Periphyton – Environment Relationships in the Horizons Region

March 2018

Horizons Report 2018/EXT/1600

Prepared for:

Abby Matthews
Science & Innovation Manager

March 2018
Report No. 2018/EXT/1600
ISBN 978-1-98-853756-6

Prepared by:

Cathy Kilroy, Michelle Greenwood, Janine Wech,
Tom Stephens (Dairy NZ), Logan Brown (Horizons),
Abby Matthews (Horizons), Maree Patterson (Horizons),
Mike Patterson (Horizons)

Client Report No. 2018123CH
project DNZ17501

For any information regarding this report please contact:
Cathy Kilroy
Freshwater Ecologist
+64-3-343 7883
cathy.kilroy@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 8602
Riccarton
Christchurch 8011
Phone +64 3 348 8987

Periphyton - environment relationships in the Horizons region

Analysis of a seven-year dataset

Prepared for DairyNZ and Horizons Regional Council

March 2018

Prepared by:
Cathy Kilroy
Michelle Greenwood
Janine Wech
Tom Stephens (Dairy NZ)
Logan Brown (Horizons)
Abby Matthews (Horizons)
Maree Patterson (Horizons)
Mike Patterson (Horizons)

For any information regarding this report please contact:

Cathy Kilroy
Freshwater Ecologist

+64-3-343 7883
cathy.kilroy@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 8602
Riccarton
Christchurch 8011

Phone +64 3 348 8987

NIWA CLIENT REPORT No: 2018123CH
Report date: March 2018
NIWA Project: DNZ17501

Quality Assurance Statement		
	Reviewed by:	Clive Howard-Wiliams
	Formatting checked by:	Fenella Falconer
	Approved for release by:	Helen Rouse

Contents

Tables

Figures

Summary

This report describes a set of analyses on time-series data (at least monthly) of periphyton and associated environmental variables collected by Horizons Regional Council (hereafter Horizons) at over 60 river sites thoughout the Manawatu-Whanganui region since late 2008. This dataset is the longest and most comprehensive of its type in New Zealand and is probably unusual worldwide in terms of length of coverage and data resolution. Accompanying hydrological data (a continuous flow record) was available for 50 of the sites.

The research was planned and executed as a collaboration between DairyNZ, Horizons and NIWA and was jointly funded by DairyNZ and Horizons. Its aim was to provide information to assist in better management of the impacts of periphyton on river health and other river values throughout the region, thereby contributing to fulfilling community expectations for water quality (including periphyton), as established within Horizons' One Plan. The contract for the work specified three objectives:

- Objective 1 – Establish the significance and strength of relationships between environmental factors and periphyton standing crop (max or 92nd% Chl-*a* between stations, observed Chl-*a* within station time-series);
- Objective 2 – Establish if the resolution of sampling affects the performance of environmental drivers identified in Objective 1;
- Objective 3 – Classify stations on the basis of their within-station environmental drivers, to permit generalisation of earlier driver findings, and comparison of findings to between-station inferences across all stations.

This summary comprises a compliation of the key messages from different components of the research. The messages are also presented at the beginning of nine sections of the report, following a summary of the available data.

Sensitivity of periphyton to flows at each site

- The aim of the analysis was to quantify the effect of flows on periphyton at each of 50 sites in the Horizons dataset with a linked flow record and enough data to identify differences between sites.
- Sites were classified into flow-sensitive and flow-insensitive sites based on the periphyton–flow relationships. Flow-sensitive sites were those at which more than 20% of variance in periphyton chlorophyll *a* was explained by accrual period for a distinct flow threshold. At flow-insensitive sites either accrual period explained less than 20% of the variance in chlorophyll *a*, or no distinct threshold was identiifiable.
- At 31 sites, we were able to define a distinct threshold for the size of flow event that would typically remove periphyton to a low level. At three more sites, removal typically occurred over a wide range of flows.
- The thresholds varied between 1.5 and 15 x median flow.
- We compiled a new variable from this analysis: the ***effective flow***, and accrual period calculated from the effective flow (i.e., the time periphyton has to grow without being washed away in a high flow). This is a new idea: up until now a rule-

of-thumb has been that events >3 x median flow generally remove periphyton in rivers.

- Accrual period calculated using the effective flow explained up to 53% of the variance in chlorophyll *a* within a site (and over 40% at 14 of the 42 sites), supporting previous research conclusions, that flow variability is commonly the dominant driver of variability in periphyton.
- Effective flow thesholds were exceeded for between 4% and 33% of the time at flow-sensitive sites with low flow thesholds (up to 5 x median flow) and for lower percentages of time (0.6 to 4.2%) at sites with thresholds less than 5 x median flow.
- Previous work using the Horizons dataset has identified that hydraulic and geomorphological characteristics determine the effective flow. Simple field techniques to determine the effective flow at new sites are under development.

Patterns of periphyton nutrient limitation

- The aim of the analysis was to use the dissolved inorganic N (DIN) and dissolved reactive P (DRP) data to assess potential nutrient limitation of periphyton growth at each site in the dataset. N or P limitation occurs if one or other of these nutrients is in short supply. In that case additions of that nutrient could cause periphyton to increase.
- We assessed limitation by looking at DIN : DRP ratios and DIN and DRP concentrations. DIN and DRP can be correlated with river flow, usually positively. Therefore, we also took this effect of river flow into account.
- Based only on DIN : DRP ratios and at all flows and seasons (including sites with no linked flow record), 50% of the sites were P-limited (mostly in the Manawatu, Makotuku, Mangawhero and Ohau Rivers); 33% were N-limited (most headwater sites, and all sites in the Rangitikei River), and 17% were limited by both N and P (co-limited).
- At over 60% of the 47 sites with a flow record and enough data, highest DIN typically occurred in high flows, when periphyton is being sloughed.
- When samples collected during high flows were excluded, three sites with flow records shifted from P-limitation to co-limitation, and five from co-limitation to N-limitation. In this smaller dataset, 51% of sites were P-limited, 40% N-limited, 9% co-limited.
- When concentrations of saturating DIN and DRP (i.e., enough for maximum growth rates) were taken into account as well as flows, 28% of sites (with flow records) were assessed as predominantly P-limited, 8.5% as N-limited, 55% as co-limited, and 8.5% as limited by neither N nor P.
- Regardless of how it was calculated, the limiting nutrient varied over time at all sites. This implies that additions of either N or P could potentially stimulate periphyton growth at different times, over much of the regional monitored stream network.

- Seasonality of nutrient limitation was not accounted for in the analysis, but further analysis of this aspect of variability would likely be useful.

Test of the Biggs (2000a) relationships

- We used the Horizons periphyton dataset as independent data to test published equations linking annual maximum chlorophyll *a* to DIN or DRP concentrations and accrual period calculated using the frequency of flows exceeding 3 x median flow (Biggs 2000a). Our aim was to determine whether prediction and ultimately management of periphyton chlorophyll *a* simply requires knowledge of DIN and DRP concentrations and accrual period. A further question was whether using accrual period based on ***effective flow*** would improve predictions.

- Across all sites, predictions of maximum chlorophyll *a* from the Biggs (2000a) equations were only weakly or not correlated with observed chlorophyll *a*. Relationships between nutrient concentrations and periphyton standing crop are unlikely to be accurately characterised in the Horizons region by the Biggs (2000a) equations.

- Using the effective flow to calculate accrual period did not improve the predictions.

- The Biggs (2000a) equations were expected to perform weakly given that almost half the annual mean DIN values in the Horizons dataset exceeded the range underpinning the relationship. Predicting beyond the range of the original data is unlikely to be accurate.

- The Biggs (2000a) equations were derived using data from a smaller range of river type than is found in the Horizons region, where hydro-physical characteristics are variable.

- The weak performance of the Biggs (2000a) equations in predicting annual maximum chlorophyll *a* across the Horizons region indicated the need for new predictive relationships, with additional variables considered for inclusion.

Between-site relationships

- Aims were to explore (a) correlations between peak chlorophyll *a* and averaged DIN or DRP across sites and years; and (b) relationships between peak chlorophyll *a* and a combination of environmental variables, in both cases applying a space-for-time approach using linear regression. We tested relationships in annual and 3-year datasets and a 7-year dataset.

- "Peak chlorophyll *a*" was annual maximum chlorophyll *a* for annual datasets, and the 92nd percentile of chlorophyll *a*, the for multi-year datasets. The latter is the metric used in the periphyton attribute of the National Policy Statement for Freshwater Management (NPS-FM).

- For aim (a):

  - DIN was significantly and positively related to peak chlorophyll *a* in most time periods. Relationships were especially strong across sites classed as flow-insensitive (i.e., where a threshold for flows effective for removing periphyton could not be identified);

- despite significant relationships, separate tests of predictive ability (cross-validation tests) showed that mean DIN in isolation from other variables was not a good predictor of peak chlorophyll *a* across sites within the Horizons region;
- mean DRP was weakly or not correlated with peak chlorophyll *a* in all periods.

▪ For aim (b):

- in addition to mean DIN and DRP, potential predictor variables included water conductivity, river bed sediment composition, mean water temperature, and mean accrual period (based on both 3 x median and effective flow);
- generally, the strongest models in each time period included ***DIN, conductivity*** and ***accrual period*** as predictors. The initial models explained at least 50% of the variance in peak chlorophyll *a* across sites in all time periods;
- accrual period calculated from the effective flow always produced stronger relationships than accrual period from 3 x median flow;
- some models also included terms for substrate, water temperature and DRP;
- leave-one-out cross-validation (a robust method for evaluating the predictive ability of models) produced encouraging results, with high proportions of variance in observed chlorophyll *a* explained by predicted chlorophyll *a* for some periods (e.g., 75% in 2012 - 2015).
- Models for the 3-year datasets performed better than the annual datasets with means of 63% vs. 55% explained respectively across all 3-year and annual periods, and 67% for the 7-year dataset;
- substituting total nitrogen (TN – all N in a sample including organic particles) for DIN produced slightly stronger relationships, which, again, were optimised if accrual period based on effective flow was included;
- substituting land-cover variables such as percentage of the catchment under intensive farmland (which is correlated with mean DIN) for DIN did not improve the models;
- the best models included all of the available sites (not smaller subsets).

▪ The models may be useful for (a) predicting likely chlorophyll *a* at new sites or at the same sites under different scenarios, such as reduced flood frequency or increased nutrient concentrations; and (b) setting nutrient limits. The error in each model was determined.

▪ Conductivity was highlighted in all strongly performing models as having a positive effect on chlorophyll *a*. This points to either a direct effect on periphyton chlorophyll *a* (e.g., via algal community composition) and/or a positive feedback into other factors such as nutrient availability. Conductivity was weakly associated with DIN across the region, suggesting that the cause of conductivity variation was not strongly linked to catchment DIN losses (e.g., from land use practices) but was more likely a function of underlying catchment hydrogeology.

