and Punakitere river sites are indicative of good ecological conditions during all years when continuous DO data were collected (A-band or B-band; Table 1). The minimum 1-day DO concentrations at Opouteke, Otaika and Waitangi are also indicative of good (B-band) conditions during all years, while minimum 7-day DO indicate some concerns (C-band) (Table 1). Minimum DO measured at the Mangere site was low during most years, but in 2020 minimum 7-day dissolved oxygen concentrations dropped below national bottom line levels (D-band) (Table 1). We understand that there was a severe drought in Northland in summer 2020, which may help to explain the particularly low DO concentrations that year.

## 4.2. Ecosystem metabolism

Metabolic estimates were calculated for the 6 river sites using the continuous data provided by NRC. Differences in ecosystem metabolism among sites can be summarised with key statistics (Table 2) and by comparing the distributions of daily metabolism values (Figure 12). Time series plots of metabolism at each site are provided in Appendix 3. Median GPP was highest at Mangere and lowest at Otaika (Table 2). Similarly, median R was greatest (most negative) at Mangere, but lowest at Punakitere (Table 2).

Table 2. Metabolism summary statistics for the 6 river sites in Northland. Units are  $\text{g O}_2 \text{ m}^{-2} \cdot \text{day}^{-1}$  for Gross Primary Production (GPP), Respiration (R) and Net Ecosystem Metabolism (NEM), and  $\text{day}^{-1}$  for the reaeration coefficient (k).

	Site	Hatea	Mangere	Opouteke	Otaika	Punakitere	Waitangi
GPP	mean	4.4	28.9	6.6	3.6	4.2	6.4
	median	4.0	19.0	6.2	3.4	4.1	6.2
	SD	2.6	54.2	4.5	4.5	2.4	2.2
	minimum	0.2	0.2	0.3	0.1	0.1	0.6
	maximum	28.5	580.3	87.2	85.3	31.2	16.9
	n	525	185	623	498	386	312
R	mean	-3.6	-46.5	-4.3	-6.1	-2.3	-5.4
	median	-3.4	-27.2	-3.8	-4.8	-2.0	-4.6
	SD	2.8	92.4	4.3	9.7	3.2	3.8
	minimum	-50.5	-914.9	-89.0	-196.5	-49.9	-34.2
	maximum	-0.2	-8.5	0.0	-1.7	0.0	-0.6
	n	525	185	623	498	386	312
NEM	mean	0.8	-17.8	2.4	-2.3	2.0	0.9
	median	0.6	-9.1	2.4	-1.2	1.9	1.9
	SD	2.0	44.0	1.8	6.3	2.4	3.7
	minimum	-23.8	-334.7	-11.3	-115.4	-23.2	-19.1
	maximum	13.1	31.9	8.3	2.9	6.4	6.8
	n	525	185	623	498	386	312
k	mean	11	12	11	7	6	6
	median	9	7	9	5	6	5
	SD	9	24	16	9	2	4
	minimum	3	2	3	2	2	1
	maximum	112	261	363	191	27	26
	n	525	185	623	498	386	312

The distribution of daily estimates of GPP, R and NEM for each site demonstrate metabolic differences among the sites (Figure 12). As discussed above, rates of respiration and GPP were highest at Mangere (Figure 12). Annual variability in R was relatively low at Hatea, Opouteke, Otaika, Punakitere and Waitangi, but larger at Mangere (Figure 12). Annual variability in GPP was also higher at Mangere than at the other sites. NEM, which is the difference between GPP and R, was centred on zero for Hatea, Otaika, Punakitere and Waitangi, suggesting reasonably well balanced systems with rates of GPP similar to rates of R. NEM at the Mangere site tended to be negative most of the time suggesting that inputs of organic matter into the system

from upstream or the surrounding catchment were outweighing the local production of organic matter via macrophytes and algae (Figure 12). In contrast, rates of NEM at Opouteke were typically slightly positive, suggesting that local production of organic matter from aquatic plants and algae was larger than local respiration of organic matter. This means that this reach of the river was likely a net exporter of organic matter downstream (Figure 12).

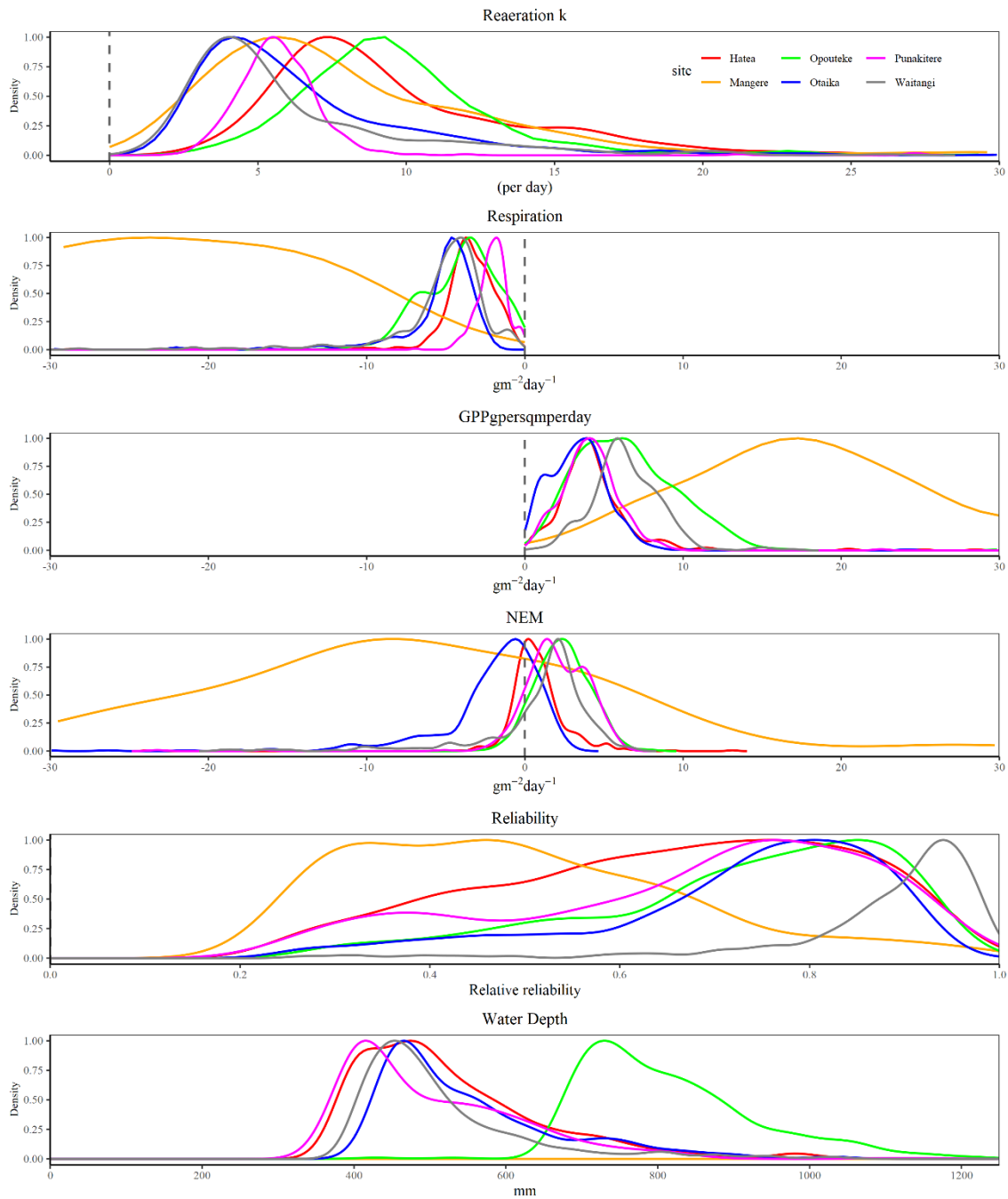


Figure 12. Comparing distributions of metabolic values between rivers. Y axis is the relative frequency of the value shown on x axis.

The distribution of daily relative reliability of the metabolism estimates (as shown in Figure 12) are also informative regarding the accuracy/ease by which metabolism could be estimated from the continuous DO records. The relative reliability of the metabolism estimates at Waitangi was consistently high throughout the data record. Relative reliability was more variable at the other sites, but typically relatively high (> 0.6) (Figure 12). The exception to this was the Mangere site where relative reliability was typically between 0.3–0.5. A visit to this site during the workshop found that the oxygen logger was deployed at the edge of the river channel, among aquatic plants, and the DO measurements in this location were probably not a good representation of DO in the main flowing part of this river channel. This may explain the typically low relative reliability of the metabolism estimates at this site.

The distribution of daily reaeration coefficient estimates at each site is also informative and likely represents the physical characteristics of each site and their response to changing flow conditions. Reaeration coefficients at all sites were typically between 3–10 day<sup>-1</sup> (Figure 12). Opouteke and Hatea typically had the highest reaeration coefficients of the rivers (Figure 12). In our experience, these reaeration rates are low-moderate, reflecting the relatively low channel slope of these Northland rivers, slow water velocity and relatively deep water. For comparison, reaeration coefficients have been estimated to be up to 34 day<sup>-1</sup> in the upper reaches of the Taieri River in Otago (Young & Huryn 1996) and over 100 day<sup>-1</sup> in some small steep Otago streams (Young & Huryn 1999).

Bands to interpret metabolism were not included in the NPS-FM. However, a potential banding system for ecosystem metabolism measurements, based on Young et al. (2008), was discussed by the Science and Technical Advisory Group that helped inform the NPS-FM (2020) (Freshwater Science & Technical Advisory group 2019; and Figure 13). Classifications using this banding system, and based on median ecosystem metabolism measurements, are shown in Table 3. Median rates of GPP at Otaika were indicative of excellent health (A-band) while rates at Hatea and Punakitere (B-band), Opouteke and Waitangi (C-band) were indicative of good to satisfactory ecosystem health according to this banding system (Table 3). In contrast, rates of GPP at Mangere were indicative of unsatisfactory health (D-band) using this banding system (Table 3). Rates of R at the Mangere site were also indicative of severely impaired ecosystem health (D-band), whereas rates of R at all the other sites were indicative of excellent health (A-band). These health assessments based on ecosystem metabolism are broadly similar to those calculated from the minimum DO assessments (Table 1), with impaired ecosystem health at the Mangere site and excellent to satisfactory health at the other sites. This is encouraging, but ideally further work is required to inform the development of criteria that are used to determine how ecosystem metabolism values should be interpreted.

Value	Ecosystem health				
Freshwater Body Type	Rivers				
Attribute	Ecosystem metabolism				
Attribute Unit	g O <sub>2</sub> m <sup>-2</sup> d <sup>-1</sup> (grams of dissolved oxygen per square metre per day)				
Attribute State	Numeric Attribute State <sup>1</sup>				Narrative Attribute State
	Gross primary production		Ecosystem respiration		
	Non-wadeable	Wadeable	Non-wadeable	Wadeable	
A	≤3.0	≤3.5	1.6-3.0	1.6-5.8	No evidence of an impact on ecosystem metabolism.
B	>3.0 and <5.5	>3.5 and <5.0	>1.0 and <1.6 Or >3.0 and >8	>1.2 and <1.6 Or >5.8 and <7	Mild effect on ecosystem metabolism.
C	≥5.5 and ≤8.0	≥5.0 and ≤7.0	≥0.6 and ≤1.0 Or ≥8.0 and ≤13.0	≥0.8 and ≤1.2 Or ≥7.0 and ≤9.5	Moderate effect on ecosystem metabolism.
D	>8.0	>7.0	<0.6 or >13.0	<0.8 or >9.5	Severely impaired ecosystem metabolism.

1. Derived from 7 consecutive days of continuous dissolved oxygen monitoring. Objective applies year-round.

Figure 13. Potential banding system for ecosystem metabolism as discussed by the Freshwater Science and Technical Advisory Group 2019 that helped inform the National Policy Statement-Freshwater Management (2020).

Table 3. Median ecosystem metabolism measurements for the 6 Northland sites. Colours represent the metabolism bands for wadeable rivers discussed by the STAG but not eventually included in the NPS-FM (2020): A-band (blue), B-band (green), C-band (orange) and D-band (red).

Site	GPP (gO <sub>2</sub> ·m <sup>-2</sup> ·day <sup>-1</sup> )	R (gO <sub>2</sub> ·m <sup>-2</sup> ·day <sup>-1</sup> )
Hatea	4.0	-3.4
Mangere	19.0	-27.2
Opouteke	6.2	-3.8
Otaika	3.4	-4.8
Punakitere	4.1	-2.0
Waitangi	6.2	-4.6



## 5. RECOMMENDATIONS REGARDING DO MONITORING AND METABOLISM CALCULATION

### 5.1. National Environmental Monitoring Standard for dissolved oxygen

The National Environmental Monitoring Standard for dissolved oxygen (2016) provides detailed guidance on measurement of dissolved oxygen (DO) and approaches for processing and archiving of DO data. There is considerable detail in this National Environmental Monitoring Standard document, and we support the guidance and recommendations included. In our opinion, the National Environmental Monitoring Standard document should be the primary guide on DO monitoring and we do not intend to repeat the detailed guidance in this report. Rather, we focus on key details at a relatively high level and provide more focus on factors relating to DO monitoring that have specific relevance to subsequent ecosystem metabolism calculation.

### 5.2. Monitoring considerations

#### 5.2.1. *Sensor types*

There are a wide variety of DO sensors available. We strongly recommend the use of optical DO sensors, which have become commonly available over the last 10 years. Optical DO sensors are less prone to fouling than electrochemical sensors, and thus require less frequent calibration. Optical DO sensors don't consume DO as part of the measurement process, so are able to record accurate DO measurements in slow moving water without the need for an additional stirring mechanism to refresh the water at the surface of the sensor.

We have considerable experience with YSI, Hydrolab and D-opto DO sensors/loggers, but do not consider one brand of sensor/logger necessarily better than another. We understand that the data from 5 of the Northland sites were collected using YSI EXO sondes, while data from the Waitangi site were collected using a HOBO optical DO logger. The data collected from all sites was consistent with that we have seen in other parts of New Zealand and are suitable for metabolism calculation (assuming that appropriate deployment, calibration and sensor maintenance practices have been employed, as discussed further below).