Within-site analyses

- A long periphyton dataset (>7 years) may enable an alternative approach to determining factors associated with variation in periphyton standing crop, by exploring relationships over time within sites.
- Explanatory variables were coeval, lagged and averaged DIN and DRP (over the previous 4 and 6 months). Lagged and averaged data were included because periphyton on a particular date has been influenced by preceding conditions.
- Regardless of the DIN metric used, most relationships between chlorophyll *a* and DIN were negative: high chlorophyll *a* was associated with low DIN.
- Negative correlations persisted even when the data were filtered to remove samples associated with high flows (i.e., when DIN tends to be high but chlorophyll *a* is low because of flood-removal).
- At some sites, low DIN at times of high periphyton could indicate high rates of uptake of DIN from the water. For example, at sites in the lower Rangitikei River, the negative correlation between chlorophyll *a* and DIN became stronger as data associated with high flows were removed from the dataset.
- Correlations between chlorophyll *a* and DRP were much weaker than for DIN and were positive or negative (but with low coefficient of determination).
- Reducing the dataset to annual peak chlorophyll *a* revealed shifts in the direction of the relationships between chlorophyll *a* and DIN or DRP from negative/neutral (using all data) to more positive. For DIN, 4% of sites with positive correlations increased to 33%; for DRP 22% increased to 41%.
- Adding in other variables (water temperature, accrual period) to predictive relationships for chlorophyll *a* (using all of the data) led to reasonably strong predictive models at some sites (e.g., cross-validated R^2 up to 0.6), although some models still included negative terms for DIN or DRP or both.
- Accrual period based on the effective flow was the only predictor that operated consistently across sites (using the between-site approach) and within sites, with a positive effect on chlorophyll *a*.
- There was no clear and simple linear relationship between periphyton standing crop and nutrient availability (as DIN or DRP) throughout the year, either across all flows, or in low flows only .

Effect of using fortnightly vs monthly datasets

- Horizons has collected periphyton data at fortnightly intervals at a subset of the monitoring sites. Data from 12 sites were used to compare the predictive ability of within-site models derived from fortnightly and monthly datasets at these sites. Data were available at each site from between 17 and 24 months.

- The models from fortnightly data performed similarly to or better than the models using monthly data in 11 of the 12 sites tested. Only at ohau_gladstone did monthly data predict periphyton biomass more accurately than fortnightly.
- Poor performance of monthly data over the period of fortnightly surveys (17 to 24 months) may be attributable to low numbers of samples.
- The 17- to 24-month time series of fortnightly data from the 12 sites generally did not yield stronger predictive relationships than using the complete (up to 7 years) monthly dataset at the same sites (although noting that the datasets were not strictly comparable in numbers of samples or variables included).
- We concluded that fortnightly data in some cases could allow relationships to be developed over a shorter time period. Fortnightly datasets have other applications including more accurate estimation of accrual rates, and testing of mechanistic models of periphyton growth.

Relationships between chlorophyll *a* and percentage cover

- In addition to data on periphyton chlorophyll *a*, data on periphyton cover were available from all sites, in six categories (bare rock, film, sludge, mats, slimy green filaments, other (coarse) filaments). For the analysis, sludge and mats were combined into "Mats", and slimy green and other (coarse) filaments into "Fils".
- Correlations between chlorophyll *a* and metrics of percentage cover were investigated using between-site and within-site approaches. The purpose of the analysis was to see whether it is possible to make robust conversions from visual estimates to chlorophyll *a*. If that proved to be the case, then it can be inferred that management of the environmental factors that affect chlorophyll *a* will apply to visual cover by periphyton in an equivalent way.
- We explored relationships between mean and maximum chlorophyll *a* and cover, between sites and within sites, using a range of cover metrics (in particular, weighted composite cover (WCC) and the combination of Film, Mats and Fils in a multiple regression).
- For the between-site analysis, predictive ability of the relationships with mean chlorophyll *a* varied across years and was often poor (NSE < 0.3), with the exception of later years (2013–14, 2014–15). Relationships were broadly equivalent in performance between WCC, Mats or Fils, and in general were weaker for annual maxima then annual mean cover estimates.
- Within sites, the multiple regression using Film, Mats and Fils produced the strongest relationships with chlorophyll *a*. 44% of sites had RSE (a measure of model performance) > 0.55. All the sites with strong predictive ability were in the wider Manawatu catchment, or in the Ohau catchment, and did not include headwater sites.

- We concluded that use of a single region-wide relationship to predict chlorophyll *a* from cover is unlikely to be robust for the Horizons dataset. Only within the mid- to lower Manawatu River and Ohau River did strong chlorophyll *a* – cover relationships suggest that the drivers of chlorophyll *a* are likely to have corresponding effects on percentage cover by periphyton.

Classification of sites and implications for predicting chlorophyll *a* and setting nutrient limits

- We aimed to determine the scope for grouping sites, based on output from earlier chapters in this report, to discriminate between differing site responses of chlorophyll *a* (e.g., on the basis of effective flow, correlations between measures of standing crop, within-site relationshps with environmental factors, nutrient limitation status, conductivity, geology and catchment land use).

- Sites were assigned to groups (i.e., classified) on the basis of site characteristics, catchment characteristics, and the variables included in, or strength of, within-site relationships between chlorophyll *a* and environmental variables. We also considered the strenths of within-site relationships between chlorophyll *a* and periphyton cover.

- The strongest pattern noted was that sites with strong within-site relationships between chlorophyll *a* and cover also had the strongest within-site relationships with environmental variables (including with accrual time (days since an effective flow)). These sites included most sites in large rivers, had higher DIN and finer sediment, and were in catchments with high proportions of their area in farmland and low proportions in indigenous forest. All these variables were generally intercorrelated and it was not clear what was driving the pattern.

- Grouping sites by their within-site chlorophyll *a* – environment relationships did not generate a pattern aligned with catchment geology or (life-supporting capacity (LSC) class. Treating sites within each LSC or geological class alike in terms of management actions is therefore unlikely to deliver equivalent periphyton (chlorophyll *a*) outcomes.

- However, the strengths of the within-site relationships again showed some patterns across catchment geology classes. Sites with AL (alluvium) and SS (soft sedimentary) geology had stronger within-site relationships than those with HS (hard sediemtnary) or VA (volcanic acidic) geology.

1 Introduction

Horizons Regional Council (Horizons) commenced monthly monitoring of periphyton cover and biomass at 48 river sites in the Manawatu-Whanganui Region in late 2008. The monitoring programme had multiple aims, including assessment of regional compliance with periphyton targets specified in the One Plan (http://www.horizons.govt.nz/about-us/one-plan/) and development of a regional model for predicting periphyton at unmonitored river sites and in response to catchment changes.

The data were reviewed after one and three years of data collection, as joint projects involving Horizons staff and NIWA (Kilroy et al. 2010, 2012). These reports each addressed several questions including compliance with the Proposed One Plan (at the time), accuracy of chlorophyll *a* determination, relationships between periphyton, flows and other environmental variables, and conversion of visually assessed periphyton cover into a chlorophyll *a* equivalent. In a further report after six years, the data were used to (a) assess the state of periphyton in the region's rivers relative to regional targets and national standards / thresholds; and (b) identify any trends (declining or increasing cover or chlorophyll *a*) from the data (Kilroy et al. 2016). That report was jointly funded by Horizons and DairyNZ.

With over seven years of data now available, the Horizons dataset is the most comprehensive of its type in New Zealand and possibly elsewhere in the world. The number of sites included in the programme has increased over the years to at least 61 currently, on 25 rivers. There is therefore considerable scope to expand the analyses carried out in previous years, and to verify and begin to explain the patterns observed. Starting in late 2016, and based on a preliminary analysis of the six-year dataset (referred to in Kilroy et al. (2016)), DairyNZ and NIWA worked together to define an approach for analysing the seven-year dataset, to be carried out with input from Horizons staff. The research was planned and executed as a collaboration between DairyNZ, Horizons and NIWA. Its aim was to provide information to assist in better management of the impacts of periphyton on river health and other river values throughout the region, thereby contributing to fulfilling community expectations for water quality (including periphyton), as established within the One Plan.

The discussions with Dairy NZ resulted in a structured approach to the analysis under three objectives:

- Objective 1 – Establish the significance and strength of relationships between environmental factors and periphyton standing crop (max or 92nd% Chl-*a* between stations, observed Chl-*a* within station time-series).
- Objective 2 – Establish if the resolution of sampling affects the performance of environmental drivers identified in Objective 1.
- Objective 3 – Classify stations on the basis of their within-station environmental drivers, to permit generalisation of earlier driver findings, and comparison of findings to between-station inferences across all stations.

The detailed wording in the contract is reproduced as Appendix A. As the work proceeded, we reviewed the suggested methods and in some cases revised the approach, considering (a) the features of the dataset, and (b) literature related to the methods. Appendix B provides background information on the specified analytical approaches, issues encountered during the analysis, and our approaches to solving the issues.

In this report, the results of the analyses are reported in approximately the order listed in the contract (Table 1-1). A minor change is that classification of sites in terms of their nutrient limitation status is presented in its own section so that the classification can be referred to in subsequent sections. Sections in the report are as follows.

Following this Introduction (Section 1), Section 2 comprises a summary of the data available, including information on missing data, and a list of all explanatory variables used in subsequent analyses.

Section 3 describes the analyses carried out to determine the hydrological sensitivity of periphyton at each site.

Section 4 uses the data from all sites to explore patterns of nutrient limitation status. This part of the analysis was specified as part of the between-site analysis, particularly the test of existing relationships. The resulting classification is used in both within- and between-site analyses.

Table 1-1: Summary of tasks in the contract in relation to Sections in the report. The contract wording is reproduced in Appendix A.

Objective	Task	Sub-task	Section in report	Comments
Objective 1. Between and within station analyses	1a. Effect of river flows on periphyton		3	See also Section 10 for discussion of classification
	1b. Between-station analyses	Chlorophyll *a* vs. cover	9.2, 9.3	
		Test of Biggs (2000) relationships	5	
		Chlorophyll *a* vs. DIN and DRP	6.2	
		Patterns of N or P limitation and effect on relationships	4	Limitation in own section then cross-referenced
		Multiple stepwise linear regression	6.3	See also Appendix B
		Quantile regression		See Appendix B
	1c. Within-station relationships	Chlorophyll *a* vs. cover	9.4, 9.5	
		Chlorophyll *a* vs. DIN and DRP	7.2	
		Multiple stepwise linear regression	7.3	See also Appendix B
		Quantile regression		See Appendix B
Objective 2. Effect of sampling resolution			9	
Objective 3. Classification of sites			10	See Appendix K for classifications at all sites

Section 5 covers testing the ability of equations in Biggs (2000a) to accurately predict annual maximum periphyton. Reasons for failure of the equations to provide realistic predictions at many sites are discussed.

Section 6 describes between-site analyses carried out to detect relationships that could be potentially be used to predict chlorophyll *a* metrics (summarised from monthly time-series) at new sites, or at the same sites in different periods.

Section 7 covers within-site analyses to detect within-site relationships between periphyton and nutrients; then moving to regression using multiple variables.

Section 8 describes periphyton chlorophyll *a* vs. percentage cover relationships as a stand-alone analysis that relates to the chlorophyll *a* – environment relationships by addressing the question: will the information on drivers of chlorophyll *a* (from Section 6) be useful in managing percentage cover?

Section 9 comprises analyses carried out to determine the effect on outcomes of using data collected at fortnightly intervals rather than data collected at the usual monthly intervals.

Section 10 summarises site classifications that have emerged from the analysis, and discusses use of the classifications to understand the effects of nutrients (DIN and DRP) on periphyton and to assist in river management.

Section 11 provides a brief synthesis of all of the results in relation to the three main objectives in the contract.

2 Data availability and preparation

Key messages

- The Horizons periphyton monitoring dataset comprises monthly chlorophyll *a* and periphyton visual assessments from over 60 monitoring sites, starting in December 2008.
- Additional environmental data collected at the time of periphyton surveys includes bed substrate cover assessments, nutrient concentrations (nitrogen, phosphorus), conductivity, water temperature, and other water quality variables.
- A continuous flow record is available for over 50 of the monitoring sites.
- Higher resolution data (fortnightly surveys) have been collected at a subset of the sites (currently 13 sites that also have a flow record for periods of up to two years.
- The length and detail of the programme makes the dataset the most comprehensive of its type in New Zealand, and probably worldwide.

Locations of all sites in the Horizons periphyton monitoring programme are shown in Figure 2-1. A list of all monitored sites with locations, Horizons' site classifications, and the date of the first periphyton survey is presented in Table 2-1. Throughout this report we refer to individual sites by their site abbreviation, for brevity. In Figures and Tables, sites are generally listed in order of their Horizons site number, which arranges sites within catchments, from upstream to downstream.

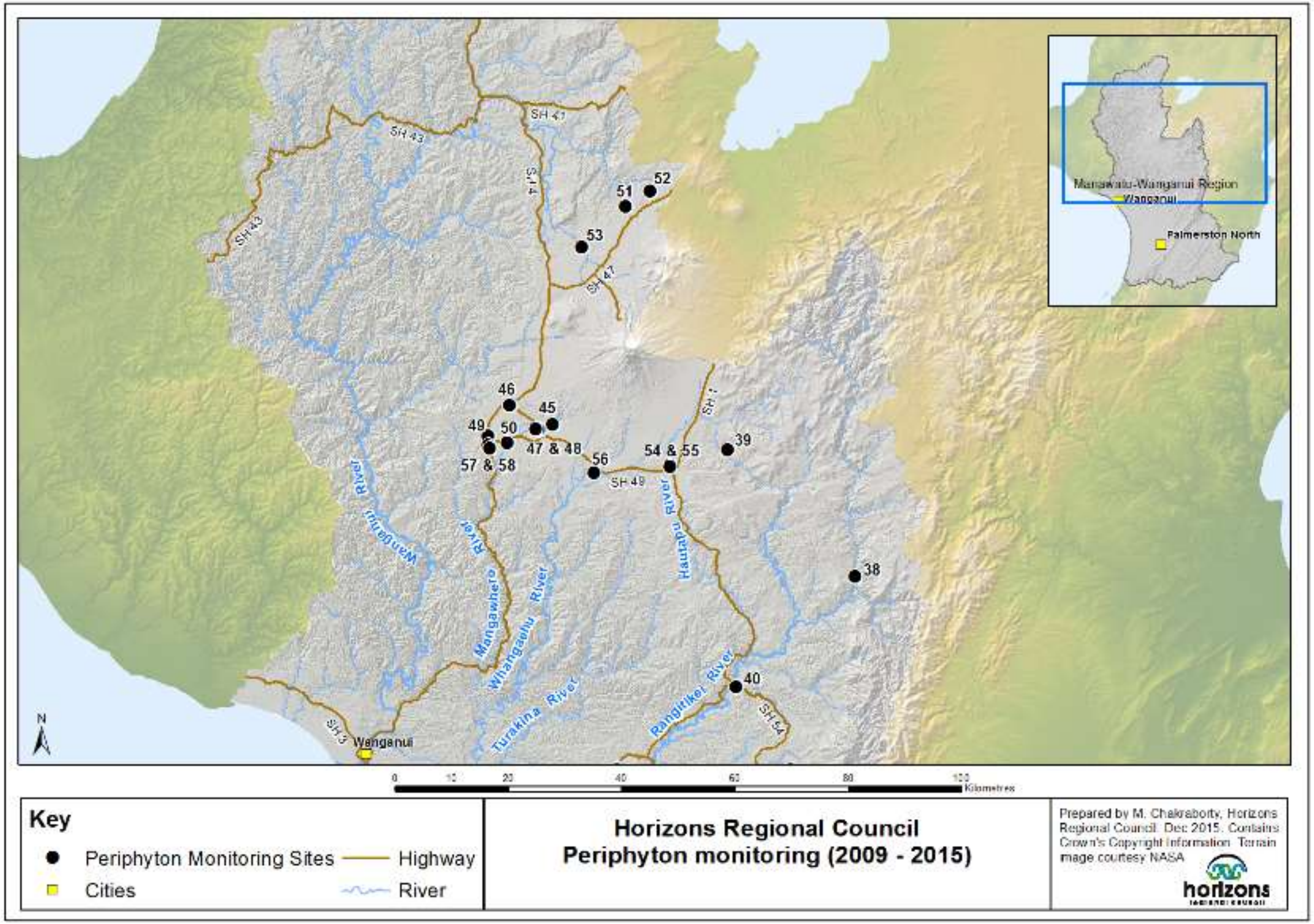
2.1 Periphyton data

2.1.1 The Horizons periphyton dataset

Data on chlorophyll *a*, periphyton visual assessments and bed substrate cover from 66 monitoring sites were provided by Horizons. Periphyton data collection at most sites started in December 2008, with data supplied to April 2017. For detailed sample collection methods refer to Kilroy et al. (2016).

One of the 66 sites was omitted from the analyses because of a short record (tiraumea_ds_mangat, less than 2 years data). A further four sites were omitted because they had no associated water quality data and only a few periphyton observations (see below). Tokiahuru at Karioi (tokiahuru_kar) was included in some analyses but we note that there is no data later than October 2014 at this site.

All sites had dates some dates with missing data, when flows were too high or water clarity too low to conduct a survey or collect samples. The overall rate of missing data (across all sites) was 10% for chlorophyll *a* samples, and 17% for periphyton visual assessments. The highest proportions of missing visual assessments data (up to 46%) were from sites in main stems of the Manawatu, Oroua, Rangitikei and Tiraumea Rivers.


SH 41
SH 43
SH 43
SH 4
SH 47
52
51
53
46
45
49
50
47 & 48
57 & 58
56
54 & 55
39
38
40
SH 1
SH 49
SH 3
SH 54
Wanganui River
Mangawhero River
Whangaehu River
Turakina River
Hautapu River
Rangitikei River
Wanganui
Manawatu-Wanganui Region
Wanganui
Palmerston North
N
0
10
20
40
60
80
100
Kilometres
Key
Periphyton Monitoring Sites
Highway
Cities
River
Horizons Regional Council
Periphyton monitoring (2009 - 2015)
Prepared by M. Chakraborty, Horizons Regional Council. Dec 2015. Contains Crown's Copyright Information. Terrain image courtesy NASA
horizons

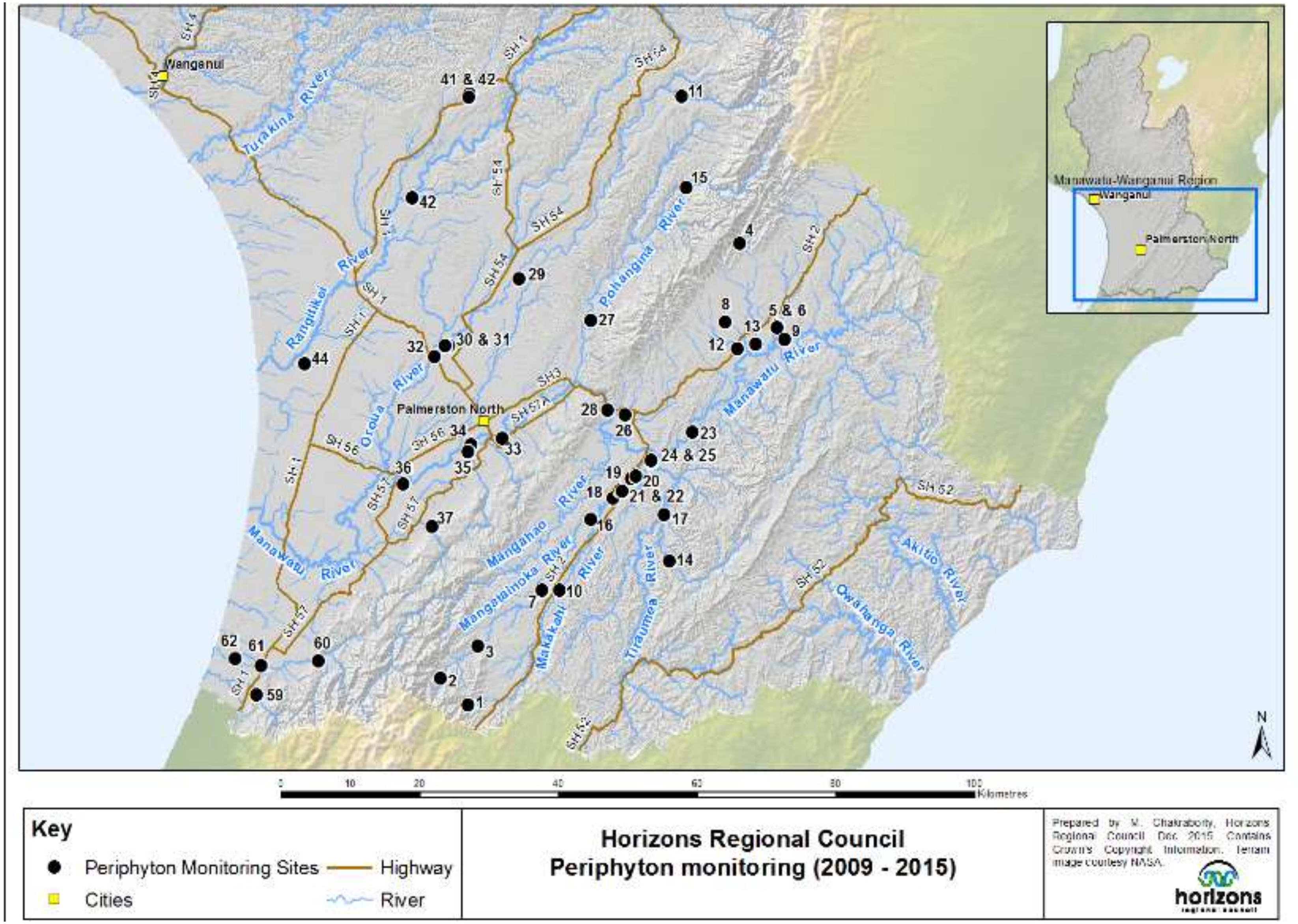


Figure 2-1: Locations of periphyton monitoring sites in the Manawatu-Whanganui Region northern area (top) and southern area (bottom). Site numbers on the maps are the Horizons site numbers listed in Table 2-1.

Table 2-1: List of periphyton monitoring sites used in the analysis, with location details. LSC class is the Life-supporting capacity class assigned by Horizons (see text) and Sub-region is the One Plan management unit for the site. PSD = yes means a point-source discharge is upstream of the site. The site abbreviation (*) is used in the text when referring to periphyton sites. Monitoring continued until April 2015 at all sites except for Site 25, tiraumea_ds_mangat. (**), where monitoring ceased in October 2010. Refer to map (Figure 2-1) for location within the region. Sites are in order of the HRC site number (HRCn), which sorts sites from upstream to downstream in successive catchments.