Some sensor brands incorporate 'wipers' that reduce the likelihood of sensor fouling. This feature is preferable but not essential if a regular maintenance programme with approximately monthly sensor cleaning is followed.

### **5.2.2. Deployment locations**

Oxygen sensors should be deployed in a location where the water passing over the sensor reflects well-mixed conditions representative of the main water column. This is typically close to the thalweg of the flow. For logistical reasons, it is also important to deploy the sensor in a location where it will not be washed out or fouled by debris during a flood (and safe from vandalism). These placement requirements for good readings and for safety can be conflicting, but the primary consideration should be placing the logger in a position where well-mixed water can be measured. There is no point having DO measurements from a secure location that is a 'backwater' not influenced by the main flow.

Sensors are best placed where there is limited or no sunlight on the sensor surface. This will help to discourage fouling caused by growth of algae on the sensor surface.

### **5.2.3. Deployment duration**

Ideally, sensors/loggers should be deployed continuously to gather information on annual, seasonal and daily fluctuations in DO and ecosystem metabolism.

If resources make long-term continuous logging difficult, then a focus on collection of data during the summer/autumn period is an alternative approach. Such deployments should be done for a minimum of 7 days, but longer deployment periods are preferred. DO and metabolism data collected from summer/autumn are likely to represent a 'worst-case' scenario where warm temperatures, low flows, aquatic plant accumulations and long sunlight hours will contribute to annual DO and metabolism extremes.

Comparisons in DO and metabolism data among sites need to consider the duration and timing of data collection. It is not appropriate to compare data collected in winter at one site with data collected in summer at another site.

### **5.2.4. Measurement frequency**

We recommend that DO is monitored every five minutes, although a measurement interval anywhere between 2 minutes and 15 minutes is appropriate for metabolism calculation.

## **5.3. Calibration and maintenance**

The National Environmental Monitoring Standard document on DO provides detailed guidance on calibration and maintenance procedures that should be followed. Appropriate calibration and maintenance procedures will vary to some extent

depending on the brand of sensor/logger. Therefore, it is also important to follow the manufacturer's instructions when calibrating the sensors.

Some brands of sensor involve 1-point calibration, typically at 100% saturation. Other brands require 2-point calibration, typically at 0 mg·L<sup>-1</sup> and 100% saturation. Two point calibrations give more confidence across a wider range of possible DO measurements.

For the calibration at 100% saturation, some sensors use air-saturated water, while instructions for other models suggest using water-saturated air. If possible, we suggest using water-saturated air as it is often difficult to aerate water to exactly 100% saturation, which can lead to subsequent data bias.

Ideally, sensor calibration should be done at a temperature similar to that which will be experienced during the subsequent deployment. The same advice applies with salinity.

Barometric pressure varies with weather conditions (high pressure versus low pressure systems). More importantly, it also varies with site altitude. Some sensors/loggers have built in barometric pressure sensors that will help to compensate for the variations due to weather conditions. However, significant care needs to be taken when monitoring sites at different altitudes. Ideally, sensors should be calibrated at the altitude that will be experienced during subsequent deployment. However, an alternative approach is to calibrate all sensors at a fixed altitude (e.g., NRC offices) and correct the data later with the known difference in altitude and barometric pressure between sites and calibration location.

Sensors should be regularly checked, cleaned and if necessary re-calibrated. We recommend this is done monthly, but this might need to vary depending on the conditions at individual sites and the sensors used. For example, sensors at some sites might foul more quickly than at other sites, while sensors with built in 'wipers' are unlikely to be affected by fouling even over long deployment periods.

Re-calibration, if necessary, should be done with an independent reference DO sensor. The calibration of this independent sensor is of course critical and also needs to consider issues with water temperature, salinity, barometric pressure and altitude that apply to the deployed sensors.

Detailed records of cleaning, maintenance and re-calibration need to be kept as metadata alongside the DO records.

## 5.4. Other data required

### 5.4.1. Water temperature

Water temperature is a critical parameter to determine how DO concentration in  $\text{mg}\cdot\text{L}^{-1}$  is related to DO saturation. Therefore, all reputable DO sensors will also have water temperature sensors. Water temperature sensors need to be checked in a similar way to DO sensors but are typically much more robust and not subject to fouling. Therefore, re-calibration of temperature sensors is not typically required or even possible with most loggers.

### 5.4.2. Mean depth upstream

An estimate of mean depth in the river reach upstream of the DO logger location is required to convert volumetric measurements to areal measurements, so metabolism estimates can be compared among sites.

We recommend taking about 10–20 depth measurements at equal offset distances across each of 5 transects upstream of the logger location. The transects should be located at intervals of about 5 times the stream width apart over a reach of about 20 times the stream width, upstream of the logger location, so that the variation in channel form at the site is captured in the depth measurements. If the channel morphology is very variable, then depth measurements across more transects should be taken.

If a water stage height recorder is at the site, make sure that the stage is measured when the depths are measured so change in average depth with changing flow can be estimated. Ideally, a relationship between stage height and mean depth should be developed at each site (using multiple measures of mean reach depth at different flows/water levels), so an estimate of mean depth on any day can be used to accurately convert the volumetric metabolism estimates to areal measures for every day. In the absence of such a relationship, mean depth can be assumed to vary linearly with stage height (if stage height data are available) or assumed to be constant despite changes in flow and water level.

### 5.4.3. Light

Light measurements are not required for the calculation of metabolism using the method that is discussed in this report. However, differences in light caused by cloud cover and riparian shading are likely to contribute to differences in metabolism over time and among sites. Having this information is helpful when interpreting the data, but not essential.

#### 5.4.4. Water quality

Information on standard water quality parameters like water clarity, nutrient concentrations, conductivity and pH are useful in helping to interpret differences in metabolism among sites. However, they are not an essential part of metabolism measurement.

### 5.5. Data processing

The National Environmental Monitoring Standard for dissolved oxygen provides detailed guidance on how independent measurements taken during the regular check/clean/re-calibration can be used to correct data from the logger, if required. Drift in instrument measurements can be an issue if fouling of the sensor occurs. Sudden step-changes in DO measurements associated with cleaning and maintenance are a signal that corrections are required.

The National Environmental Monitoring Standard also provides guidance on the assignment of quality codes to the data. We support the recommendations in the National Environmental Monitoring Standard regarding the use of quality codes, although note that measurements of DO above 100% saturation are real and equally accurate to DO measurements at a similar distance to the calibration points for the instrument. It seems harsh to 'downgrade' the quality code for data just above 100% saturation, as recommended in the National Environmental Monitoring Standard, although we agree that there is increasing uncertainty about measurements that are increasingly above 100% saturation.

As discussed above, raw data collected from a site may need to be corrected for differences in altitude between calibration location and deployment location.

Metabolism calculations, particular for respiration, are strongly influenced by the accuracy of the DO measurements. Therefore, it is critical that consistent bias associated with poor instrument calibration, effects of site altitude, or instrument drift are addressed before metabolism calculations are attempted. Any datasets with minimum daily DO saturation above 100% saturation should be checked carefully to see if there is a likely data bias.

NRC are currently using the following data processing steps:

1. Both temperature and DO  $\text{mg}\cdot\text{L}^{-1}$  are used as the starting raw data.
2. No adjustments are made to temperature as we are then unable to establish the exact corresponding change in DO.
3. Where there is an obvious temperature issue, both temperature and DO data are removed for the affected period.

4. Where temperature data are out of tolerance but within trend, NRC do not ramp the temperature but instead downgrade the QC code in the corresponding DO data.
5. DO data are ramped only when required.
6. DO% is calculated using a virtual measurement.
7. DO% is adjusted for barometric pressure also via a virtual measurement.
8. Altitude corrections are applied to U26 HOBO data (using the National Environmental Monitoring Standard equations).
9. Quality codes are applied to DO mg·L<sup>-1</sup>, DO% and data archived.
10. Temperature data are archived as RAW (QC200).

Step 1 will depend on the type of sensor used. Some sensors read the temperature and DO and calculate the DO%. Others read the temperature and DO% and calculate the DO. Depending on the sensors some also perform the required barometric adjustments. Knowledge of equipment specifications are really important prior to processing the data.

## 5.6. Metadata

Recording and saving metadata alongside the DO and temperature data are obviously important. Metadata might include details such as site information, site location, sensor type and calibration history, observations during cleaning/maintenance trips, corrections to raw data that have been applied, and National Environmental Monitoring Standard quality code.

Assuming that metabolism results are also saved in a database, it would be important to include information on the relative reliability of the calculated metabolism estimates (the R<sup>2</sup> regression statistic), the reaeration coefficient, the estimated mean depth upstream, and the number of data points used in the night-time regression.

NRC record all metadata for a specific site and time period in an excel processing document which is then attached to the dataset in their Hilltop database.

## 5.7. Data storage and archiving

Several sets of data need to be stored, including the unedited raw DO and water temperature data, the final checked and verified DO and temperature data, and the associated metadata mentioned above, including National Environmental Monitoring Standard quality code.

Once metabolism calculations have been conducted on the checked and verified data we recommend that the daily metabolism estimates (GPP and R) are stored, along with information on the relative reliability of the calculated metabolism estimates (the  $R^2$  regression statistic), the reaeration coefficient, the estimated mean depth upstream, and the number of data points used in the night-time regression to calculate metabolism on each day.

Supporting data such as flow, barometric pressure and light should also be stored, if available.

## 5.8. Curation of code for metabolism calculations

The code used to collate and analyse the data and create the plots presented in this report is provided in the appendices. Some parts of this code are specific to Northland Regional Council's database structure and bespoke to how they store data. Other sections of the code are more general and focus on the calculation of metabolism and associated summary statistics assuming data is structured in a particular way.

One issue that has become apparent with the delivery of code via an appendix to a report, as is done here, is that people will apply and modify this code, meaning it will evolve over time independently for each group using and modifying the code. This can be very problematic when trying to identify and understand the cause of different results produced by different versions of code.

The code provided with this report is accurate and up-to-date currently, but it is likely that further improvements to the code will be incorporated in the future. To avoid multiple versions of the code circulating and evolving independently, we intend to provide a centrally curated version of the code in the future that incorporates any updates. This may be done through the GitHub code repository system with potential for users to be notified about updates to the code. Therefore, if readers are wanting to develop systems for automated metabolism calculation, we recommend contacting the authors of this report first rather than just immediately applying the code provided in the appendices to this report.

## 6. SUMMARY

This report provides details on the development of an R-code based approach for estimating ecosystem metabolism from continuous DO records, the results of an analysis of ecosystem metabolism from 6 sites in Northland with continuous DO data, and recommendations to help guide future monitoring of DO and ecosystem metabolism in Northland.

R-code has been developed to calculate ecosystem metabolism from long-term continuous DO records. This is a significant step forward since metabolism calculation using this method has only been possible on a day-by-day basis using a spreadsheet, previously.

Minimum DO concentrations were at good or satisfactory levels at most of the sites, but concentrations were consistently in the C-band at the Mangere site. The 7-day minimum concentration was below the national bottom line at Mangere during the 2020 drought. Rates of GPP and R were high at the Mangere site, and potentially indicative of poor ecosystem health, although the location of the DO logger at this site influenced the reliability of the metabolism estimates, particularly during low flow periods. Rates of GPP and R at the other sites were indicative of excellent to satisfactory ecosystem health. Further work is ideally required to inform further development of criteria that are used to determine how ecosystem metabolism values should be interpreted.

The National Environmental Monitoring Standard for DO measurement is a very useful document and should be considered as the primary guide on DO monitoring. The recommendations provided here on DO monitoring are focussed on the collection of datasets suitable for subsequent ecosystem metabolism calculation. Smart deployment of well-calibrated DO sensors in the right locations is critical. A regular monthly monitoring site check and instrument clean is recommended, alongside recalibration of the DO sensors, if required. Careful data processing will sometimes be required to address calibration issues and sensor drift. Sudden step-changes in DO measurements associated with cleaning and maintenance are a signal that corrections are required. An estimate of mean depth in the river reach upstream of the DO logger location is required to convert volumetric measurements to areal measurements, so metabolism estimates can be compared among sites. Other data such as flow, light and water quality are useful to help understand variation in metabolism over time or among sites but is not essential. Recording and storing metadata (e.g., site location, sensor type and calibration history, and observations during cleaning/maintenance trips) alongside the DO and temperature data is obviously important.

The code used to collate and analyse the data and create the plots presented in this report is provided in the appendices. To avoid multiple versions of the code circulating



and evolving independently, we intend to provide a centrally curated version of the code in the future that incorporates any updates.

## 7. ACKNOWLEDGEMENTS

We thank Northland Regional Council Freshwater Scientist Manas Chakraborty for initiating this work and helping to progress DO and ecosystem metabolism assessment in Northland and throughout New Zealand. We also thank Preston Ferreira and the wider NRC team for their assistance with data provision and interactions on code development and their ideas and knowledge that were shared during the workshop and subsequently. We also acknowledge the assistance of Paula Casanovas (Cawthron) for her assistance with code development. This work was mainly funded by an Envirolink medium advice grant (NLRC225) with additional funding by Northland Regional Council.

## 8. REFERENCES

- Freshwater Science & Technical Advisory group 2019. Report to the Minister for the Environment. Publication reference number: CR 372. 58 p.  
<https://environment.govt.nz/publications/freshwater-science-and-technical-advisory-group-report-to-the-minister-for-the-environment/>
- Kilpatrick FA, Rathbun RE, Yotsukura N, Parker GW, De Long LL 1989. Determination of stream reaeration coefficients by use of tracers. In Techniques of water-resources investigations 3-A18. United States Geological Survey, Denver, Colorado, USA.
- National Environmental Monitoring Standard 2016. Dissolved Oxygen. Measuring, processing and archiving of dissolved oxygen data. Version: 2.0.  
<https://www.nems.org.nz/documents/dissolved-oxygen/>
- [National Policy Statement for Freshwater Management 2020.](https://environment.govt.nz/acts-and-regulations/national-policy-statements/national-policy-statement-freshwater-management/)  
<https://environment.govt.nz/acts-and-regulations/national-policy-statements/national-policy-statement-freshwater-management/>. May 2021.  
[Accessed September 2021.](#)
- Phinney HK, McIntire CD 1965. Effect of temperature on metabolism of periphyton communities developed in laboratory streams. *Limnology & Oceanography* 10: 341-344.
- Young RG, Huryn AD 1996. Interannual variation in discharge controls ecosystem metabolism along a grassland river continuum. *Canadian Journal of Fisheries and Aquatic Sciences* 53: 2199–2211.
- Young RG, Huryn AD 1999. Effects of land use on stream metabolism and organic matter turnover. *Ecological Applications* 9: 1359–1376.
- Young RG, Matthaei CD, Townsend CR 2008. Organic matter breakdown and ecosystem metabolism: functional indicators for assessing river ecosystem health. *Journal of the North American Benthological Society* 27: 605–625.