HRCn	Site name	Abbreviation*	E	N	LSC	Sub-zone	PSD	start	Flow site
1	Makakahi at DOC Reserve	makakahi_doc	2729456	6051399	HM	Mana_8d	no	13-Aug-13	Makakahi at Hamua
2	Mangatainoka at Putara	mangatainoka_putara	2725500	6055099	UHS	Mana_8a	no	9-Dec-08	Mangatainoka at Larsens Br
3	Mangatainoka at Larsons Road	mangatainoka_lars	2730878	6059626	UHS	Mana_8a	no	13-Aug-13	Mangatainoka at Larsons Road
4	Tamaki at Reserve	tamaki_res	2768598	6115899	UHS	Mana_3	no	10-Dec-08	Tamaki at Water Supply Weir
5	Mangatera u/s Dannevirke STP	mangatera_us_dan	2773957	6104367	HM	Mana_2b	no	10-Dec-08	NO FLOW SITE
6	Mangatera d/s Dannevirke STP	mangatera_ds_dan	2773970	6104182	HM	Mana_2b	yes	10-Dec-08	NO FLOW SITE
7	Mangatainoka at Hukanui	mangatainoka_huk	2740072	6067395	HM	Mana_8b	no	13-Aug-13	Mangatainoka at Larsons Road
8	Kumeti at Te Rehunga	kumeti_tr	2766500	6104991	UHS	Mana_4	no	10-Dec-08	Kumeti at Te Rehunga
9	Manawatu at Weber Road	manawatu_weber	2775096	6102500	HM	Mana_1a	no	10-Dec-08	Manawatu at Weber Rd
10	Makakahi at Hamua	makakahi_ham	2742599	6067399	HM	Mana_8d	no	9-Dec-08	Makakahi at Hamua
11	Oroua at Apiti Gorge	oroua_apiti	2760199	6136499	HM	Mana_12a	no	11-Dec-08	Oroua at Almadale
12	Tamaki at Stephensons	tamaki_ste	2770914	6101859	HM	Mana_5b	no	10-Dec-08	Tamaki at Stephensons
13	Oruakeretaki at SH2	oruakeretaki_sh2	2768237	6101204	HM	Mana_5d	no	10-Dec-08	Oruakeretaki at SH2(Napier)
14	Makuri at Tuscan Hills	makuri_tuscan	2758500	6071501	ULi	Mana_7d	no	19-Dec-08	Makuri at Tuscan Hills
15	Pohangina at Piripiri	pohangina_pir	2760843	6123817	UHS	Mana_10b	no	15-Dec-08	NO FLOW SITE
16	Mangatainoka at Scarborough Konini Road	mangatainoka_scarb	2747160	6077271	HM	Mana_8b	no	13-Aug-13	Mangatainoka at Larsens Br
17	Tiraumea at Ngaturi	tiraumea_nga	2757748	6077929	HSS	Mana_7b	no	19-Dec-08	Tiraumea at Ngaturi
18	Mangatainoka at Pahiatua Town Bridge	mangatainoka_pahiatua	2750283	6080248	HM	Mana_8c	no	13-Aug-13	Mangatainoka at Pahiatua
21	Mangatainoka u/s Pahiatua STP	mangatainoka_us_pah	2751269	6081437	HM	Mana_8c	no	9-Dec-08	Mangatainoka at Pahiatua
22	Mangatainoka d/s Pahiatua STP	mangatainoka_ds_pah	2751656	6081282	HM	Mana_8c	yes	9-Dec-08	Mangatainoka at Pahiatua
19	Mangatainoka at SH2	mangatainoka_sh2	2753015	6082998	HM	Mana_8c	no	9-Dec-08	Mangatainoka at Pahiatua
20	Mangatainoka d/s DB Breweries	mangatainoka_ds_db	2753600	6083400	HM	Mana_8c	yes	9-Dec-08	Mangatainoka at Pahiatua
23	Manawatu at Hopelands	manawatu_hop	2761799	6089499	HM	Mana_5a	no	11-Dec-08	Manawatu at Hopelands
24	Mangatainoka u/s Tiraumea confl	mangatainoka_us_tir	2755838	6085354	HM	Mana_8c	no	14-Jan-11	Mangatainoka at Pahiatua

HRCn	Site name	Abbreviation*	E	N	LSC	Sub-zone	PSD	start	Flow site
25	Tiraumea d/s Mangatainoka confl**	tiraumea_ds_mangat	2755829	6085578	HSS	Mana_7b	no	19-Dec-08	Tiraumea at Ngaturi
26	Mangapapa at Troup Road	mangapapa_troup	2752115	6092008	HM	Mana_9b	no	19-Dec-08	Mangapapa at Troup Rd
27	Pohangina at Mais Reach	pohangina_mais	2747118	6105154	HM	Mana_10c	no	15-Dec-08	Pohangina at Mais Reach
28	Manawatu at Upper Gorge	manawatu_ug	2749590	6092568	HM	Mana_9a	no	11-Dec-08	Manawatu at Upper Gorge
29	Oroua at Almadale	oroua_almadale	2736799	6110997	HM	Mana_12a	no	11-Dec-08	Oroua at Almadale
30	Oroua u/s Feilding STP	oroua_us_fei	2726681	6101660	HM	Mana_12b	no	11-Dec-08	NO FLOW SITE
31	Oroua d/s Feilding STP	oroua_ds_fei	2726109	6101599	HM	Mana_12b	yes	11-Dec-08	NO FLOW SITE
32	Oroua at Awahuri Bridge	oroua_awahuri	2724600	6100103	LM	Mana_12c	no	11-Dec-08	NO FLOW SITE
33	Manawatu at Teachers College	manawatu_tc	2734398	6088683	HM	Mana_10a	no	15-Dec-08	Manawatu at Teachers College
34	Manawatu u/s PNCC STP	manawatu_us_pncc	2729885	6087742	HM	Mana_11a	no	15-Dec-08	Manawatu at Teachers College
35	Manawatu d/s PNCC STP	manawatu_ds_pncc	2729400	6086801	HM	Mana_11a	yes	15-Dec-08	Manawatu at Teachers College
36	Manawatu at Opiki	manawatu_opik	2720025	6082268	HM	Mana_11a	no	18-Dec-08	Manawatu at Teachers College
37	Tokomaru at Horseshoe Bend	tokomaru_hb	2724295	6076368	LM	Mana_13c	no	18-Dec-08	Tokomaru at RiverlandFarm
38	Rangitikei at Pukeokahu	rangitikei_puk	2771500	6170599	UHS	Rang_2a	no	16-Dec-08	Rangitikei at Pukeokahu
39	Moawhango at Waiouru	moawhango_waiouru	2749046	6193020	UVM	Rang_2d	no	23-Sep-10	Moawhango at Waiouru
40	Rangitikei at Mangaweka	rangitikei_man	2750500	6151099	HM	Rang_3a	no	16-Dec-08	Rangitikei at Mangaweka
41	Porewa u/s Hunterville STP	porewa_us_hun	2729637	6136845	HSS	Rang_4c	no	3-Oct-12	NO FLOW SITE
42	Porewa d/s Hunterville STP	porewa_ds_hun	2729508	6136457	HSS	Rang_4c	yes	3-Oct-12	NO FLOW SITE
43	Rangitikei at Onepuhi	rangitikei_one	2721393	6122388	HM	Rang_3a	no	19-Dec-08	Rangitikei at Onepuhi
44	Rangitikei at McKelvies	rangitikei_mk	2705863	6099094	HM	Rang_4a	no	19-Dec-08	Rangitikei at McKelvies
45	Mangawhero at DoC	mangawhero_doc	2718100	6197500	UVA	Whau_3d	no	17-Dec-08	Mangawhero at Pakihi Rd Br
46	Makotuku at SH49	makotuku_sh49	2710500	6200899	UVA	Whau_3b	no	17-Dec-08	Makotuku at SH 49A Br
47	Mangawhero u/s Ohakune STP	mangawhero_us_oha	2715636	6196590	UVA	Whau_3d	no	17-Dec-08	Mangawhero at Pakihi Rd Br
48	Mangawhero d/s Ohakune STP	mangawhero_ds_oha	2715200	6196694	UVA	Whau_3d	yes	17-Dec-08	Mangawhero at Pakihi Rd Br
49	Makotuku at Raetihi	makotuku_rae	2706701	6195500	UVA	Whau_3c	no	17-Dec-08	Makotuku at Raetihi
50	Mangawhero at Pakihi Road Bridge	mangawhero_pakihi	2710100	6194301	UVA	Whau_3d	no	17-Dec-08	Mangawhero at Pakihi Rd Br
51	Mangatepopo d/s Genesis Intake	mangatepopo_gi	2731007	6236021	UVA	Whai_1	no	24-Sep-10	Mangatepopo Intake at Spillweir

HRCn	Site name	Abbreviation*	E	N	LSC	Sub-zone	PSD	start	Flow site
52	Whanganui d/s Genesis Intake	whanganui_ds_gen	2735298	6238634	UVA	Whai_1	no	24-Sep-10	Whanganui R. at D/S Intake
53	Whakapapa d/s Genesis Intake	whakapapa_ds_gen	2723315	6228846	UVA	Whai_2b	no	24-Sep-10	Whakapapa at Footbridge
54	Waitangi u/s Waiouru STP	waitangi_us_wai	2738867	6190310	UVM	Whau_1b	no	16-Dec-08	NO FLOW SITE
55	Waitangi d/s Waiouru STP	waitangi_ds_wai	2738879	6190109	UVM	Whau_1b	yes	16-Dec-08	NO FLOW SITE
56	Tokiahuru at Karioi	tokiahuru_kar	2725435	6188945	UVA	Whau_1c	no	17-Dec-08	NO FLOW SITE
57	Makotuku u/s Raetihi STP	makotuku_us_rae	2706764	6193797	UVA	Whau_3c	no	19-Jul-10	Makotuku at Raetihi
58	Makotuku d/s Raetihi STP	makotuku_ds_rae	2707001	6193299	UVA	Whau_3c	yes	17-Dec-08	Makotuku at Raetihi
59	Waikawa at North Manakau Road	waikawa_nmr	2698900	6052801	HM	West_9a	no	18-Dec-08	Waikawa at Nth Manakau Rd
60	Ohau at Gladstone Reserve	ohau_gladstone	2707799	6057500	UHS	Ohau_1a	no	18-Dec-08	Ohau at Rongomatane
61	Ohau at SH1	ohau_sh1	2699599	6056900	HM	Ohau_1b	no	18-Dec-08	Ohau at Rongomatane
62	Ohau at Haines Farm	ohau_haines	2695804	6057886	HM	Ohau_1b	no	17-Dec-12	Ohau at Rongomatane

Periphyton data at 13 of the sites with linked flow records (see Section 2.2) had fortnightly data, either starting in August 2013 and running until August 2015 (nine sites, on the Makakahi (2), Mangatainoka (6), Manawatu (1)) or starting in August 2015 and ongoing (four sites, on the Ohau (3) and Waikawa (1)). Therefore, the data available for carrying out Objective 2 comprised two years of data at nine sites, and less than two years of data (August 2015 to April 2017) at four sites.

The dataset also included fortnightly data starting in August 2015 at four further sites that did not have a linked flow record (Makaretu Above Ohau Confluence, Makorokio at Tirohanga Station, Ohau at u/s Makakahi Confluence, Waikawa at u/s Manakau Confluence).

2.1.2 Data preparation

The primary independent variable in most of the analyses was chlorophyll *a*. In the following analyses chlorophyll *a* is used to represent periphyton standing crop at the time of sampling collection (i.e., periphyton abundance net of growth and loss processes, see review in Kilroy et al. 2016).