## 9. APPENDICES

### APPENDIX 1. R CODE

The code below was used to collate and analyse the data and create the plots presented in this report. Some parts of this code are specific to Northland Regional Council's database structure and bespoke to how they store their data. Other sections of the code are more general and focus on the calculating of metabolism and associated summary statistics. To avoid multiple versions of the code circulating and evolving independently, we intend to provide a centrally curated version of the code in the future that incorporates any updates. Therefore, if readers are wanting to develop systems for automated metabolism calculation, we recommend contacting the authors of this report first rather than just immediately applying this code.

```
#Definition----
# This code extracts continuous time series DO data from a Hilltop file,
then
# performs high level data cleaning (removes spikes) and some data
wrangling
# before calculating DO metabolic parameters. This is all
# done for sites that are part of a predefined list of sites

#_Input----
#A list of DO sites: This is a csv file with four columns:
#1) first column is the site name as defined in Hilltop, any
# misspelling of the site name would result in error (script will not
extract
# data)
# 2) second column is short name for the site, the script uses this in
naming
#of variables and objects.
# 3) and 4) are reference stage and depth values used to convert
continuous
# stage data to river depth, which is used in the script to metabolic
parameters
#path: Q:\Automation NRC\R Scripts\Water Quality\Reference Files\...
# List of DO sites used in metabolic calculation.csv

#A Hilltop file (hts) with stage, DO concentration and DO percent,
#water temperature and pressure data for each site. This file is currently
# set to a combined process file once this data is archived, will point
# to the final archive
#path: Q:\Projects and Requests\Freshwater quality and ecology\working
area\...
# Envirolink DO Project\DO Data Set_Respiration Calculation.hts

#_Output----
# The outputs are:
# 1) a csv file for each site (with both the input data and calculated
# metabolic coefficients).
# 2) As well as a set of four plots
# per site:
# i. Site Raw Data
# ii. A plot of the sites metabolic parameters
# iii. A distribution plot of the different metabolic parameters
# iv. A plot of the DO minimum plot, is a time series plot of
```

```
# -7day mean of the DO minimum
# - daily DO minimum
# - DO concentration mg/l
# 3) Statistical Summary of metabolic results, this is a csv file.
# 4) Two distribution plots comparing the metabolic distribution across all
sites
# 5) summary of the 7day minimum and minimum of the daily minimums for each
# hydro year. output is a csv file called SOE DO minimum Stats
# the outputs are saved to
# Q:\Projects and Requests\Freshwater quality and ecology\working area\...
# Envirolink DO Project\output_from_script

#_Contact Details----
# This script was based on the R-script developed by Eric Goodwin and was
# extracted from Section3.7&3.8 of the Metabolism Calculation
Report_21_April.
# However, simplified by Steve Kitto and Preston Ferreira.

# Date last modified on 11th of May 2022 by
# Preston Ferreira (prestonf@nrc.gov.nz)

#####
++++
# Install and activate packages----
# List of packages used in script besides Hilltop package, that will have
to
# be installed from the zip file.
packages <- c("tidyverse", "caTools", "parallel", "doParallel", "mgcv",
"zoo", "pROC",
"ggplot2", "gridExtra")

# Installing packages that have not been installed
install.packages(setdiff(packages, rownames(installed.packages())))

# Activate packages
library(tidyverse, warn.conflicts = F) # used for data handling
library(caTools) # used to calculate a 7 day rolling minimum
library(lubridate) # handling time conversion
library(parallel) # perform parallel computing to speed up script
library(doParallel) # to implement parallel processing
library(ggplot2) # graphically represent data
library(gridExtra) # plot grid
library(mgcv) # used to calculate smoothed averages
library(pROC)
library(zoo) #required by Hilltop
library(Hilltop) # used to extract time series data from a hts file
```

```

#####
++++
# Input Parameters----
#Ensure R does not convert Date objects from NZST to NZDT
Sys.setenv(TZ = "Pacific/Wallis")

# define working directory
path_wd <- paste0("Q:/Projects and Requests/Freshwater quality and
ecology/",
                 "working area/Envirolink DO
Project/output_from_script/update from Paula")

setwd(path_wd)

# import DO sites
path_dosite <- paste0("Q:/Automation NRC/R Scripts/Water Quality/",
                    "Reference Files/List of DO sites used in metabolic
",
                    "calculation.csv")

dosites <- read_csv(path_dosite)

# define Date range
startdate = "" # extract data from when first data was recorded

# define end date as today
enddate <- as.POSIXct(ceiling_date(Sys.time() - days(1), unit = "days"))

# Hilltop file. Current set to the processing file
hilltop_path <- paste0("Q:/Projects and Requests/Freshwater quality and
ecology",
                      "/working area/Envirolink DO Project/DO Data ",
                      "Set_Respiration Calculation.hts")

#####
++++
# User Defined Function----
stdise <- function(x){
  x=x-min(x,na.rm=T)
  x=x/max(x,na.rm=T)
  return(x)
}

stDensity = function(x,...){
  denis = density(x,na.rm=T,...)
  denis$y = stdise(denis$y)
  return(denis)
}

```

```
}

qtdise=function(x,probs=c(0.01,0.99)){
  x=x-quantile(x,na.rm=T,probs=probs[1])
  x=x/quantile(x,na.rm=T,probs=probs[2])
  return(x)
}

minza<-function(x){
  min(x[which(x>0 & !is.na(x))])
}

sumX <- function(x){
  return(round(c(mean(x,na.rm=T),
                median(x,na.rm=T),
                sd(x,na.rm=T),
                min(x,na.rm=T),
                max(x,na.rm=T),
                sum(!is.na(x)),
                sum(is.na(x))),2))
}

# Theme function, ensure theme used across all ggplots are consistent.
theme_pt <- function(base_family = "serif"){
  theme_bw(base_family = base_family) +
    theme(text = element_text(family = "serif"),
          panel.grid.major.x = element_blank(),
          panel.grid.minor.x = element_blank(),
          panel.grid.major.y = element_blank(),
          panel.grid.minor.y = element_blank(),
          panel.border = element_rect(color = "grey20"),
          plot.title = element_text(size = 12, hjust = 0.5),
          axis.text = element_text(size = 8),
          axis.title.y = element_text(size = 8),
          axis.line = element_line(color = "black"),
          legend.text = element_text(size = 8),
          legend.title = element_text(size = 10))
}

ts_plot <- function(plot_data, col_ind, ylab = ""){

  #ts_plot is a times series plot. which is used to generate time series
  plot
  # of Site data.

  #input:
  #1) plot_data, a data frame/tibble of at least two columns, one of the
  columns
  # is a date column
  #2) col_ind is the column index in plot_data of the y-variable to plot
  #3) ylab is the y-axis title, it is an optional variable

  #output is a times series plot
}
```

```

ggplot(plot_data, aes(x = date, y = plot_data[, col_ind])) +
  geom_point(alpha = 0.2, color = "grey60", size = 0.15) +
  scale_x_datetime(breaks = "1 month",
                  date_labels = "%b-%y") +
  scale_y_continuous(limits = quantile(plot_data[, col_ind],
                                      probs = c(0.001, 0.999),
                                      na.rm = T),
                    expand = c(0, 0)) +
  labs(x = "", y = ylab) +
  theme_pt()+
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x = element_text(angle = 90, size = 10))
}

#####
++++
#Connect to Hilltop----
hilltop_file <-HilltopData(hilltop_path)

for(i in 1:nrow(dosites)){

  dosite <- dosites$`Hilltop Site name`[i]

  sitename <- dosites$`Script Site name`[i]

  #Get all data from Hilltop
  DO_Perc_data <-
    tryCatch(GetData(hilltop_file, dosite,
                    "Dissolved Oxygen % Continuous Calculated [Dissolved
Oxygen % Continuous Calculated]",
                    startdate,enddate), error=function(err) NA)

  DO_mg_data <-
    tryCatch(GetData(hilltop_file, dosite,
                    "Dissolved Oxygen Continuous [Dissolved Oxygen
Continuous]",
                    startdate,enddate), error=function(err) NA)

  Stage_data <-
    tryCatch(GetData(hilltop_file, dosite, "Stage", startdate,enddate),
              error=function(err) NA)

  Temp_data <-
    tryCatch(GetData(hilltop_file, dosite,
                    "Water Temperature Continuous [Water Temperature
Continuous]",

```

```

        startdate,enddate), error=function(err) NA)

Baro_data <-
  tryCatch(GetData(hilltop_file, dosite,
                  "Barometric Pressure Continuous [Barometric Pressure
Continuous]"),
           startdate,enddate), error=function(err) NA)

All_Data <- merge (DO_Perc_data, DO_mg_data, Stage_data, Temp_data,
Baro_data)

siteData <- fortify.zoo(All_Data)

# find when continuous DO monitoring started and remove dates before this
point
frstxDate = min(siteData$Index[which(!is.na(siteData$DO_mg_data) |
!is.na(siteData$DO_Perc_data))])
siteData = siteData[siteData$Index > frstxDate,]

#####
#####
#REMOVE THIS SECTION ONCE WE ARE WORKING WITH PROCESS DATA

# High Level Filtering based on code from Eric
if('Baro_data'%in%names(siteData)){
  siteData$Baro_data[siteData$Baro_data<500] <- NA
}#if('Baro_data'%in%names(siteData))

siteData$DO_mg_data[siteData$DO_mg_data<2] <- NA
siteData$DO_mg_data[siteData$DO_mg_data>13.3]=NA
siteData$DO_Perc_data[siteData$DO_Perc_data<20]=NA
siteData$DO_Perc_data[siteData$DO_Perc_data>160]=NA
#####
#####

# Adding SatC
siteData$satC <- 100*(siteData$DO_mg_data/siteData$DO_Perc_data)

#create a deficit column
siteData$deficit <- (100*(siteData$DO_mg_data/siteData$DO_Perc_data))-
siteData$DO_mg_data

#create a depth column from stage using site stage data
siteData$Depth_data <- siteData$Stage_data -
  dosites$Stage_mm[dosites`Script Site name` == sitename] +
  dosites$Depth_mm[dosites`Script Site name` == sitename]

#additional filtering
siteData$satC[siteData$satC>16] <- NA
siteData$deficit[siteData$deficit>30] <- NA

```



```

# adding time related variables to siteData
siteData$timeofday = hour(siteData$Index) + minute(siteData$Index)/60
siteData$dayofyear = yday(siteData$Index)
siteData$jday = round(as.numeric(julian(siteData$Index)))
siteData$date <- ymd_hms(siteData$Index)
siteData <- as.data.frame(siteData)

#Create mg per day column
complete <- which(!is.na(siteData$DO_mg_data))
siteData$DO_mg_data_perday <- NA
siteData$DO_mg_data_perday[complete] <- c(NA,

diff(siteData$DO_mg_data[complete])/diff(as.numeric(ymd_hms(siteData$Index[
complete]))/60/60/24)) #g/m3 Per day

#make sure there are no NA in the DO_Perc_data and deficit values
#before calculating the running mean for smoxy, smdoxy and smdef
siteData <- siteData[!is.na(siteData$DO_Perc_data),]

# Calculating the smooth mean for dissolved oxygen and deficit
# this is a 5 point running mean smoother
siteData$smoxy <- runmean(siteData$DO_mg_data, k=5, endrule='keep')
siteData$smdoxy <- runmean(siteData$DO_mg_data_perday,
k=5, endrule='keep')
siteData$smdef <- runmean(siteData$deficit, k=5, endrule='keep')

# Re-ordering siteData, used in plotting the data
siteData <- siteData%>%select(date, jday, dayofyear, timeofday,
Stage_data,
                                Temp_data, DO_mg_data, DO_Perc_data,
                                everything(), -Index)

# Save csv file
write.csv(siteData, file = paste0('./', sitename, '-
MetaData.csv'), row.names=F)
cat('+\\n')
}

disconnect(hilltop_file)

#####
####
#Setup Parallel Processing----
workers = makeCluster(min(c(7, nrow(dosites))))
registerDoParallel(workers)
clusterCall(workers, function() {
  library(tidyverse)
  library(lubridate)
  library(caTools)

```

```

require(mgcv)
library(zoo)
})

#####
####
#Calculate Metabolic parameters----
foreach(dosite = dosites`Hilltop Site
name`,.combine=rbind,.errorhandling="stop")%dopar%{

  sitename <-
    dosites`Script Site name`[which(dosite == dosites`Hilltop Site
name`)]

  #import site data
  siteData <- read.csv(paste0('./',sitename,'- MetaData.csv'),
                      stringsAsFactors = F)

  #ensure date time is in correct datetime format
  siteData$date <- ymd_hms(siteData$date)
  siteData$nightofyear <- NA
  siteData$dayofyear <- NA

  # Calculating the number of days in data set
  ydaysMod <- unique(siteData$jday[which(!is.na(siteData$deficit) &
!is.na(siteData$DO_mg_data_perday))])
  nDays <- length(ydaysMod)

  # creating an empty data.frame which will be used as a place holder for
  # calculated metabolic parameters
  daily <- data.frame(jday = ydaysMod,
                    date = as.Date(ydaysMod,origin=lubridate::origin),
                    darkStop=rep(NA,nDays),
                    darkStart=rep(NA,nDays),
                    lightStop=rep(NA,nDays),
                    lightStart=rep(NA,nDays),
                    Rgppersqmperday=rep(NA,nDays), #respiration
                    k=rep(NA,nDays), #reaeration
                    GPPgppersqmperday=rep(NA,nDays),
                    NEMgppersqmperday=rep(NA,nDays),
                    reliability=rep(0,nDays))

  dailyDepth <- siteData%>%
    group_by(jday)%>%
    summarise(Depth_data = median(Depth_data,na.rm=T))%>%
    ungroup

  daily$depth <- dailyDepth$Depth_data[match(daily$jday,dailyDepth$jday)]
  rm(dailyDepth)

```

```

thisyday=1
# Find dark period and light period
for(thisyday in thisyday:length(ydaysMod)){
  #rudimentary progress bar
  cat('.')
  if(thisyday%in%((1:100)*80))cat('\n')
  #day changes at lunchtime 12 midday, so all of one night shares a jday
number
  daySpanning <- which(siteData$jday==(ydaysMod[thisyday]))

  # change this line to the original line that takes into account the
actual time difference between records (P!)
  # this is important because if there are NAs in between records, the
calculation of the day spanning won't work
  timeResn =
as.numeric(names(sort(table(diff(siteData$date)),decreasing=T)))[1]

  if(length(daySpanning)<(16/(timeResn/60))) {next}

  darkStart=NA
  darkStop=NA
  lightStart=NA
  lightStop=NA

  # The established method uses the most negative dO/dt to determine the
timing of dusk.
  darkStart <-
siteData$timeofday[daySpanning[which.min(siteData$smdoxy[daySpanning])]]
  lightStart <-
siteData$timeofday[daySpanning[which.max(siteData$smdef[daySpanning])]]

  daily$darkStart[thisyday] <- darkStart
  daily$darkStop[thisyday] <- lightStart
  daily$lightStart[thisyday] <- lightStart
  daily$lightStop[thisyday] <- darkStart
  rm(darkStart,lightStart)

  if(daily$darkStop[thisyday]<12){
    theseDark <- which((siteData$timeofday>daily$darkStart[thisyday] &
siteData$jday==ydaysMod[thisyday]) |
                      (siteData$timeofday<daily$darkStop[thisyday] &
siteData$jday==(ydaysMod[thisyday])))
  }else{
    theseDark <- which((siteData$timeofday>daily$darkStart[thisyday]) &
                      (siteData$timeofday<daily$darkStop[thisyday]) &
                      (siteData$jday==ydaysMod[thisyday]))
  }
  theseLight <- which(siteData$timeofday>daily$lightStart[thisyday] &
siteData$timeofday<daily$lightStop[thisyday] &
siteData$jday == ydaysMod[thisyday])
  theseDay <- which(siteData$jday == ydaysMod[thisyday])

  siteData$nightofyear[theseDark] <- thisyday
  siteData$dayofyear[theseLight] <- thisyday

  rm(daySpanning)

```

```

    rm(theseDark)
  }

  #Replace NAs with mean of flanking
  if(any(is.na(daily$darkStart))){
    daily$lightStop[which(is.na(daily$lightStop))] <-
      daily$darkStart[which(is.na(daily$darkStart))] <-
      sapply(which(is.na(daily$darkStart)),FUN = function(i){
mean(c(daily$darkStart[rev(which(!is.na(daily$darkStart[1:i])))] [1]),
daily$darkStart[(i:dim(daily)[1])[which(!is.na(daily$darkStart[i:dim(daily)
[1]])][1]]),na.rm=T)
    })
  }

  if(any(is.na(daily$darkStop))){
    daily$lightStart[which(is.na(daily$lightStart))] <-
      daily$darkStop[which(is.na(daily$darkStop))] <-
      sapply(which(is.na(daily$darkStop)),FUN=function(i){
      mean(c(daily$darkStop[rev(which(!is.na(daily$darkStop[1:i])))] [1]),
daily$darkStop[(i:dim(daily)[1])[which(!is.na(daily$darkStop[i:dim(daily)[1
]])][1]]),na.rm=T)
    })
  }

  #Through again having figured out the start/stop times
  siteData$NEMgpersqmpersday <- 0
  siteData$GPPgpersqmpersday <- 0
  siteData$Rgpersqmpersday <- NA
  siteData$k <- NA
  siteData$Rgpersqmpersdayadj <- NA
  siteData$kadj <- NA
  siteData$reliability <- NA

  for(thisyday in seq_along(ydaysMod)){
    cat('.')
    if(thisyday%in%((1:100)*80))cat('\n')
    # jday changes at NZ midday
    if(daily$darkStop[thisyday]<12){
      theseDark <- which((siteData$timeofday>daily$darkStart[thisyday] &
siteData$jday==ydaysMod[thisyday]) |
      (siteData$timeofday<daily$darkStop[thisyday] &
siteData$jday==ydaysMod[thisyday]))
    }else{ #darkStop is before midnight,
      theseDark <- which((siteData$timeofday>daily$darkStart[thisyday]) &
      (siteData$timeofday<daily$darkStop[thisyday]) &
      (siteData$jday==ydaysMod[thisyday]))
    }
    theseLight <- which(((siteData$timeofday>=daily$lightStart[thisyday]) &
(siteData$timeofday<12) &
      (siteData$jday == (ydaysMod[thisyday]))) |
      ((siteData$timeofday<=daily$lightStop[thisyday])
& (siteData$timeofday>12) &

```

```

                                (siteData$jday == (ydaysMod[thisyday+1])))
theseDay <- which((siteData$jday ==
ydaysMod[thisyday]&siteData$timeofday<12) |
                                (siteData$jday ==
ydaysMod[thisyday+1]&siteData$timeofday>=12))

siteData$nightofyear[theseDark] <- thisyday
siteData$dayofyear[theseLight] <- thisyday

if(length(theseDark)>2){
  toModDat <- droplevels(siteData[theseDark,])

  if(!all(is.na(toModDat$smdoxy))&!all(is.na(toModDat$deficit))){

    # fitting linear model to the night time data for each day
    kR_regn <- lm(data=toModDat,smdoxy~smdef) #g/cube

    daily$Rgpersqmpday[thisyday] <- coef(kR_regn)[1] *
(daily$depth[thisyday]/1000) #respiration ("r_nightavg") (g per square
metre per day)
    daily$k[thisyday] <- coef(kR_regn)[2] #reaeration ("k_nightavg")
(per day)
    daily$reliability[thisyday] <- summary(kR_regn)$r.squared

    siteData$Rgpersqmpday[theseDay] <- daily$Rgpersqmpday[thisyday]
    siteData$k[theseDay]<- daily$k[thisyday]

    # used correlation from "Interannual variation in discharge
controls
# ecosystem metabolism along a grassland river continuum" by
# Roger G. Young and Alexander D. Huryn
if('Temp_data'%in%names(siteData)){
  meanTemp <- mean(siteData$Temp_data[theseDark],na.rm=T)
  siteData$Rgpersqmpdayadj[theseDay] <-
siteData$Rgpersqmpday[theseDay] *
  (1.07177^(siteData$Temp_data[theseDay] - meanTemp))
  siteData$kadj[theseDay] <- siteData$k[theseDay] *
  (1.0241^(siteData$Temp_data[theseDay] - meanTemp))
  rm(meanTemp)
}else{
  siteData$Rgpersqmpdayadj[theseDay] <-
siteData$Rgpersqmpday[theseDay]
  siteData$kadj[theseDay] <- siteData$k[theseDay]
}

#This will get written over for the light period below
siteData$NEMgpersqmpday[theseDay] <-
siteData$Rgpersqmpdayadj[theseDay]
  rm(kR_regn)
}else{
  if(all(is.na(toModDat$DO_per_Sat))){
    daily$reliability[thisyday]=-0.1 #Code to alert dodgy %DO data
  }
}

```

```

    }
    rm(theseDark,toModDat)
  }

  if(length(theseLight)>2){
    toModDat <- droplevels(siteData[theseLight,])
    toModDat$timeStep <- (toModDat$date-toModDat$date[1])/60/60/24
    toModDat$smoxy <- lowess(toModDat$timeStep,toModDat$DO_mg_data,f =
1/4)$y
    toModDat$smDO <-
c(NA,diff(toModDat$smoxy)/diff(as.numeric(toModDat$timeStep)))
    toModDat$smdef <- lowess(toModDat$timeStep,toModDat$deficit,f=1/4)$y
    toModDat$smDOSat <-
lowess(toModDat$timeStep,toModDat$DO_Perc_data,f=1/4)$y
    # dO/dT = P + kD + R
    # P = dO/dT -kD - R
    toModDat$GPPgpersqmpersday <- (toModDat$DO_mg_data_persday -
toModDat$kadj*toModDat$deficit -
toModDat$Rgpersqmpersdayadj/(toModDat$Depth_data/1000))*(toModDat$Depth_data
/1000)
    toModDat$smGPPgpersqmpersday <- (toModDat$smDO -
toModDat$kadj*toModDat$smdef -
toModDat$Rgpersqmpersdayadj/(toModDat$Depth_data/1000))*(toModDat$Depth_data
/1000)

    toModDat$NEM <- (toModDat$Rgpersqmpersdayadj +
toModDat$GPPgpersqmpersday)
    toModDat$smNEM <- (toModDat$Rgpersqmpersdayadj +
toModDat$smGPPgpersqmpersday)

    siteData$GPPgpersqmpersday[theseLight] <- toModDat$smGPPgpersqmpersday
    siteData$NEMgpersqmpersday[theseLight] <- toModDat$smNEM

    daily$GPPgpersqmpersday[thisyday] <-
mean(toModDat$smGPPgpersqmpersday,na.rm=T)
    daily$NEM[thisyday] <- mean(toModDat$smNEM,na.rm=T)

    #add reliability
    siteData$reliability[theseDay]<- daily$reliability[thisyday]
  }
  rm(theseLight,theseDay)
}

#save Site Metabolic output
write.csv(siteData,file = paste0('./',sitename,'-
MetaData.csv'),row.names=F)

# save data to be extracted and used in graphical analysis
eval(parse(text=paste0(sitename,'=list(siteData,daily)'))))

eval(parse(text=paste0("save(",sitename,",file='",sitename,"Metab.rData'")
))

cat('+\\n')

```

```

}#foreach

stopCluster(workers)
rm(workers)

#+++++
++++
#Visualise:Plot----
#Generate multiple parameters plots per site
sumStat = data.frame(site=dosites$`Script Site name`, meanR=NA,medianR=NA,
  SDR=NA, minR=NA, maxR=NA, nR=NA, nNAR=NA, meank=NA,
  mediank=NA, SDk=NA, mink=NA, maxk=NA, nK=NA, nNAk=NA,
  meanGPP=NA, medianGPP=NA, SDGPP=NA,
  minGPP=NA,maxGPP=NA,nGPP=NA,nNAGPP=NA, meanNEM=NA,
  medianNEM=NA,SDNEM=NA,minNEM=NA,maxNEM=NA,nNEM=NA,
  nNANEM=NA)

allsite_distcomp = NULL

#create data frame to store DO minimums
DOminima = expand.grid(site = dosites$`Script Site name`,
  hyear = paste0('hy',2015:year(enddate)),
  stringsAsFactors = F)

#add new column Min7Day and Min1Day to DO minima data frame
DOminima$Min7Day=NA
DOminima$Min1Day=NA

#plot site: raw data, metabolic parameters, distribution plots and DO
minimums
for(dosite in dosites$`Hilltop Site name`){

  sitename <-
    dosites$`Script Site name`[which(dosite == dosites$`Hilltop Site
name`)]

  #load data
  load(paste0(sitename, 'Metab.rData'), verbose=T)

  eval(parse(text=paste0("contData=", sitename, "[[1]]")))
  eval(parse(text=paste0("daily=", sitename, "[[2]]")))

```

```

#####
+++
#Calculate DO minimums----
#7-day mean minimum
#"The 7-day mean minimum, which is the mean value of 7 consecutive daily
#minimum values"
# "The 1-day minimum, which is the lowest daily minimum across the whole
#summer period"

#add Julian day, day of the year, and define whether date falls within
the
#summer
contData$day <- as.numeric(floor(julian(contData$date)))
contData$yday <- yday(contData$date)
contData$summer <- contData$yday>335 | contData$yday<60

#add year column and hydrology year column
contData$year <- year(contData$date)
contData$hydYear <- contData$year
contData$hydYear[contData$yday > 180 & contData$yday < 370] <-
contData$hydYear[contData$yday > 180 & contData$yday < 370] + 1

#calculate minimum DO per day
dailyMin <- contData%>%
group_by(day)%>%
summarise(minDO = min(DO_mg_data, na.rm = T)) %>%
mutate(minDO = ifelse(is.finite(minDO), minDO, NA)) #add, returns inf
when no data

#calculate 7 day mean of the minimums
dailyMin$day7meanMin <- runmean(x = dailyMin$minDO, k = 7, align =
'right')

# add results to site data
contData$dailyMin <- dailyMin$minDO[match(contData$day, dailyMin$day)]
contData$day7meanMin <- dailyMin$day7meanMin[match(contData$day,
dailyMin$day)]

#calculate the minimum of the 7 day mean of the minimums across each
#hydrological year
day7meanmin <- contData %>%
group_by(hydYear) %>%
summarise(min7daymin = min(day7meanMin, na.rm = T))

#calculate the minimum DO for each year across summer period only
dailyMin <- contData %>%
filter(summer) %>%
group_by(hydYear) %>%
summarise(minDailyMin = min(dailyMin, na.rm = T))

#add data to DMinma data frame, find site row

```



```

siteRows <- which(DOminima$site == sitename)

#7 day mean
DOminima$Min7Day[siteRows[match(paste0('hy', day7meanmin$hydYear),
                                DOminima$hyeare[siteRows])]] <-
  day7meanmin$min7daymin

# daily minimum
DOminima$Min1Day[siteRows[match(paste0('hy', dailyMin$hydYear),
                                DOminima$hyeare[siteRows])]] <-
  dailyMin$minDailyMin

rm(dailyMin)

#Find out-of-range values, both by logical flag, and by index
spuriousSign = (daily$k<0 | is.na(daily$k) |
                daily$Rgpersqmpersday>0 | is.na(daily$Rgpersqmpersday) |
                daily$GPPgpersqmpersday<0 |
is.na(daily$GPPgpersqmpersday))
toNuke = which(spuriousSign)

#set reliability ("r squared") minimum to 0
daily$reliability[daily$reliability<0] <- 0

#Also remove any estimates from cases with reliability below a selected
#threshold The value used here, 0.25, is arbitrarily chosen. We could
more
#methodically use the value identified by the ROC analysis above
#(stored in CND)
toNuke = unique(c(toNuke,which(daily$reliability<0.25)))

# creating data for generic distribution plot
#site to daily
daily$site <- sitename

#filter data based on toNuke
daily_f <- daily[-toNuke, ]

#save a copy of daily summary
write.csv(daily_f, file = paste0("./",sitename,"- daily data.csv"),
row.names = F)

#merge with other site data, to create a data frame with all site data
#filtered by toNuke, will be used to compare the distributions of
metabolic
#parameters across sites
allsite_distcomp <- rbind(daily_f, allsite_distcomp)