Chlorophyll *a* data were log(n+1)-transformed for the analyses to meet requirements for normally distributed data and errors in parametric analyses. Depending on the analysis, we used individual values (within-site analyses), maximum annual values (for between-site analyses in different years), or the 92nd percentile (calculated from all of the available data across multiple years). Years were defined from July to June.

Periphyton cover data comprised visual assessments of the percentage cover of the stream bed by periphyton in six categories: no algae (bare rock), thin film, sludge (loose, unconsolidated mat algae), mats (more consolidated cover from 3 mm thick), green slimy filaments, coarse filamentous algae (various colours). Sludge and mats were combined into a single category Mats, and the two types of filamentous algae into Fils. These two combined categories were used to calculate weighted composite cover (WCC), which is %Fils + (%Mats /2).

2.2 Hydrological data

Fifty-one of the 62 sites with periphyton and water quality data had a linked flow record. In some cases, the same flow record was linked to more than one site on the same river (see Table 2-1). Complete flow records from 31 flow recording sites were provided by Horizons Regional Council. The flow records were from January 2000 (or the earliest date of the record, if later) to February 2017.

2.2.1 Data preparation

Daily mean flows were calculated from the complete flow records. At each periphyton site with a linked flow record, we extracted a suite of flow metrics for every monitoring date. The metrics included days since floods of N_m x median flow, where N_m = 1.5, 2, 3, 4 ….. 14, 15, mean and median flows on the day of the survey and in the preceding 7, 15, 30 and 45 days, and percentages of time flows were increasing or decreasing (npos and nneg).

For the between-site analyses, the following flow metrics were calculated for each period considered: mean flow, median flow, number of events exceeding N_m x median flow (where N_m = 1.5, 2, 3, 4 …. 14, 15) (i.e., FRE1.5, FRE2, etc.), mean duration of periods between events of each magnitude (days of accrual, or Da_$N_m$med), total duration in days when flows exceeded each magnitude.

Note that FRE1.5, FRE2, etc. can be calculated in various ways (Booker 2013). For the test of the Biggs (2000a) equations (Section 5) we took FRE3 to be the number of times the threshold was exceeded in the year of interest, regardless of the time between each event (as calculated by Biggs 2000a). For the between-site analyses, we assumed that events less than or equal to 5 days apart were the same event in terms of their effect on periphyton, so that frequencies were smaller for these analyses.

2.3 Water quality data

Water quality data up to December 2016 were supplied for all 50 periphyton sites with linked flow records (i.e., the 51 sites, excluding the short record at Tiraumea d/s Mangatainoka confluence), and for 15 of the sites with no flow record. No water quality data were provided for the four sites with fortnightly periphyton data starting in August 2015, and no flow record (see Section 2.1.1 above). These sites were not included in the analyses.

All the water quality data were linked to the periphyton data by sites and date. In about 90% of cases, the periphyton and water quality data were collected on the same day. Cases where there were discrepancies (usually just one day) were noted in the master file.

The main water quality variables used were dissolved inorganic nitrogen (DIN, the sum of nitrate-N and ammoniacal-N), dissolved reactive phosphorus (DRP) and water conductivity (Cond). Other available variables were total N and P (TN, TP), total dissolved P (TDP), total suspended solids (TSS), and water temperature.

2.3.1 Data preparation

For the within-site analyses nutrient data were log-transformed to achieve normal distributions of data and errors. For the between site analyses the we used mean data at each site calculated over the period of interest (annual, three-year or seven-year periods). Geometric means were used, to reduce the influence of occasional very high outlier values.

Conductivity data were square-root transformed and water temperature data were used untransformed. Arithmetic means were calculated for the between-site analyses.

2.4 Catchment data

The primary objective of these analyses was to look for relationships between periphyton and instream variables. However, in previous studies catchment landcover and/or geology have proved to be good predictors of periphyton abundance (as chlorophyll *a*) (Biggs and Gerbeaux 1993, Biggs 1995), probably because landcover and/or geology carry information about a suite of variables that can directly influence periphyton growth and community composition, which in turn may influence chlorophyll *a*. In this study we included landcover variables derived from New Zealand's land cover database (LCDB version 3), calculated for the entire catchment upstream of each site. Selected land cover variables were simplified to create three uncorrelated variables representing the percentage of the catchment under intensive farmland/pasture, low-producing grassland, and indigenous forest.

We also used classifications for catchment geology and life-supporting capacity (LSC), supplied by Horizons. Refer to Section 10 for details.

Table 2-2: Periphyton, flow, water quality, physical habitat, and catchment variables used in the analyses.

Variable code	Units	Description	Comments
Chlorophyll *a*	mg/m^2	Spectrophotometrically determined in the laboratory using hot ethanol as extractant, from quantitative field samples	Composite from 10 rocks Log(N+1)-transformed when single values used
Mats	%	Periphyton mats (sum of mats and sludge, from visual estimates)	Fourth-root transformation
Fils	%	Periphyton filaments (sum of slimy green and coarse filaments, from visual estimates)	Fourth-root transformation
Film	%	Periphyton film (thin algal covering)	Untransformed
Da_$N_m$med	days	Mean interval (days of accrual) between floods exceeding *n* x median flow, where N_m = 1.5, 2, 3, 4, 5, …. 14,15.	Flow data available at 53 of 61 sites
FREn	index	Annual frequency of flood events exceeding n x median flow, where n = 2, 3, 5, 7 and 10	All flow metrics calculated from records of daily mean flows
mean flow	m^3/s	Mean flow for the period of interest	
nneg, npos	%	Percentage of time the flow was declining or increasing	Used as initial potential variable
DIN	mg/m^3	Mean dissolved inorganic nitrogen (NO_3 + NO_2 + NH_4-N)	Log-transformed
DRP	mg/m^3	Mean dissolved reactive phosphorus (surrogate for inorganic phosphate, PO_4^{3-})	Log-transformed
TDP	mg/m^3	Total dissolved phosphorus (from filtered sample, but includes organically bound component, using a digestion step). Note that TDN data were not available.	Not available at all sites in all years. Data complete from 2012-13
TN, TP	mg/m^3	Total nitrogen and total phosphorus (from unfiltered water samples; includes all organic N and P, including suspended algae, and material adsorbed onto suspended sediment)	Not available at all sites in all years
TSS	mg/L	Total suspended solids in water column. Reflects fine sediment within bed (low flows), and in runoff (high flows)	Not available at all sites in all years
Conductivity	µS/cm	Electrical conductivity of water	Field measurement
Temperature	°C	Water temperature (with conductivity measurement)	Spot field measurement
%coarse	%	Mean percentage of streambed covered by bedrock, boulders and large cobbles combined	From visual assessments

Variable code	Units	Description	Comments
%fine	%	Mean percentage of streambed covered by silt, sand and fine gravel combined	
%sand	%	Mean percentage of streambed covered by sand	
%farm	%	Percentage of upstream catchment in high-producing grassland, horticulture, orchards/vinyards.	From LCDB3 database
%lo_grass	%	Percentage of upstream catchment under low-productivity grassland	From LCDB3 database
%indig_forest	%	Percentage of upstream catchment under indigenous forest	From LCDB3 database

3 Sensitivity of periphyton to flows at each site

Key messages

- The aim of the analysis was to quantify the effect of flows on periphyton at each of 50 sites in the Horizons dataset with a linked flow record and enough data, to identify differences between sites.
- Sites were classified into flow-sensitive and flow-insensitive sites based on the periphyton–flow relationships. Flow-sensitive sites were those at which more than 20% of variance in periphyton chlorophyll *a* was explained by accrual period for a distinct flow threshold. At flow-insensitive sites either accrual period explained less than 20% of the variance in chlorophyll *a*, or no distinct threshold was identiifiable.
- At 31 sites, we were able to define a distinct threshold for the size of flow event that would typically remove periphyton to a low level. At three more sites, removal typically occurred over a wide range of flows.
- The thresholds varied between 1.5 and 15 x median flow.
- We compiled a new variable from this analysis: the ***effective flow***, and accrual period calculated from the effective flow (i.e., the time periphyton has to grow without being washed away in a high flow). This is a new idea: up until now a rule-of-thumb has been that events >3 x median flow generally remove periphyton in rivers.
- Accrual period calculated using the effective flow explained up to 53% of the variance in chlorophyll *a* within a site (and over 40% at 14 of the 42 sites), supporting previous research conclusions, that flow variability is commonly the dominant driver of variability in periphyton.
- Effective flow thesholds were exceeded for between 4% and 33% of the time at flow-sensitive sites with low flow thesholds (up to 5 x median flow) and for lower percentages of time (0.6 to 4.2%) at sites with thresholds less than 5 x median flow.
- Previous work using the Horizons dataset has identified that hydraulic and geomorphological characteristics determine the effective flow. Simple field techniques to determine the effective flow at new sites are under development.

3.1 Introduction

River flows and the effects of floods generally have the strongest influence on periphyton standing crop at any given site over time (Biggs and Close 1989, Biggs 1995). Flows influence periphyton standing crop though the effects of scouring leading to biomass removal (Francouer and Biggs 2006) and the effects of water velocity on community composition and growth rates (Hart et al. 2013). The effect of high water velocities in removing periphyton by shear stresses on the algae is modified by the geomorphological characteristics of a river site because substrate particles mobilised by high flows cause additional abrasion and scouring (Francouer and Biggs 2006, Hoyle et al. 2017). While river flow regimes are generally variable over time, geomorphology reflects the geological setting and broad-scale climate of the catchment and tends to be characteristic of a site. Consistent geomorphological characteristics over time combined with a less predictable flow regime can be expected to lead to somewhat predictable effects of high flows at a given site. The objective of the analysis described in this section was to identify at each site whether there was a characteristic flow magnitude (threshold) at each site that typically removed periphyton.

3.2 Methods

The approach used to identify flow thresholds for effective removal of periphyton (as chlorophyll *a*) was to extract from the flow record the number of days since flows of specified magnitudes (in multiples of median flow, N_m) and then run regressions between chlorophyll *a* and days since the event of each magnitude for the time-series of periphyton at each site. The number of days since a high flow event is potentially the accrual time available for periphyton development, assuming that smaller flow perturbations during that time have no or only a minor effect on biomass. Relationships in which accrual time explains a high proportion of the variability in chlorophyll *a* would indicate that the flow threshold defining the accrual time approximates the threshold that removes periphyton to low levels. This method isolates the effective flow because if N_m is too low, high chlorophyll *a* could occur after short accrual times because some high flows would fail to remove biomass, leading to low explanatory power; if the selected flow size is too high, then low chlorophyll *a* could occur after long accrual periods after being removed by smaller flows, again leading to low explanatory power. Only at flow sizes close to the threshold for removal would we expect a strong correlation between chlorophyll *a* and days since the high flow, with the slope of the relationship approximating the rate of accrual.

A caveat to this method is that care needs to be taken in interpreting relationships when accrual times are very long, because spontaneous sloughing can lead to unexpectedly low biomass (Biggs and Close 1989). It is also acknowledged that the condition of the periphyton can influence the effect of a particular high-flow event (Katz et al. 2018).

At each site with a flow record and sufficient data (49 sites), we extracted the time in days since a high flow greater than N_m x median flow (where N_m = 1.5, 2, 3, 4, 5, …. in steps of 1, up to 15). Median flow was calculated using long-term flow data at each site (2000 to 2016), or all the available data if the record was shorter. Linear regressions were run on log-transformed chlorophyll *a* versus log-transformed time in days since an event of each magnitude. The linear regression results (particularly the adjusted R^2, hereafter R^2) and plots of the relationships at each site were reviewed.

To identify whether a particular flow threshold, or a range of thresholds, removed periphyton, we also plotted the R^2 of each relationship against multiples of median flow. A periphyton-removal flow threshold was indicated by those relationships showing a definite maximum value of R^2. If equivalent maximum values extended over a range of *N*, the lowest value of *N* at which the relationship was strongest (and also significant) was taken to indicate the removal threshold.

3.2.1 Results

Summary results (R^2) of all the regressions run at each site are presented in Appendix C. The "effective flow" was determined for each site from the highest R^2 of the relationships between chlorophyll *a* and time in days since a high flow greater than N_m x median flow. Effective flow was highly variable and ranged from 1.5 to 15 x median flow (Table 3-1). Plots of R^2 against N_m shows how the relationship between flow magnitude and periphyton varied across sites. At some sites R^2 peaked at a clear value of N_m. For example, at manawatu_opik, maximum R^2 occurred at N_m = 2, and at mangatainoka_lars at N_m = 14. At other sites, a maximum R^2 was identified in Table 3-1, but the value did not vary substantially across the whole range of N_m.

Varying shapes of the relationships between flow magnitude and the R^2 of the chlorophyll *a* – days since flow threshold relationships enabled a classification into groups of sites based on the overall response to flows. Plots for all sites are shown in Figure 3-1, along with their groups (A to D, see below).

Table 3-1: Flow thresholds most likely to remove periphyton to a low level (effective flows) at each site. Site in order of their HRC site number (HRCn). Effective flow was defined by running linear regressions of $\log_{10}$chla versus $\log_{10}$ accrual time, where accrual time was the number of days since a high flow of N_m x median flow, where N_m = 1.5, 2, 3, 4, … 15. Effective flow (in multiples of median flow) was taken as N_m in the relationship with maximum R^2. Group = Removal groups A to D, assigned after inspection of the plots in Figure 3-1. Refer to definitions in the text. ? means that there was some uncertainty about the group assignment. NA no group because of insufficient data range. Flood frequencies (with no and a 5-day window – see text), and the percentage of time flows exceeded the effective flow were calculated from the record from 2009 to 2016.

HRCn	Periphyton site	Derivation of effective flow				Flood freq. (FREeff)		% time effective flow exceeded
		Max R^2	P	Effective flow	Group	No window	5-day window	
1	makakahi_doc	0.14	0.003	11	C	5.0	4.0	1.8
2	mangatainoka_putara	0.14	<0.001	10	C	11.6	8.9	4.2
3	mangatainoka_lars	0.53	<0.001	14	B	7.6	6.2	2.0
7	mangatainoka_huk	0.31	<0.001	10	B	11.6	8.9	4.2
8	kumeti_tr	0.46	<0.001	1.5	A	9.8	6.6	28.5
9	manawatu_weber	0.39	<0.001	5	D	9.3	6.6	7.0
10	makakahi_ham	0.13	<0.001	13	C	4.1	3.3	1.3
11	oroua_apiti	0.40	<0.001	3	A	13.5	9.2	9.8
12	tamaki_ste	0.46	<0.001	3	A	12.4	8.3	11.1
13	oruakeretaki_sh2	0.49	<0.001	2	A	16.3	9.9	21.2
14	makuri_tuscan	0.31	<0.001	3	A	12.8	9.3	6.7
16	mangatainoka_scarb	0.42	<0.001	15	B	5.2	4.4	1.6
17	tiraumea_nga	0.52	<0.001	4	A	14.3	9.4	12.0
18	mangatainoka_pahiatua	0.28	<0.001	4	D	16.8	10.1	10.3
19	mangatainoka_sh2	0.43	<0.001	10	B	5.1	4.3	2.0
20	mangatainoka_ds_db	0.32	<0.001	11	B	4.3	3.6	1.6

HRCn	Periphyton site	Derivation of effective flow				Flood freq. (FREeff)		% time effective flow exceeded
		Max R^2	P	Effective flow	Group	No window	5-day window	
21	mangatainoka_us_pah	0.45	<0.001	10	B	5.1	4.3	2.0
22	mangatainoka_ds_pah	0.37	<0.001	11	B	4.3	3.6	1.6
23	manawatu_hop	0.47	<0.001	1.5	A	15.2	8.2	33.3
24	mangatainoka_us_tir	0.35	<0.001	11	B	4.3	3.6	1.6
26	mangapapa_troup	0.29	<0.001	10	D	5.7	4.6	2.4
27	pohangina_mais	0.29	<0.001	4	A	12.2	8.8	6.9
28	manawatu_ug	0.42	<0.001	3	A	17.8	11.1	13.2
29	oroua_almadale	0.21	<0.001	3	A	13.5	9.2	9.8
33	manawatu_tc	0.45	<0.001	2	A	21.2	10.6	22.5
34	manawatu_us_pncc	0.51	<0.001	3	A	16.2	10.6	12.3
35	manawatu_ds_pncc	0.43	<0.001	4	A	12.2	8.4	7.3
36	manawatu_opik	0.51	<0.001	2	A	21.2	10.6	22.5
37	tokomaru_hb	0.02	0.215	15	C			
38	rangitikei_puk	0.35	<0.001	4	A	8.5	6.6	4.8
39	moawhango_waiouru	0.21	<0.001	5	NA	0.5	0.5	0.3
40	rangitikei_man	0.30	<0.001	4	A	6.8	5.0	4.0
43	rangitikei_one	0.34	<0.001	4	A	8.4	6.1	4.9
44	rangitikei_mk	0.39	<0.001	4	A	9.2	6.4	6.6
45	mangawhero_doc	0.06	0.015	8	C			
46	makotuku_sh49	0.13	<0.001	8	C	9.8	7.1	4.1
47	mangawhero_us_oha	0.08	0.004	11	C			
48	mangawhero_ds_oha	0.13	<0.001	8	C			
49	makotuku_rae	0.09	0.007	13	C			
50	mangawhero_pakihi	0.31	<0.001	8	B?	1.9	1.7	0.7
51	mangatepopo_gi	0.35	<0.001	15	B	1.8	1.7	0.6
52	whanganui_ds_gen	0.02	0.295	2	NA			
53	whakapapa_ds_gen	0.32	<0.001	3	NA	4.2	3.7	1.9
57	makotuku_us_rae	0.04	0.105	12	C			
58	makotuku_ds_rae	0.15	<0.001	12	C			
59	waikawa_nmr	0.23	<0.001	2	A?	24.3	13.6	21.5
60	ohau_gladstone	0.22	<0.001	6	C	9.1	6.8	3.7
61	ohau_sh1	0.35	<0.001	3	A	21.1	12.4	12.4
62	ohau_haines	0.32	<0.001	5	A	11.8	8.8	5.1

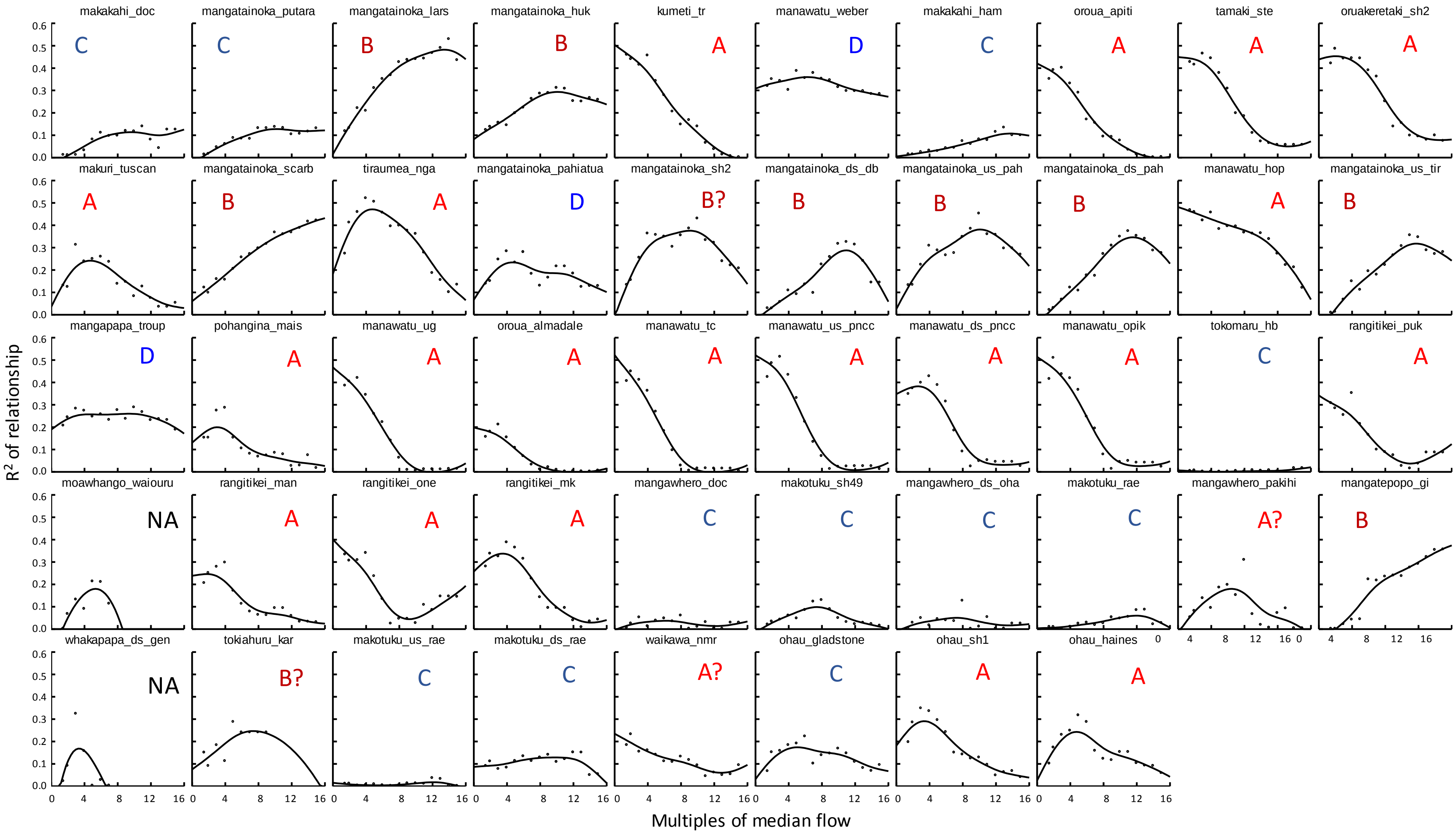


Figure 3-1: Plots of the R^2 of the relationship between chlorophyll a and days since a high flow vs the flow magnitude in multiples of median flow. Median flow was taken as the long-term median at each site (2000 to 2016). The flow threshold above which periphyton was effectively removed was the flow at which R^2 was maximum provided the relationship was statistically significant. The plots help to classify sites into: (A) unstable (low removal thresholds); (B) stable (high removal thresholds); (C) unresponsive (no clear threshold, and R^2 values generally low); and (D) undefined (no clear threshold, but R^2 values generally significant), by looking at the shape of the relationship as well as the maximum R^2 value. The shapes of the relationships are shown by fitting distance-weighted least squares smoothing lines through the data points. Sites arranged in order of the Horizons site number.