```

```

daily$toNuke = seq_along(daily$reliability)%in%toNuke

# calculate statistic summary for Rgpersqmpерday
sumStat[which(sumStat$site==sitename),2:8] <-
  sumX(daily$Rgpersqmpерday[!daily$toNuke])

# calculate statistic summary for k
sumStat[which(sumStat$site==sitename),9:15] <-
  sumX(daily$k[!daily$toNuke])

# calculate statistic summary for GPPgpersqmpерday
sumStat[which(sumStat$site==sitename),16:22] <-
  sumX(daily$GPPgpersqmpерday[!daily$toNuke])

#calculate statistic summary for NEM
sumStat[which(sumStat$site==sitename),23:29] <-
  sumX(daily$NEM[!daily$toNuke])

spanlength = diff(range(daily$jday))
loessspan=(180/spanlength)

rm(spanlength)

#fit smooth curves to the data that is weighted to reliability
darkDurnLoess <- loess(data=daily[-toNuke,], formula=I((darkStop-
darkStart)%%24)~jday,
  span=loessspan, na.action=na.omit,
control=loess.control(surface='direct'))
lightDurnLoess <- loess(data=daily[-toNuke,], formula=I((lightStop-
lightStart)%%24)~jday,
  span=loessspan, na.action=na.omit,
control=loess.control(surface='direct'))
reaeloess<-loess(data = daily[-toNuke,],formula = k~jday,weights =
reliability, span = loessspan, na.action = na.omit, control =
loess.control(surface='direct'))
resploess<-loess(data = daily[-toNuke,], formula =
Rgpersqmpерday~jday,weights = reliability, span = loessspan, na.action =
na.omit,control = loess.control(surface='direct'))
reliloess<-loess(data = daily[-toNuke,],formula = reliability~jday, span
= loessspan, na.action = na.omit,control = loess.control(surface='direct'))
depthloess<-loess(data = daily[-toNuke,],formula = depth~jday, span =
loessspan/4.5, na.action = na.omit, control =
loess.control(surface='direct'))
gpploess<-loess(data = daily[-toNuke,],formula = GPPgpersqmpерday~jday,
span = loessspan, na.action = na.omit, control =
loess.control(surface='direct'))
nemloess<-loess(data = daily[-toNuke,],formula = NEM~jday, span =
loessspan,na.action = na.omit,control = loess.control(surface='direct'))

# generating a plot of input data (Water depth, water temperature, DO
concentration,

```

```
# DO %) for each site
nMeas <- dim(contData)[2]-4

#_Plot Site Data----
# variables/parameters to include under site plot
var_plot <-
  data.frame(plot_var = c("Depth_data", "Temp_data", "DO_mg_data",
                        "DO_Perc_data"),
            ylab = c("Water Depth (mm)", "expression", "expression",
                    "DO Percent Saturation (%)"))

for(i in 1:nrow(var_plot)){

  # variable to plot
  var <- var_plot$plot_var[i]

  #ylablel
  ylab <- ifelse(var == "Temp_data",
                expression("Water Temperature " ( degree~C)),
                ifelse(var == "DO_mg_data",
                        expression(Oxygen~(" *gm^-3~")),
                        var_plot$ylab[i]))

  #find column index for variable to plot
  col_ind <- which(colnames(contData) == var)

  #assign data to plot data
  plot_data <- contData

  #create plot
  figure <- ts_plot(plot_data, col_ind, ylab)

  #create object, to allow plot to be placed on grid
  eval(parse(text=paste0("f", i,"<- figure")))

  #added suppress warning as data had NA and with limits had warnings
  # points were removed.
  suppressWarnings(print(figure))

}

#create a grid
plot_grid <- suppressWarnings(grid.arrange(f1, f2, f3, f4, nrow = 4))

#save plot of site parameters
ggsave(filename = paste0(sitename,' - Site Data.tiff'), plot = plot_grid,
        device = 'tiff', width = 10, height = 12, dpi = 300, units = 'in',
        type = 'cairo', compression='lzw' )
```

```

# generating plot of output metabolic parameters for each site, where
outputs
#   are Reaeration coefficient (k), Resipartion (R), GPP, NEM, relative
# reliability and depth
#_Plot Metabolic parameters-----
plot_data <- daily[-toNuke,]

##__Plot Dusk Dawn and dark duration ----
dark_duration <- ggplot(data = daily, aes(x = date, y = darkStart)) +
  geom_point(shape = 1, color = "red", size = 0.5) +
  geom_line(data = daily, aes(x = date,
                              y = predict(darkDurnLoess, newdata =
jday))) +
  geom_point(data = daily, aes(x = date, y = darkStop), color = "blue",
            shape = 1, size = 0.5) +
  scale_y_continuous(expand = c(0.02, 0.1),
                    limits = c(0, 24),
                    breaks = c(0, 6, 12, 18, 24),
                    labels = c("", "6", "12", "18", "24")) +
  scale_x_date(breaks = "2 month", date_labels = "%b-%y") +
  labs(x = "", y = "'Time (Hours)", title = "Dusk, Dawn & dark duration")
+
  theme_pt() +
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x = element_text(angle = 0, size = 10))

##__Plot Dusk Dawn and light duration ----
light_duration <- ggplot(data = daily, aes(x = date, y = lightStart)) +
  geom_point(shape = 1, color = "salmon", size = 0.5) +
  geom_line(data = daily, aes(x = date,
                              y = predict(lightDurnLoess, newdata =
jday))) +
  geom_point(data = daily, aes(x = date, y = lightStop), color =
"salmon3",
            shape = 1, size = 0.5) +
  scale_y_continuous(expand = c(0.02, 0.1),
                    limits = c(0, 24),
                    breaks = c(0, 6, 12, 18, 24),
                    labels = c("", "6", "12", "18", "24")) +
  scale_x_date(breaks = "2 month", date_labels = "%b-%y") +
  labs(x = "", y = "'Time (Hours)", title = "Dawn, Dusk & light
duration") +
  theme_pt() +
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x = element_text(angle = 0, size = 10))

##__Plot Reaeration-----
reaeration_plot <- ggplot(plot_data, aes(x = date, y = k)) +
  geom_point(shape = 1, color = "grey60", size = 0.5) +
  geom_line(data = daily, aes(x = date,
                              y = predict(reaeloess, newdata = jday))) +
  geom_hline(yintercept = 0, color = "black") +
  geom_rug(data = daily[toNuke, ], aes(x = date), color = "red" ,
          length = unit(0.03, "npc"), sides = 'b') +

```

```

scale_y_continuous(expand = c(0.02, 0.1),
                   limits = c(0, quantile(plot_data$k,
                                           probs = 0.99,
                                           na.rm = T))) +
scale_x_date(breaks = "2 month", date_labels = "%b-%y") +
labs(x = "", y = "(per day)", title = "Reaeration") +
theme_pt() +
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x = element_text(angle = 0, size = 10))

##__Plot Respiration----
respiration_plot <- ggplot(plot_data, aes(x = date, y = Rgpersqmpday))
+
  geom_point(shape = 1, color = "grey60", size = 0.5) +
  geom_line(data = daily, aes(x = date,
                              y = predict(resploess, newdata = jday))) +
  geom_hline(yintercept = quantile(plot_data$Rgpersqmpday,
                                   probs = 0.01,
                                   na.rm = T),
            color = "black") +
  geom_rug(data = daily[toNuke, ], aes(x = date), color = "red" ,
          length = unit(0.03, "npc"), sides = 'b') +
  scale_y_continuous(expand = c(0.02, 0.1),
                    limits = c(quantile(plot_data$Rgpersqmpday,
                                         probs = 0.01,
                                         na.rm = T), 0)) +
  scale_x_date(breaks = "2 month", date_labels = "%b-%y") +
  labs(x = "", y = expression(gm^-2*day^-1), title = "Respiration") +
  theme_pt() +
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x = element_text(angle = 0, size = 10))

##__Plot GPPgpersqmpday----
gpp_plot <- ggplot(plot_data, aes(x = date, y = GPPgpersqmpday)) +
  geom_point(shape = 1, color = "grey60", size = 0.5) +
  geom_line(data = daily, aes(x = date,
                              y = predict(gpploess, newdata = jday))) +
  geom_hline(yintercept = quantile(plot_data$GPPgpersqmpday,
                                   probs = 0.01,
                                   na.rm = T),
            color = "black") +
  geom_rug(data = daily[toNuke, ], aes(x = date), color = "red" ,
          length = unit(0.03, "npc"), sides = 'b') +
  scale_y_continuous(expand = c(0.02, 0.1),
                    limits =
                      quantile(plot_data$GPPgpersqmpday,
                               probs = c(0.01, 0.99),
                               na.rm = T)) +
  scale_x_date(breaks = "2 month", date_labels = "%b-%y") +
  labs(x = "", y = expression(gm^-2*day^-1), title = "GPPgpersqmpday")
+
  theme_pt() +
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x = element_text(angle = 0, size = 10))

```

```

##__Plot NEM----
nem_plot <- ggplot(plot_data, aes(x = date, y = NEM)) +
  geom_point(shape = 1, color = "grey60", size = 0.5) +
  geom_line(data = daily, aes(x = date,
                              y = predict(nemloess, newdata = jday))) +
  geom_hline(yintercept = min(c(quantile(plot_data$NEM,
                                         probs = 0.01,
                                         na.rm = T), 0)),
            color = "black") +
  geom_rug(data = daily[toNuke, ], aes(x = date), color = "red" ,
          length = unit(0.03, "npc"), sides = 'b') +
  scale_y_continuous(expand = c(0.02, 0.1),
                    limits =
                      quantile(plot_data$NEM,
                                probs = c(0.01,0.99),
                                na.rm = T)) +
  scale_x_date(breaks = "2 month", date_labels = "%b-%y") +
  labs(x = "", y = expression(gm^-2*day^-1), title = "NEM") +
  theme_pt() +
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x= element_text(angle = 0, size = 10))

##__Plot Relative reliability----
rr_plot <- ggplot(plot_data, aes(x = date, y = reliability)) +
  geom_point(shape = 1, color = "grey60", size = 0.5) +
  geom_line(data = daily, aes(x = date,
                              y = predict(relioloess, newdata = jday))) +
  scale_y_continuous(expand = c(0.02, 0.1),
                    limits = c(0, 1)) +
  scale_x_date(breaks = "2 month", date_labels = "%b-%y") +
  labs(x = "", y = "Reliability", title = "Relative reliability") +
  theme_pt() +
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x= element_text(angle = 0, size = 10))

##__Plot depth----
depth_plot <- ggplot(plot_data, aes(x = date, y = depth)) +
  geom_point(shape = 1, color = "grey60", size = 0.5) +
  geom_line(data = daily, aes(x = date,
                              y = predict(depthloess, newdata = jday))) +
  scale_y_continuous(expand = c(0.02, 0.1),
                    limits = c(0, quantile(daily$depth,
                                           probs = 0.99,
                                           na.rm = T))) +
  scale_x_date(breaks = "2 month", date_labels = "%b-%y") +
  labs(x = "", y = "Depth", title = "Depth") +
  theme_pt() +
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x= element_text(angle = 0, size = 10))

#create a grid plot of all metabolic plost

```

```

plot_grid <- suppressWarnings(grid.arrange(dark_duration,
                                           light_duration,
                                           reaeration_plot,
                                           respiration_plot,
                                           gpp_plot,
                                           nem_plot,
                                           rr_plot,
                                           depth_plot,
                                           nrow = 8))

#save plot of site parameters
ggsave(filename = paste0("./", sitename, ' - Metab Parameters.tif'), plot
= plot_grid,
        device = 'tiff', width = 10, height = 12, dpi = 300, units = 'in',
        type = 'cairo', compression='lzw' )

#_Plot Metabolic distribution----
##__Plot Reaeration distribution----
plot_data_dist <- as_tibble(density(plot_data$k, na.rm = T, from =
0)[c("x", "y")])

reaeration_dist <- ggplot(plot_data_dist, aes(x, y)) +
  geom_line() +
  scale_x_continuous(limits = quantile(plot_data$k,
                                      probs = c(0.01,0.99),
                                      na.rm = T)) +
  labs(y = "Density", x = "(per day)", title = "Reaeration") +
  theme_pt()

##__Plot Respiration distribution----
plot_data_dist <- as_tibble(density(plot_data$Rgpersqmpday, na.rm = T,
to = 0)[c("x", "y")])

respiration_dist <- ggplot(plot_data_dist, aes(x, y)) +
  geom_line() +
  scale_x_continuous(limits = quantile(plot_data$Rgpersqmpday,
                                      probs = c(0.01,0.99),
                                      na.rm = T)) +
  labs(y = "Density",
       x = expression(gm^-2*day^-1),
       title = "Respiration") +
  theme_pt()

##__Plot GPPgpersqmpday distribution----
plot_data_dist <- as_tibble(density(plot_data$GPPgpersqmpday, na.rm =
T, from = 0)[c("x", "y")])

gpp_dist <- ggplot(plot_data_dist, aes(x, y)) +
  geom_line() +
  scale_x_continuous(limits = quantile(plot_data$GPPgpersqmpday,
                                      probs = c(0.01,0.99),
                                      na.rm = T)) +