The site classification comprises four groups:

A. sites at which there was a maximum R^2 at a relatively small flood magnitude ($N_m < 5$). Group A included sites in the Kumeti, most sites on the Manawatu mainstem, and sites on the Ohau, Oroua, Rangitikei, Tamaki and Tiraumea Rivers;

B. sites at which there was a maximum R^2 at a relatively large flood magnitude ($N_m \geq 5$). Most group B sites were on the Mangatainoka River;

C. sites with periphyton that appeared to be generally unresponsive to high flows (at least up to the threshold tested), and there were no strong relationships between chlorophyll *a* and time since a high flow at any of the flow thresholds tested. Group C sites included sites on the Makotuku, Mangawhero, Tokomara and Waikawa;

D. a small number of sites on various rivers, which had significant relationships across the whole range of flow thresholds tested but no obvious optimum R^2.

Two sites could not be fit into the classification because the range of flow thresholds was limited (moawhango_waiouru, whakapapa_ds_gen).

The above classification can be simplified further to two categories:

- **Flow-sensitive sites**: sites at which a periphyton removal threshold can be identified, regardless of the magnitude of the threshold. This group includes groups A and B above (and in Figure 3-1);
- **Flow-insensitive sites**: sites at which no removal threshold was clear from the data. This group includes groups C and D above.

Table 3-1 also shows, for each site, the mean annual number of flow events exceeding the effective flow (FREeff). Two versions are shown. FREeff with no window is the number of times the flow exceeded the threshold regardless of how close together the events were. FREeff with a 5-day window counts successive events with 5 or fewer days between them as a single event. The rationale for 5-day window is that events occurring close together are effectively a single event from the perspective of periphyton, because no accrual is likely in a brief intervening period of 5 days or less. These definitions are referred to again in Sections 5 and 6.

The percentage of time the effective flow was exceeded is a different metric and indicates the total duration of high flows. Both FREeff and the duration of effective flow varied across sites. See Section 3.3.2 below for further comment.

3.3 Discussion

3.3.1 Methodology

Hoyle et al. (2017) identified flow thresholds for effective removal of periphyton at 18 of the sites in the Horizons dataset using an alternative empirical approach to that described above. The approach (simplified) was, for each site:

- extract from the flow record the highest flow in a designated period preceding each periphyton (chlorophyll *a*) observation;

- plot chlorophyll *a* versus maximum flow in the designated period (e.g., maximum flow in the preceding 7 days);
- identify from the plot the flow at which chlorophyll *a* was always lower than a specified low threshold (e.g., 10 mg/m^2).

This method requires selection of an appropriate period during which periphyton can accrue (typically up to about three weeks) and selection of a suitably low value of chlorophyll *a* (e.g., up to about 10 mg/m^2, which normally indicates low or no cover by mats and/or filaments).

Using a period of one week (i.e., plotting chlorophyll *a* against the maximum flow in the previous 7 days), the thresholds calculated by Hoyle et al. (2017) were similar to those identified in the present analysis at 13 of the 18 sites (within 2.5 multiples of median flow). The largest discrepancies were at makuri_tuscan (3 x median in this analysis vs. 11.7 x median in Hoyle et al. 2017) and waikawa_nmr (2 x median vs. 7.3 x median). The discrepancies could have arisen because: (a) a longer dataset was used to make the assessment in the present analysis (2009 to 2017 versus 2009 to 2013); (b) Hoyle et al. (2017) used daily maximum flows to identify flow thresholds, rather than daily mean flows, which may have identified short-lived higher flow peaks that removed biomass, that could not be identified from the record of daily mean flows; (c) the 7-day window for peak flows may not have been appropriate at all sites. For example, flow peaks that did not remove periphyton but occurred just outside the 7-day window would not have been identified. We note that the estimates of effective flow are not exact and probably cover a range of multiples of median flow rather than a single value. For example, the highest R^2 at kumeti_tr was for the relationship between chlorophyll *a* and days since a flow 1.5 x median. However, the R^2 for the 4 x median flow relationship was only marginally lower than that for 1.5 x median (see Appendix C).

As mentioned in Section 3.2, some unexplained variation is expected in the method used here (i.e., comparing relationships between chlorophyll *a* and accrual times based on different flow thresholds). Thus, the highest R^2 of all the relationships was 0.53 (at mangatainoka_lars). Even using a fine-scale two-dimensional hydraulic modelling approach in a recent study, it was still only possible to explain 49% of the variance in chlorophyll *a* over a series of high flow events (Katz et al. 2018). A potential source of variability is spontaneous sloughing of periphyton, which occurs during long periods of stable flows after periphyton develops to high levels and undergoes natural degradation. Very long accrual periods (e.g., > 90 days, Biggs and Stockseth 1995) may lead to unexpectedly low chlorophyll *a* levels following a change in periphyton structure following natural sloughing, which prevents recolonization by algae (Biggs and Close 1989). The effect of spontaneous sloughing may not be critical at many sites. However, in selecting the appropriate effective flow from the regression results, it is important to also view the plots of the relationships.

From the above, non-significant relationships between accrual time and chlorophyll *a* at two sites with regulated flows may therefore reflect unpredictable periphyton biomass responses to individual flow events, and to very long periods of accrual. Very long flood-free periods may also account for the lack of a definable effective flow in both regulated (e.g., moawhango_waiouru, whanganui_ds_gen) and unregulated rivers (e.g., makakahi_ham, makotuku_us_rae).

Some of the flow-sensitive sites with both small and large effective flows (respectively <5 x median and >5 x median; in groups A or B in Table 3-1) also had low flood frequencies but, by definition, showed relatively strong relationships between chlorophyll *a* and accrual period. In that case, site-specific factors may distinguish sites with variable periphyton during long accrual periods and sites with predictable periphyton accrual.

One advantage of the Hoyle et al. (2017) method is that there is no upper limit for the flow magnitude identified. We considered only flow up to 15 x median. Including larger flows could help to clarify the threshold at some of the sites identified as flow-insensitive (comprising sites classified as C or D in Table 3-1 and Figure 3-1). In both methods, the longer the data series, the better the chances of obtaining an accurate picture of removal patterns at a site.

Following Hoyle et al. (2017), it would be informative to repeat the present analysis using time series of maximum daily flows rather than mean daily flows. The two time-series are highly correlated. However, some subtleties of the effects of short-lived flow events on periphyton may mean that daily maximum flows provide a more accurate representation of the effective flow. A worthwhile exercise would be an evaluation of different ways of estimating the effective flow empirically, with the aim of defining a the most robust and defensible methodology.

3.3.2 Site classification

The outcome of this analysis was a suggested classification of sites into four groups based on the magnitude (in terms of multiples of median flow) of the effective flow and the strength of the chlorophyll *a* versus days of accrual relationship. The four groups can be reduced to two groups representing flow-sensitive and flow-insensitive sites.

The analysis has also provided an additional variable for use in subsequent analyses: <u>days of accrual based on the effective flow</u>, hereafter referred to as **Da_EFF**. Because the effective flow could not be clearly identified at some sites (at least using the present method), the number of sites available for analysis using this variable is restricted to a maximum of 42 (in the seven-year dataset), including some sites where the effective flow spanned a range of flows (i.e., groups C and D, see Table 3-1).

A challenge is to determine whether measurable site features could allow the effective flow to be estimated without the need for analysis of a long time-series of data, or conducting additional field measurements to determine the effective flow from hydraulic principals (as in Hoyle et al. 2017). The most likely site characteristics that would affect periphyton biomass (i.e., chlorophyll *a*) removal are bed substrate composition, site-scale hydraulic conditions (e.g., water velocity and turbulence, reach slope), flood frequency, and possibly the type of periphyton that typically grows at a site.

In a preliminary investigation, we used box plots to compare values of individual variables such as bed substrate composition and flood frequency between groups A to D (as defined in Figure 3-1, Table 3-1). No single variable stood out as distinguishing between the classes (data not shown). By combining selected variables in a multivariate ordination, we were able to distinguish most sites in group A (sites with a removal threshold of up to 5 x median flow) from most sites in groups B (sites with higher thresholds) and C, and from the few sites at which no threshold could be identified. However, groups B, C and D were not distinguishable (Figure 3-2).

Hoyle et al. (2017) provided a physical explanation for differences in removal thresholds among sites. Periphyton removal thresholds tended to approximate the threshold that mobilised the finer fractions of bed sediment at a site, and removal depended on the amount of fine sediment available and the frequency at which it was mobilised. The finding was consistent with abrasion / scouring by fine sediment as an important removal mechanism (Francouer and Biggs 2006). The empirically determined effective flow likely represents these physical processes. Since no measured site variables allow clear identification of the size of an effective flow, the simplest method may need to include one-off field measurements to determine the flow magnitude that moves fine sediment. Such a method is under development.

Further classification of sites could be based on the frequencies and durations of effective flows. The frequency of flows capable of removing periphyton is as important as the size of the flow in determining whether periphyton can develop to nuisance levels at a site. Frequencies of effective flows (with no allowance for events that occurred close together) varied from less than one per year (at moawhango_waiouru) to more than 20 per year (manawatu_opik, waikawa_nmr). When events less than 5 days apart were counted as a single event, frequencies at the higher end of the range were reduced to a maximum of about 14 events at waikawa_nmr (Table 3-1).

The percentage of time that the effective flows was exceeded was generally negatively correlated the magnitide of the effective flow and positively correlated with the event frequency. In the latter correlation two outliers were kumeti_tr and manawatu_hop, both of which were estimated to have effective flows of 1.5 x median flow. Consequently, the percentage of time when high flows was exceeded was high (up to 33%) relative to the number of events, because a river can run at slightly elevated flows for long periods. The possibility was considered that the effective flow estimates of 1.5 x median were incorrect. However, Hoyle et al. (2017) calculated similar effective flows (based on maximum daily flows). Therefore, at this stage, an effective flow of 1.5 x median is retained for these two sites.