```



```

  labs(y = "Density",
        x = expression(gm-2*day-1),
        title = "GPPgpersqmpeday") +
  theme_pt()

##__Plot NEM distribution----
plot_data_dist <- as_tibble(density(plot_data$NEM, na.rm = T)[c("x",
"y")])

nem_dist <- ggplot(plot_data_dist, aes(x, y)) +
  geom_line() +
  scale_x_continuous(limits = quantile(plot_data$NEM,
                                      probs = c(0.01,0.99),
                                      na.rm = T)) +

  labs(y = "Density",
        x = expression(gm-2*day-1),
        title = "NEM") +
  theme_pt()

##__Plot Relative reliability distribution----
plot_data_dist <- as_tibble(density(plot_data$reliability,
                                    from = 0,
                                    to = 1,
                                    na.rm = T)[c("x", "y")])

rr_dist <- ggplot(plot_data_dist, aes(x, y)) +
  geom_line() +
  labs(y = "Density",
        x = "Reliability",
        title = "Relative reliability") +
  theme_pt()

##__Plot Depth distribution----
plot_data_dist <- as_tibble(density(plot_data$depth,
                                    na.rm = T)[c("x", "y")])

depth_dist <- ggplot(plot_data_dist, aes(x, y)) +
  geom_line() +
  labs(y = "Density",
        x = "mm",
        title = "Depth") +
  theme_pt()

#create a grid plot of all metabolic plost
plot_grid <- suppressWarnings(grid.arrange(reaeration_dist,
                                           respiration_dist,
                                           gpp_dist,
                                           nem_dist,
                                           rr_dist,
                                           depth_dist,
                                           nrow = 6))

```

```

#save plot of site parameters
ggsave(filename = paste0("./", sitename, ' - Metab Distn.tif'),
        plot = plot_grid, device = 'tiff', width = 10, height = 12, dpi =
300,
        units = 'in', type = 'cairo', compression='lzw' )

#_Plot DO minimum----
do_plot <- ggplot(contData, aes(x = date, y = DO_mg_data)) +
  geom_point(size = 0.15, alpha = 0.1, color = "gray60") +
  geom_line(aes(x = date, y = dailyMin), size = 0.4, color = "red")+
  geom_point(aes(x = date, y = day7meanMin), color = "blue", size = 0.75)
+
  labs(x = "", y = expression(Oxygen~"(*gm^-3~)"),
        title = paste0(sitename, " DO concentration time series plot"))+
  theme_pt() +
  theme(axis.text = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        axis.text.x= element_text(angle = 0, size = 12))

#save DO minimum plot
filename <- paste0("./",sitename, " - DO minimum.tif")

ggsave(filename = paste0("./",sitename, " - DO minimum.tif"),
        plot = do_plot, device = 'tiff', width = 10, height = 12, dpi =
300,
        units = 'in', type = 'cairo', compression='lzw' )

cat('+\\n')
}

# Save summary data as a csv
write.csv(sumStat, file = "Statistical summary of metabolic results.csv",
        row.names = F)

#prepare DMinima Data frame
DMinima <- DMinima%>%drop_na(Min7Day)
DMinima$Min7Day <- round(DMinima$Min7Day,2)
DMinima$Min1Day <- round(DMinima$Min1Day,2)

#arrange by site
DMinima <- DMinima %>%
  arrange(site)

#save DO minima data frame
write.csv(DMinima, './SOE DO minimum Stats.csv', row.names = F)

```

```

#####
####
#_Plot distribution----
#to match colors of plot as in the report
colour_df <- data.frame(colours = c("red", "orange", "green", "blue",
"magenta", "grey50"),
                        site = c("Hatea", "Mangere", "Opouteke", "Otaika",
"Punakitere", "Waitangi"))

#define data frame to capture store distribution data across all rivers
all_dist_data <- NULL

for(dosite in dosites$`Script Site name`){

  #Get site data
  daily_site <- allsite_distcomp %>%
    filter(site == dosite)

  # get k distribution data for site using user defined stDensity function
  dist_data <- stDensity(daily_site$k, from = 0)

  all_dist_data <- rbind(all_dist_data, data.frame(x = dist_data$x,
                                                  y = dist_data$y,
                                                  site = dosite,
                                                  var = "k"))

  ## get Rgpersqmpeday distribution data for site using user defined
  #stDensity function
  dist_data <- stDensity(daily_site$Rgpersqmpeday, to = 0)

  all_dist_data <- rbind(all_dist_data, data.frame(x = dist_data$x,
                                                  y = dist_data$y,
                                                  site = dosite,
                                                  var = "Rgpersqmpeday"))

  ## get GPPgpersqmpeday distribution data for site using user defined
  #stDensity function
  dist_data <- stDensity(daily_site$GPPgpersqmpeday, from = 0)

  all_dist_data <- rbind(all_dist_data, data.frame(x = dist_data$x,
                                                  y = dist_data$y,
                                                  site = dosite,
                                                  var =
"GPPgpersqmpeday"))

  ## get NEM distribution data for site using user defined stDensity
function
  dist_data <- stDensity(daily_site$NEM)

  all_dist_data <- rbind(all_dist_data, data.frame(x = dist_data$x,

```

```

        y = dist_data$y,
        site = dosite,
        var = "NEM"))

## get reliability distribution data for site using user defined
stDensity
#function
dist_data <- stDensity(daily_site$reliability, from = 0, to = 1)

all_dist_data <- rbind(all_dist_data, data.frame(x = dist_data$x,
        y = dist_data$y,
        site = dosite,
        var = "reliability"))

## get depth distribution data for site using user defined stDensity
function
dist_data <- stDensity(daily_site$depth, from = 0)

all_dist_data <- rbind(all_dist_data, data.frame(x = dist_data$x,
        y = dist_data$y,
        site = dosite,
        var = "depth"))

## get dark duration distribution data for site using user defined
#stDensity function
dist_data <- stDensity((daily_site$darkStop - daily_site$darkStart)%%24)

all_dist_data <- rbind(all_dist_data, data.frame(x = dist_data$x,
        y = dist_data$y,
        site = dosite,
        var = "dark_duration"))

## get light duration distribution data for site using user defined
#stDensity function
dist_data <- stDensity((daily_site$lightStop -
daily_site$lightStart)%%24)

all_dist_data <- rbind(all_dist_data, data.frame(x = dist_data$x,
        y = dist_data$y,
        site = dosite,
        var = "light_duration"))

}

#define data frame to capture plot parameters to match plots generated in
#Envirolink report
plot_paratemers <- data.frame(var = c("k", "Rgpersqmpeday",
        "GPPgpersqmpeday",
        "NEM", "reliability", "depth"),
        xmin = c(-2, -30, -30, -30, 0, 0),
        xmax = c(30, 30, 30, 30, 1, 1250),
        xbreak = c(5, 10, 10, 10, 0.2, 200),

```

```

      xlab = c("(per day)", "expression", "expression",
"expression",
      "Relative reliability", "mm"),
      titlab = c("Reaeration k", "Respiration",
"GPPgpersqmp/day",
      "NEM", "Reliability", "Water Depth"),
      xint = c(0, 0, 0, 0, 0, -10))

for(i in 1:nrow(plot_paratemers)){

  #extract parameters from var data frame
  vari <- plot_paratemers$var[i]
  xmin <- plot_paratemers$xmin[i]
  xmax <- plot_paratemers$xmax[i]
  xbreak <- plot_paratemers$xbreak[i]
  xlab <- ifelse(plot_paratemers$xlab[i] == "expression",
expression(gm^-2*day^-1),
plot_paratemers$xlab[i])
  titlab <- plot_paratemers$titlab[i]
  xint <- plot_paratemers$xint[i]

  #filter data

  plot_data <- all_dist_data %>%
filter(var == vari)

  # only first figure should have legend, test if var is k else remove
  legend
  if(vari == "k"){

    figure <- ggplot(plot_data, aes(x = x, y = y, color = site)) +
geom_line(size = 0.75)+
geom_vline(xintercept = xint, color = "grey40", size = 0.75,
linetype = "dashed") +
scale_x_continuous(limits = c(xmin, xmax),
breaks = seq(from = xmin+2, to = xmax, by =
xbreak),
expand = c(0,0))+
scale_color_manual(values = colour_df$colours, labels =
colour_df$site) +
labs(x= xlab, y = "Density", title = titlab) +
theme_pt() +
theme(legend.position = c(0.80, 0.70),
legend.direction = "horizontal") +
guides(colour = guide_legend(nrow = 2))

  }else{

    figure <- ggplot(plot_data, aes(x = x, y = y, color = site)) +
geom_line(size = 0.75)+
geom_vline(xintercept = xint, color = "grey40", size = 0.75,
linetype = "dashed") +
scale_x_continuous(limits = c(xmin, xmax),
breaks = seq(from = xmin, to = xmax, by = xbreak),
expand = c(0,0))+

```

```
    scale_color_manual(values = colour_df$colours, labels =
colour_df$site) +
    labs(x= xlab, y = "Density", title = titlab) +
    theme_pt()+
    theme(legend.position = "none")

}

#create object, to allow plot to be placed on grid
eval(parse(text=paste0("f", i,"<- figure")))

suppressWarnings(print(figure))

}

#create a grid
dis_grid <- suppressWarnings(grid.arrange(f1, f2, f3, f4, f5, f6, nrow =
6))

#save plot
ggsave(filename = paste0('./RiverComparison.tif'), plot = dis_grid,
        device = 'tiff', width = 10, height = 12, dpi = 300, units = 'in',
        type = 'cairo', compression='lzw' )

#plot duration distribution
plot_paratemers <- data.frame(var = c("dark_duration", "light_duration"),
                             titlab = c("Dark duration", "Light
duration"))

for(i in 1:nrow(plot_paratemers)){

  #extract parameters from var data frame
  vari <- plot_paratemers$var[i]
  titlab <- plot_paratemers$titlab[i]

  #filter data

  plot_data <- all_dist_data %>%
    filter(var == vari)

  # only first figure should have legend, test if var is k else remove
legend
  if(vari == "dark_duration"){

    figure <- ggplot(plot_data, aes(x = x, y = y, color = site)) +
      geom_line(size = 0.75)+
      scale_x_continuous(limits = c(0, 24),
                        breaks = seq(from = 0, to = 24, by = 4),
                        expand = c(0,0))+
      scale_color_manual(values = colour_df$colours, labels =
colour_df$site) +
```

```
      labs(x= "Hours", y = "Density", title = titlab) +
      theme_pt() +
      theme(legend.position = c(0.90, 0.80))

    }else{

      figure <- ggplot(plot_data, aes(x = x, y = y, color = site)) +
        geom_line(size = 0.75)+
        scale_x_continuous(limits = c(0, 24),
                           breaks = seq(from = 0, to = 24, by = 4),
                           expand = c(0,0))+
        scale_color_manual(values = colour_df$colours, labels =
colour_df$site) +
        labs(x= "Hours", y = "Density", title = titlab) +
        theme_pt() +
        theme(legend.position = "none")

    }

    #create object, to allow plot to be placed on grid
    eval(parse(text=paste0("f", i,"<- figure")))

    suppressWarnings(print(figure))

  }

#create a grid
dis_grid <- suppressWarnings(grid.arrange(f1, f2, nrow = 2))

#save plot
ggsave(filename = paste0('./River Durations Comparison.tif'), plot =
dis_grid,
        device = 'tiff', width = 10, height = 12, dpi = 300, units = 'in',
        type = 'cairo', compression='lzw' )
```

## APPENDIX 2. RECORDING SHEET AND DEPTH MEASUREMENTS USED TO ESTIMATE MEAN REACH DEPTH UPSTREAM OF EACH DO LOGGER DEPLOYMENT SITE.

Site Name	Site ID	Date	Time (NZST)	Field personnel	Transect #	Wetted width (m)	NZTM (E)	NZTM (N)	View 1	View 2	View 3	View 4	View 5	View 6	View 7	View 8	View 9	View 10
Hatea at Whareora Road	LOC.005538	22-12-20	7:50	Sam, Suzannah	1	7.5	1720293	6048372	0.51	0.61	0.55	0.65	0.86	0.85	0.81	0.78	0.71	0.47
Hatea at Whareora Road	LOC.005538	22-12-20	7:50	Sam, Suzannah	2	8	1720303	6048410	0.09	0.08	0.25	0.26	0.35	0.28	0.18	0.07	0.03	0.09
Hatea at Whareora Road	LOC.005538	22-12-20	7:50	Sam, Suzannah	3	7.5	1720334	6048444	0.11	0.16	0.04	0.16	0.1	0.09	0.21	0.21	0.11	0.16
Hatea at Whareora Road	LOC.005538	22-12-20	7:50	Sam, Suzannah	4	11	1720357	6048459	0.47	0.32	0.32	0.34	0.27	0.19	0.18	0.23	0.22	0.17
Hatea at Whareora Road	LOC.005538	22-12-20	7:50	Sam, Suzannah	5	12.5	1720382	6048501	0.31	0.35	0.43	0.74	1	1.08	1.03	1.08	0.82	0.45
Mangere at Night Road	LOC.101625	18-12-20	9:05	Scott, Sam, Manas	1	10	1703625	6048969	0.55	0.95	1.2	1	1.2	1.4	1.5	1.6	1.14	0.83
Mangere at Night Road	LOC.101625	18-12-20	9:05	Scott, Sam, Manas	2	15	1703677	6048977	0.32	0.9	1.77	2.17	2.42	2.43	1.7	1.86	1.8	1.1
Mangere at Night Road	LOC.101625	18-12-20	9:05	Scott, Sam, Manas	3	10	1703687	6049010	0.54	1.25	1.44	1.47	1.68	1.81	1.95	2.1	2.04	0.55
Mangere at Night Road	LOC.101625	18-12-20	9:05	Scott, Sam, Manas	4	9	1703674	6049053	0.98	2.15	2.23	2.06	1.9	1.76	1.46	1.18	0.74	0.17
Mangere at Night Road	LOC.101625	18-12-20	9:05	Scott, Sam, Manas	5	11	1703717	6049078	0.82	1.24	1.56	1.54	1.25	1.2	0.98	1.01	0.58	0.48
Oputeke at Suspension Bridge	LOC.102258	18-12-20	11:10	Scott, Sam, Manas	1	19	1678505	6049533	0.26	0.35	1.3	1.5	1.35	1.58	1.5	1.36	1.1	0.31
Oputeke at Suspension Bridge	LOC.102258	18-12-20	11:10	Scott, Sam, Manas	2	24	1678496	6049488	0.91	0.97	0.83	1.04	1.25	1.6	2.3	1.8	1.93	1.31
Oputeke at Suspension Bridge	LOC.102258	18-12-20	11:10	Scott, Sam, Manas	3	8	1678524	6049442	0.32	0.34	0.51	0.56	0.66	0.59	0.54	0.48	0.49	0.41
Oputeke at Suspension Bridge	LOC.102258	18-12-20	11:10	Scott, Sam, Manas	4	10	1678523	6049406	0.15	0.29	0.38	0.41	0.35	0.57	0.5	0.55	0.57	0.5
Oputeke at Suspension Bridge	LOC.102258	18-12-20	11:10	Scott, Sam, Manas	5	11	1678504	6049352	0.1	0.13	0.24	0.33	0.37	0.44	0.63	0.73	0.52	0.18
Otaika at Otaika Valley Road	LOC.110431	18-12-20	7:48	Scott, Sam, Manas	1	6	1715513	6039913	0.18	0.69	0.6	0.58	0.5	0.38	0.27	0.17	0.13	0.09
Otaika at Otaika Valley Road	LOC.110431	18-12-20	7:48	Scott, Sam, Manas	2	8	1715515	6039905	0.4	0.29	0.22	0.22	0.28	0.49	0.49	0.48	0.46	0.42
Otaika at Otaika Valley Road	LOC.110431	18-12-20	7:48	Scott, Sam, Manas	3	7	1715502	6039930	0.5	0.51	0.52	0.5	0.62	0.73	0.86	1.12	1.3	1.37
Otaika at Otaika Valley Road	LOC.110431	18-12-20	7:48	Scott, Sam, Manas	4	4	1715483	6039922	0.56	0.72	0.71	0.66	0.55	0.45	0.36	0.33	0.28	0.11
Otaika at Otaika Valley Road	LOC.110431	18-12-20	7:48	Scott, Sam, Manas	5	8	1715460	6039921	0.22	0.73	0.74	0.68	0.47	0.25	0.3	0.45	0.52	0.14
Punakitere at Taheke	LOC.105231	21-12-20	10:30	Scott, Sam, Manas	1	16	1659944	6075428	0.04	0.16	0.31	0.58	0.67	0.58	0.8	0.69	0.67	0.48
Punakitere at Taheke	LOC.105231	21-12-20	10:30	Scott, Sam, Manas	2	16	1659955	6075459	0.38	0.38	0.32	0.52	0.38	0.53	0.39	0.44	0.28	0.09
Punakitere at Taheke	LOC.105231	21-12-20	10:30	Scott, Sam, Manas	3	14	1659994	6075484	0.2	0.29	0.35	0.44	0.42	0.41	0.49	0.45	0.21	0.05
Punakitere at Taheke	LOC.105231	21-12-20	10:30	Scott, Sam, Manas	4	17	1660014	6075533	0.26	0.38	0.37	0.51	0.65	0.72	0.92	0.83	0.93	0.95
Punakitere at Taheke	LOC.105231	21-12-20	10:30	Scott, Sam, Manas	5	14	1660024	6075587	0.28	0.38	0.49	0.75	0.8	0.77	0.56	0.28	0.16	0.22
Waitangi at Waimate North Road	LOC.103178	21-12-20	8:45	Scott, Sam, Manas	1	6.5	1682008	6093696	0.21	0.37	0.37	0.41	0.4	0.39	0.32	0.19	0.14	0.12
Waitangi at Waimate North Road	LOC.103178	21-12-20	8:45	Scott, Sam, Manas	2	5	1681976	6093712	0.63	0.73	0.82	0.92	1.03	1.1	0.96	0.99	0.15	0.17
Waitangi at Waimate North Road	LOC.103178	21-12-20	8:45	Scott, Sam, Manas	3	4	1681954	6093731	0.6	0.64	0.5	0.37	0.36	0.37	0.38	0.44	0.42	0.38
Waitangi at Waimate North Road	LOC.103178	21-12-20	8:45	Scott, Sam, Manas	4	5	1681921	6093744	0.64	0.61	0.57	0.57	0.61	0.71	0.77	0.72	0.28	0.15
Waitangi at Waimate North Road	LOC.103178	21-12-20	8:45	Scott, Sam, Manas	5	4.5	1681901	6093771	0.6	0.2	0.22	0.27	0.37	0.47	0.49	0.28	0.22	0.18



## APPENDIX 3. TIME SERIES PLOTS OF DAILY METABOLISM RESULTS

### A3.1 Hatea at Whareora Road

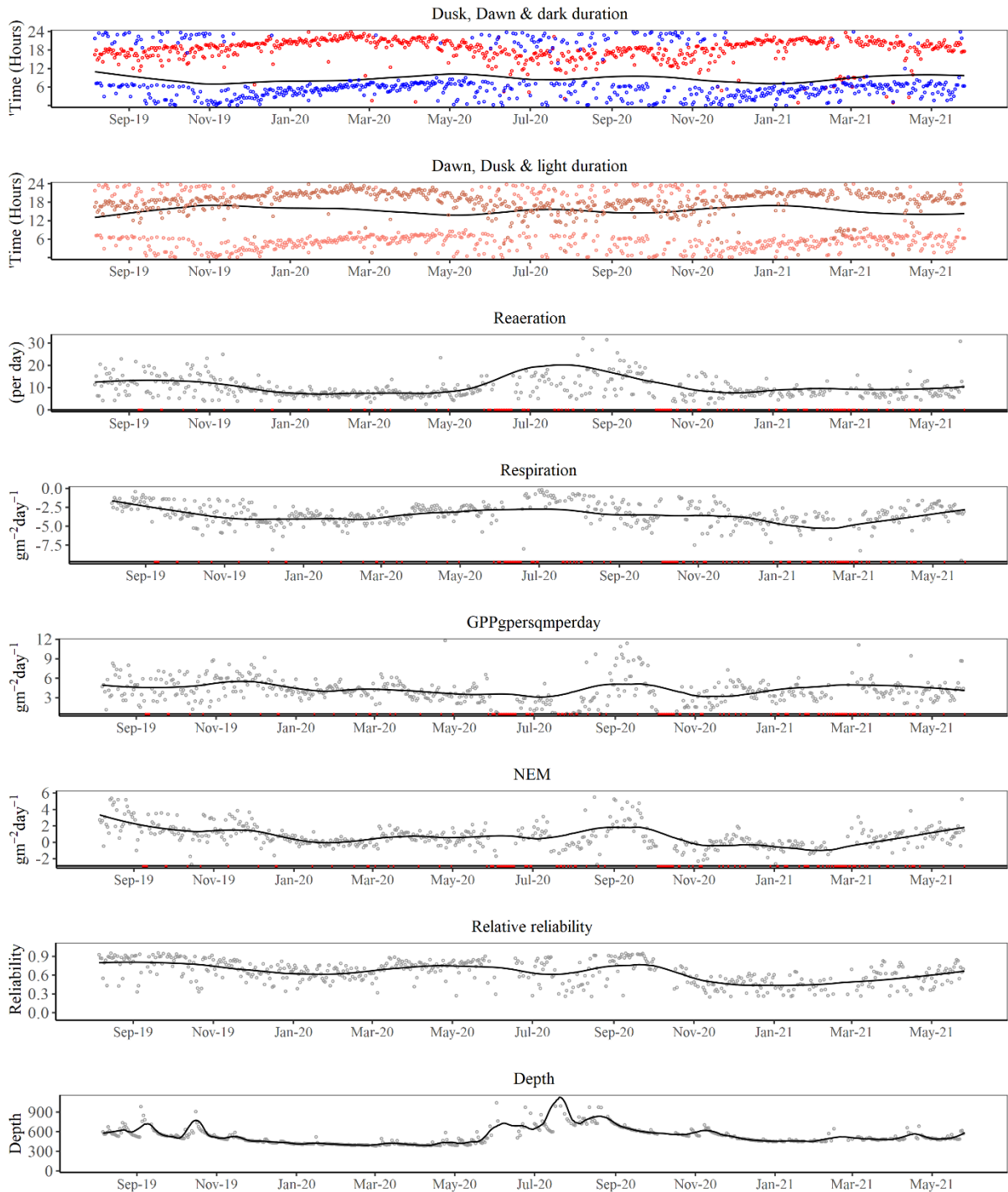


Figure A3.1. Metabolic estimate series for Hatea. Red marks on the x axis indicate days where metabolism estimates were not included in subsequent analyses because the relative reliability was below the 0.25 threshold or where estimates were considered as spurious.



### A3.2 Mangere at Knights Road

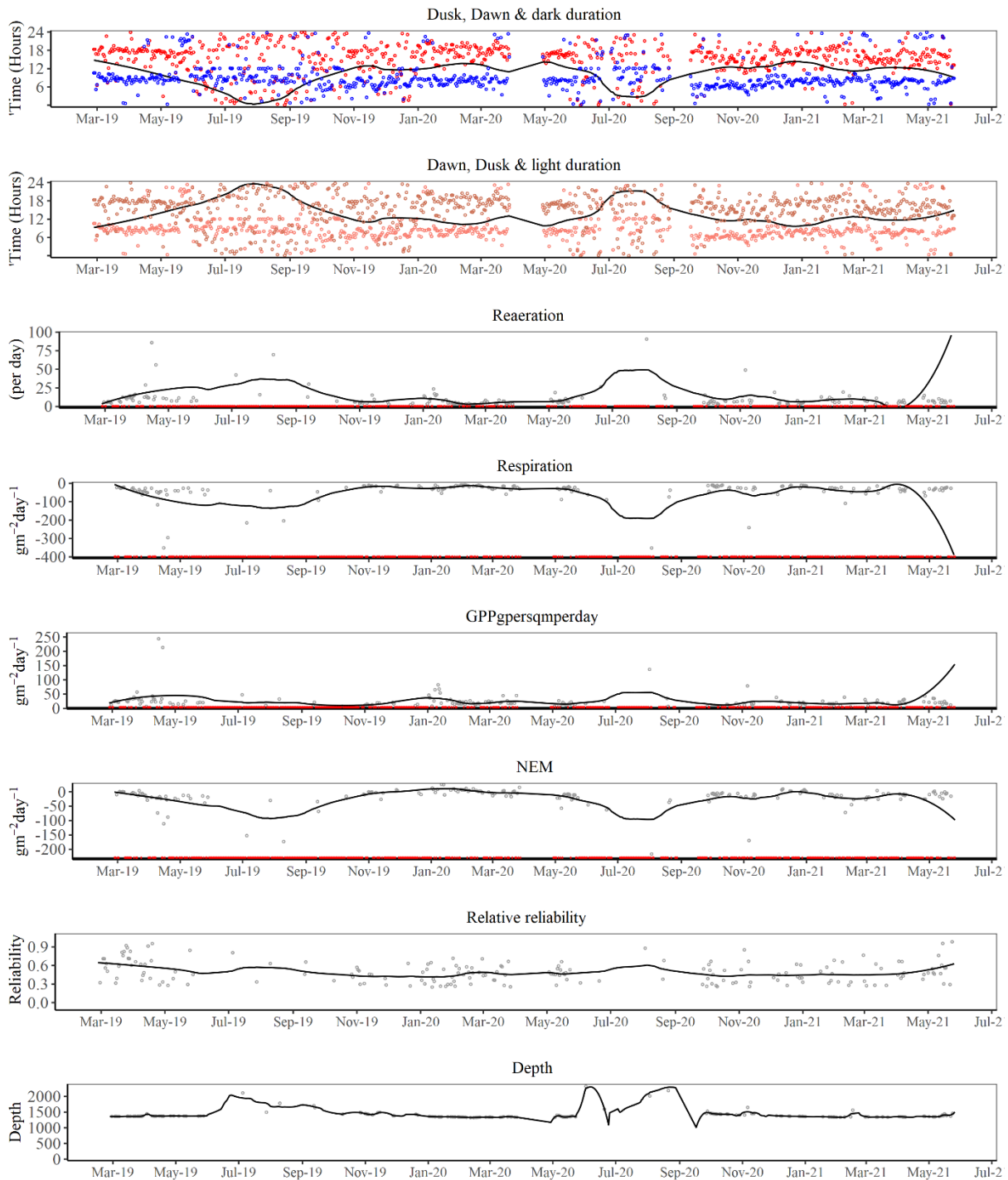


Figure A3.2. Metabolic estimates series for Mangere at Knights Road. Red marks on the x axis indicate days where metabolism estimates were not included in subsequent analyses because the relative reliability was below the 0.25 threshold or where estimates were considered as spurious.

### A3.3 Opouteke at Suspension Bridge

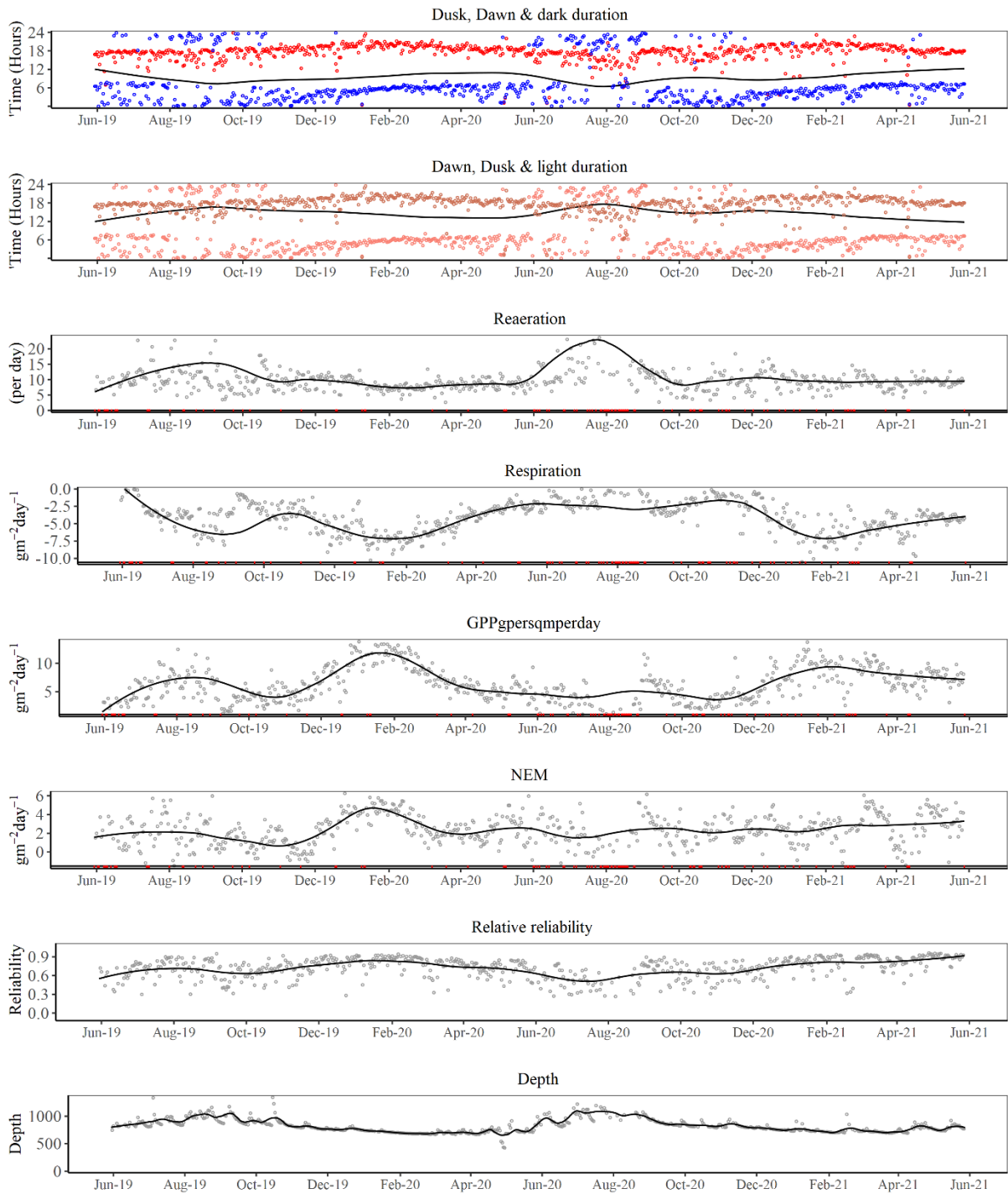


Figure A3.3. Metabolic estimates series for Opouteke at Suspension Bridge. Red marks on the x axis indicate days where metabolism estimates were not included in subsequent analyses because the relative reliability was below the 0.25 threshold or where estimates were considered as spurious.

### A3.4 Otaika at Otaika Valley Road

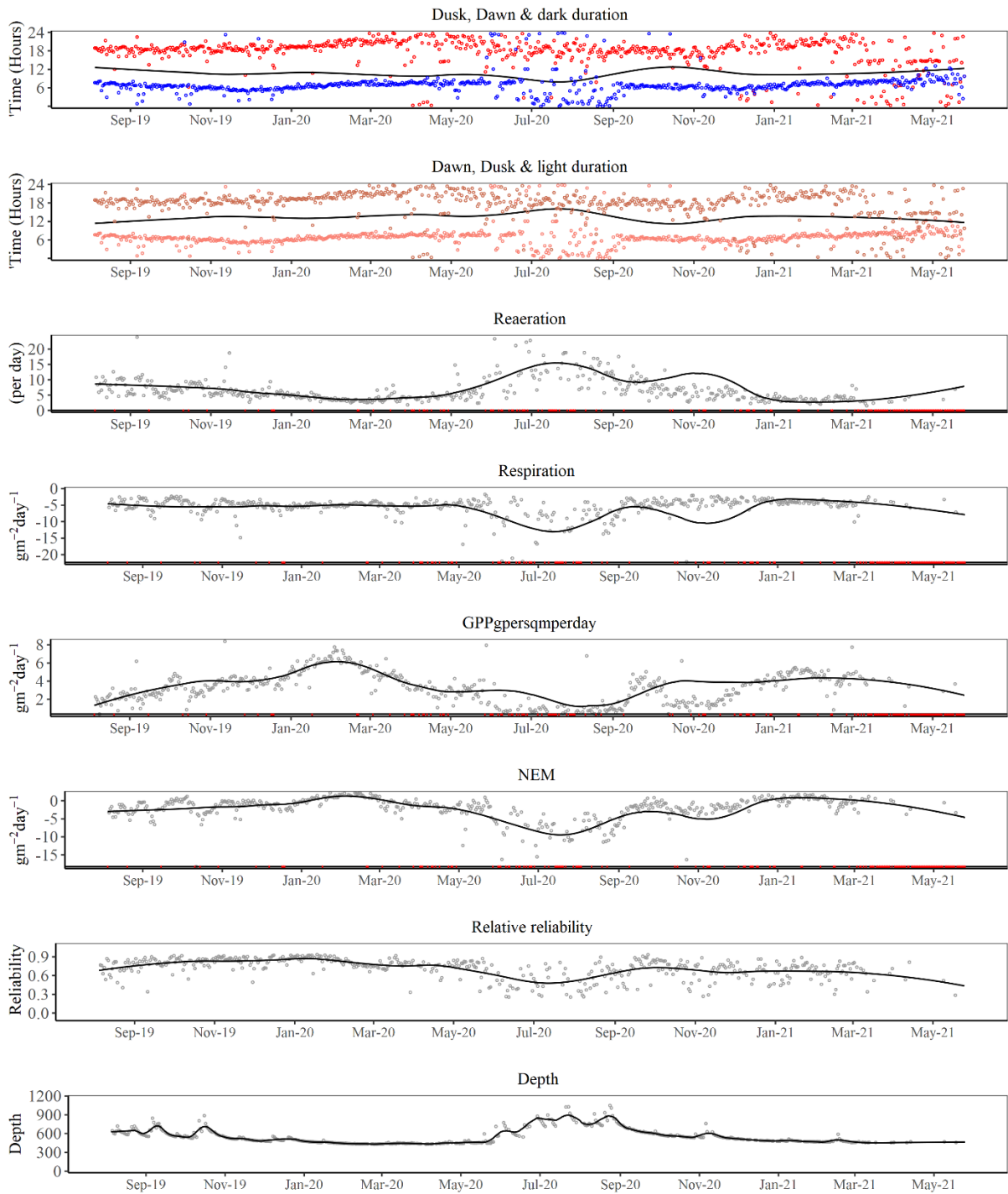


Figure A3.4. Metabolic estimates series for Otaika at Otaika Valley Road. Red marks on the x axis indicate days where metabolism estimates were not included in subsequent analyses because the relative reliability was below the 0.25 threshold or where estimates were considered as spurious.

### A3.5 Punakitere at Taheke

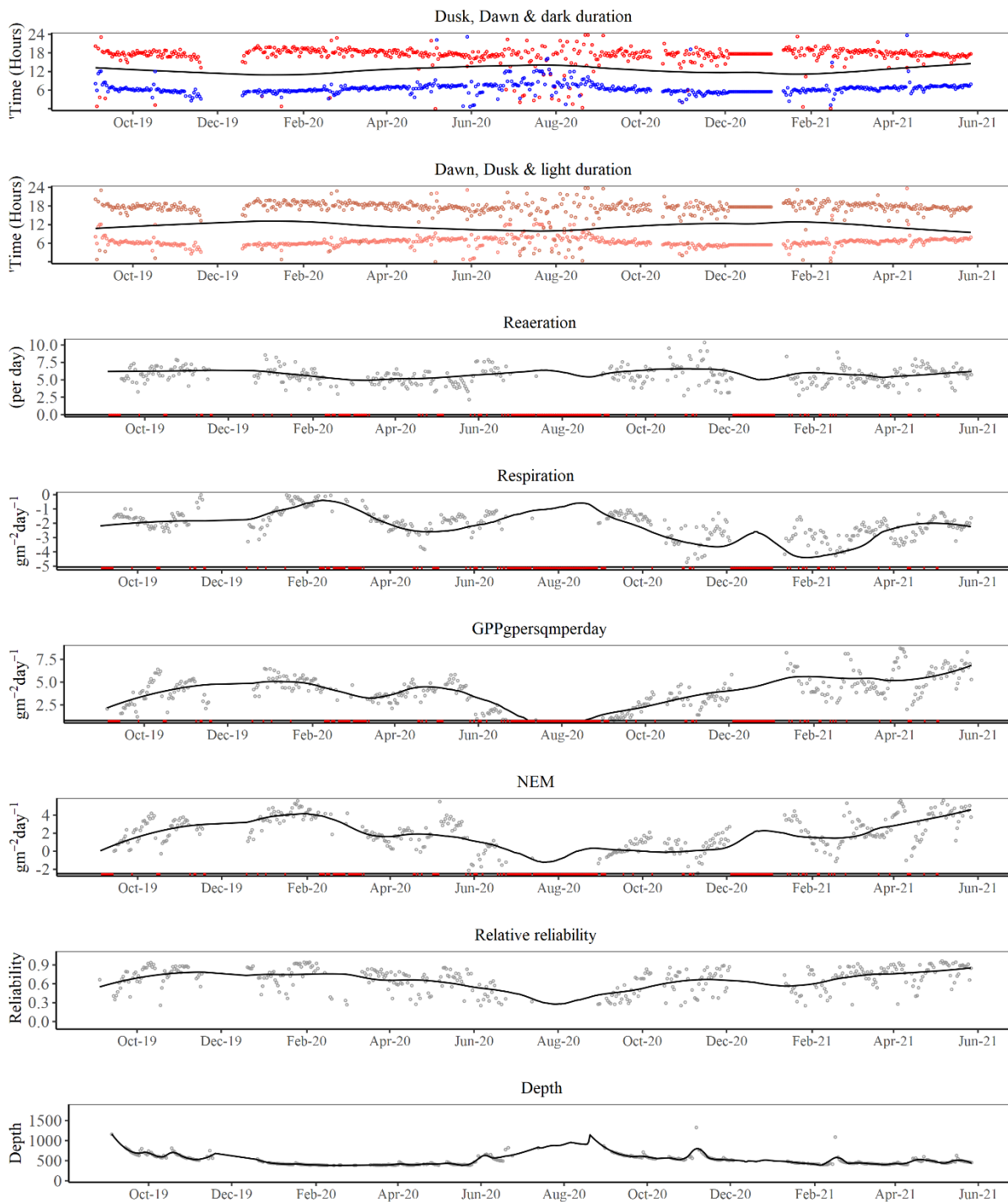


Figure A3.5. Metabolic estimates series for Punakitere at Taheke. Red marks on the x axis indicate days where metabolism estimates were not included in subsequent analyses because the relative reliability was below the 0.25 threshold or where estimates were considered as spurious.

### A3.6 Waitangi at Waimate North Road

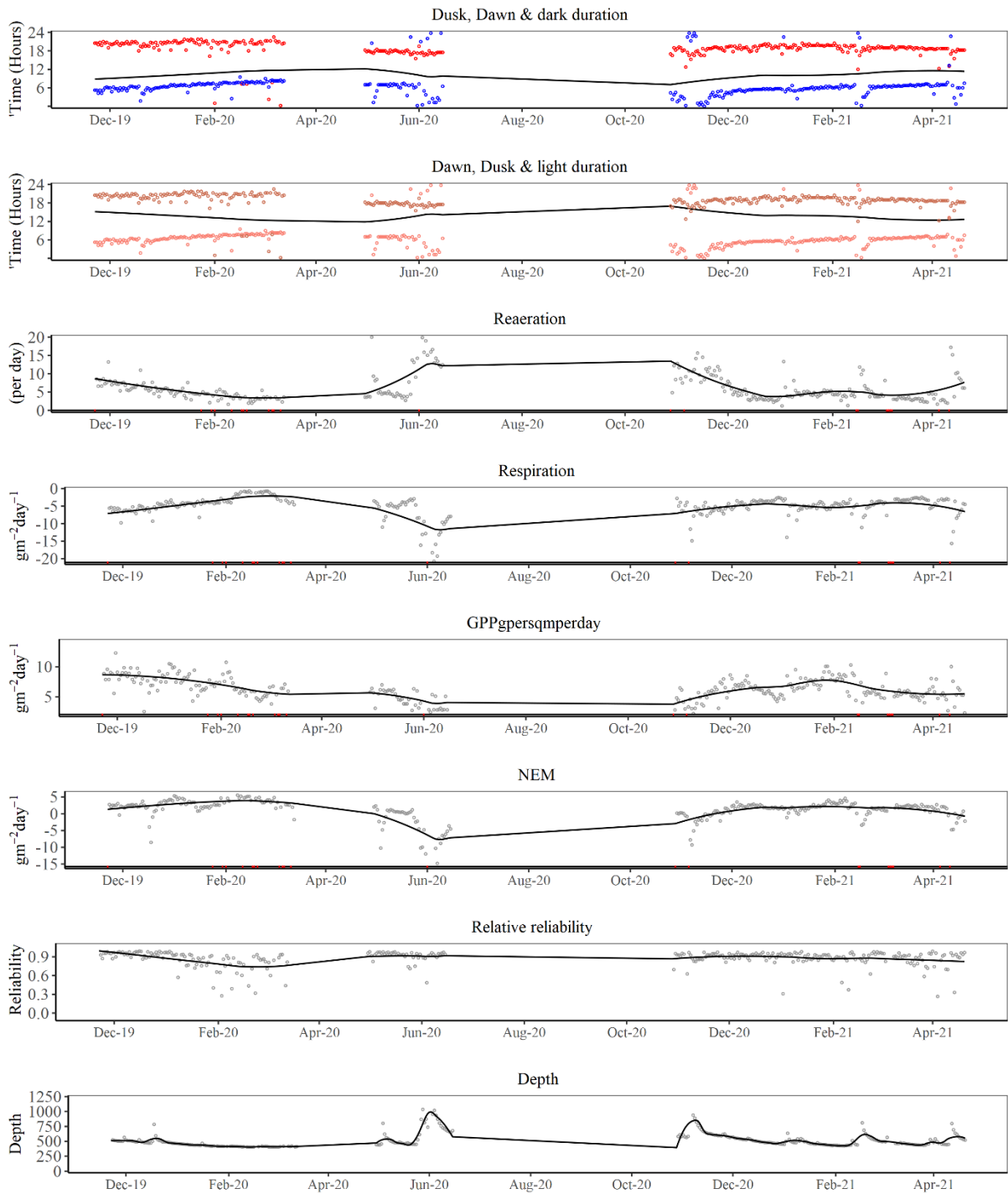


Figure A3.6. Metabolic estimates series for Waitangi at Waimate North Road. Red marks on the x axis indicate days where metabolism estimates were not included in subsequent analyses because the relative reliability was below the 0.25 threshold or where estimates were considered as spurious.