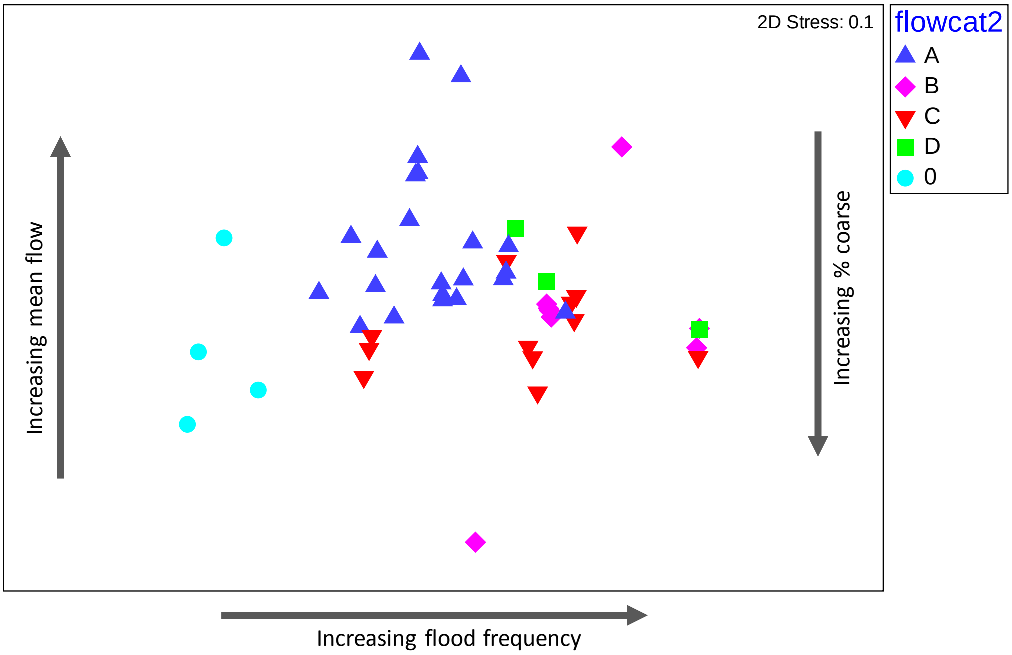


Figure 3-2: Non-metric multi-dimensional scaling ordination of sites using substrate and hydrological variables. Data from the three years July 2012 to June 2015. Sites are colour-coded to distinguish the removal groups defined in Table 3-1. Group 0 includes sites at which no flow threshold was identified. Variables included were %coarse, %fine gravel + sand, FRE3, FRE5, FRE7, FRE10 and mean flow, plus the flow metrics coefficient of variation (CVflow) and the proportion of time flow was declining (nneg). Groups A and 0 differed from groups B, C and D (ANOSIM $P < 0.05$), but groups B, C, and D could not be separated from each other. The arrows show the directions of general gradients of flood frequency, mean flow and % coarse.

3.3.3 Standardisation of flows to the median flow

In this analysis we reduced all the flow records to comparable status by converting the daily mean flows to multiples of median flow. The assumption that scouring thresholds scale with median flow is largely based on observations that the effect of any flow event on periphyton is determined by the event's size relative to preceding flows (Biggs and Close 1989, Biggs et al. 2005). It was also implied in the selection of FRE3 (the annual frequency of floods exceeding 3 x median flow) as the most ecologically useful flow variable indicator in New Zealand streams. In that analysis, periphyton biomass declined as FRE3 increased in a set of rivers spanning three orders of magnitude in mean flow (0.5 to 500 m^3/s) (Clausen and Biggs 1997). More recent comparisons of periphyton across multiple rivers have also relied on standardised flow metrics across a wide range of absolute mean flows (e.g., from 0.15 to 100 m^3/s; Schneider and Petrin 2017). It is possible that this approach to standardising flows across rivers may not be appropriate in all cases. In particular, rivers with regulated flows may have minimum flows that in effect become the median flow. The effect of multiples of the minimum flow in such regulated rivers may not be comparable with the effect of multiples of the median flow in rivers with natural flow regimes.

4 Patterns of periphyton nutrient limitation

Key messages

- The aim of the analysis was to use the dissolved inorganic N (DIN) and dissolved reactive P (DRP) data to assess potential nutrient limitation of periphyton growth at each site in the dataset. N or P limitation occurs if one or other of these nutrients is in short supply. In that case additions of that nutrient could cause periphyton to increase.
- We assessed limitation by looking at DIN : DRP ratios and DIN and DRP concentrations. DIN and DRP can be correlated with river flow, usually positively. Therefore, we also took this effect of river flow into account.
- Based only on DIN : DRP ratios and at all flows and seasons (including sites with no linked flow record), 50% of the sites were P-limited (mostly in the Manawatu, Makotuku, Mangawhero and Ohau Rivers); 33% were N-limited (most headwater sites, and all sites in the Rangitikei River), and 17% were limited by both N and P (co-limited).
- At over 60% of the 47 sites with a flow record and enough data, highest DIN typically occurred in high flows, when periphyton is being sloughed.
- When samples collected during high flows were excluded, three sites with flow records shifted from P-limitation to co-limitation, and five from co-limitation to N-limitation. In this smaller dataset, 51% of sites were P-limited, 40% N-limited, 9% co-limited.
- When concentrations of saturating DIN and DRP (i.e., enough for maximum growth rates) were taken into account as well as flows, 28% of sites (with flow records) were assessed as predominantly P-limited, 8.5% as N-limited, 55% as co-limited, and 8.5% as limited by neither N nor P.
- Regardless of how it was calculated, the limiting nutrient varied over time at all sites. This implies that additions of either N or P could potentially stimulate periphyton growth at different times, over much of the regional monitored stream network.
- Seasonality of nutrient limitation was not accounted for in the analysis, but further analysis of this aspect of variability would likely be useful.

4.1 Introduction

A part of Objective 1b, paragraph 3 in the contract (see Appendix A) was to "use classical DIN/DRP theory to classify stations into N-limited, P-limited and co-limited." This classification is set out separately in this section for ease of referral in subsequent analyses.

In theory, periphyton will respond to changes in available N or P only when one or other of these nutrients is either already in short supply or shifts into short supply status as a result of the change. When N or P are in short supply, periphyton growth is, respectively, N- or P-limited. The classical theory of nutrient limitation of the growth of plants and algae holds that plant growth is N-limited when the ratio of available N to available P is less than 7 to 1 (by weight, equivalent to a molar ratio of 16 to 1), and P-limited when the N to P ratio is greater than 7 to 1 (by weight). The ratio originated from work on marine algal cells, which were found to contain N and P in a more or less consistent molar ratio of 16 to 1 (Redfield 1958).

Although developed for marine communities, the theory is routinely applied to freshwater algae in both rivers and lakes, with the assumption that the ratio applies to N and P in the water column. Available N and P in rivers can be conveniently represented by concentrations of dissolved inorganic N (DIN) and dissolved reactive P (DRP) in the water column, respectively, although the ratio has also been calculated using total phosphorus (Dodds 2003a, Keck and Lepori 2012). Limitation of periphyton growth by N or P in rivers can be demonstrated using *in situ* trials with nutrient diffusing substrates (NDS). Such trials are short-term (up to 2 weeks) and can only reflect conditions over that brief period. Early studies in New Zealand using NDS indicated that water column ratios of DIN to DRP did not reflect the NDS results very well (Francouer et al. 1999). However, more recent studies using NDS have demonstrated that water column DIN : DRP ratios can be good indicators of which nutrient limits growth at the time of the trial (Kilroy and Wech 2015, Haidekker and Wade 2016).

While DIN : DRP ratios can reflect the nutrient limitation status at a site at a given time, defining a site-specific limiting nutrient using a single ratio could be misleading because both DIN and DRP concentrations fluctuate over time, both seasonally and with flows, and limitation status correspondingly changes over time. McArthur et al. (2010) set out an approach that accounts for changes in nutrient concentrations over time and also adopts saturating concentrations of DIN and DRP (i.e., concentrations above which no further growth stimulation can occur because algal cell growth rates are physiologically at their maximum). Therefore, in addition to using classical DIN/DRP theory to classify sites, we determined the nutrient-limitation status of each site following the general approach of McArthur et al. (2010).

4.2 Methods

We first calculated DIN : DRP at each site. With a time-series of data available at each station, there are at least two options for calculating this ratio.

(a) Calculate mean DIN and DRP over the time series and use the mean values to calculate DIN : DRP. We suggest that the geometric mean is the appropriate metric here, so that the overall mean is not strongly influenced by small numbers of extreme values.
(b) Calculate DIN : DRP for each individual survey, and the calculate the mean value across all dates (again using the geometric mean).

Seasonal patterns and flow fluctuations account for much of the variability of DIN (and sometimes DRP) over time. Seasonality of DIN and DRP at each site was obtained from Kilroy et al. (2016). A Kruskal-Wallis test (non-parametric ANOVA) was applied (in Time Trends v. 5) to determine whether there were differences among months in median values. At sites where differences between months were detected, the month(s) with maximum and minimum median chlorophyll *a* were identified from the box plots. At least 48 samples were deemed to be necessary to run the analyses on individual sites (i.e., on average four samples per month). Strongly significant results were those with

P values <0.005 (i.e., applying a correction to allow for false discovery rate of significance in multiple tests). Weaker seasonality was assumed for P > 0.005 but < 0.05.

We also ran regressions of DIN and DRP versus flow on the day of the survey to determine the relationships at each site. If flow and DIN or DRP are correlated, filtering the data to remove all high flows (e.g., greater than median flow) could result in a different nutrient limitation assessment. The DIN and DRP data were log-transformed for these analyses.

We then determined nutrient limitation status (using the classical method, based on DIN : DRP ratios) at all sites using both methods (a) and (b) above, first including data across all flows and second including only data at flows below median. The thresholds for determining nutrient limitation were those suggested by McDowell et al. (2009): N-limitation, DIN/DRP <7; P-limitation, DIN/DRP >15; co-limitation, DIN/DRP 7 – 15.

To assess extent of potential fluctuations in nutrient limitation status over time following the McArthur et al. (2010) approach, we used the time-series of data to calculate the percentages of surveys at each site on which:

- both DIN and DRP exceeded concentrations assumed to be saturating for periphyton growth (i.e., nonlimited growth);
- both DIN and DRP were less than saturating concentrations (i.e., N and P co-limited);
- DIN exceeded the saturating concentration, and DRP did not (P-limited growth);
- DRP exceeded the saturating concentration, and DIN did not (N-limited growth).

Saturating concentrations of DIN and DRP vary depending on which other factors may be growth-limiting at the site of interest. In this case the concentrations were taken to be DIN ~350 mg m^{-3} and DRP 15 mg m^{-3}. The DIN concentration was based on the international literature (Dodds et al. 2002, 2006, as discussed in Kilroy 2017; see also Rier and Stevenson 2006). The saturating concentration for DRP was assumed to be lower than the 28 mg/m^3 suggested by Bothwell (1989) because observations indicate that the lower value of 15 mg/m^3 is more realistic for New Zealand rivers. For both DIN and DRP, experiments have demonstrated that algal cellular growth rates saturate at lower concentrations than those associated with maximum standing crop. For example, DIN limited growth rates at concentrations below 86 mg/m^3 (with replete DRP), but DIN saturation relative to biomass occurred at around 300 mg/m^3 (Rier and Stevenson 2006). Bothwell (1988) demonstrated that cellular growth rates in thin diatom films saturated at extremely low DRP concentrations (< 2 mg/m^3). The differences are attributable to the fact that algae in thick periphyton mats or dense filamentous growth can only access DIN or DRP from the water column when concentrations are high enough for diffusion to occur into the deeper layers of periphyton.

The resulting percentages were used to assign an overall nutrient limitation status, or range of nutrient limitation, to each site. We assigned N-, P- and co-limitation when at least 50% of the samples were under the saturation threshold for DIN, DRP and both DIN and DRP. A fourth category indicated sites where neither DIN nor DRP were limiting because concentrations of both exceeded the assumed saturating concentrations. If no category included 55% or more of the sites, then a range of nutrient limitation was specified. For example, for a site where 54% of the samples were below the P-saturation threshold (and above the N-saturation threshold) and 42% were below both thresholds, the site was designated P-co-limited.

In summary, we assessed nutrient limitation status of each site using four methods:

1. classical DIN : DRP ratios, including data at all flows;
2. classical DIN : DRP ratios, including only data collected when flows < median flow;
3. McArthur et al. (2010) approach, accounting for saturating concentrations, including data at all flows;
4. McArthur et al. (2010) approach, accounting for saturating concentrations, including only data collected when flows < median flow.

4.3 Results

4.3.1 DIN and DRP in relation to season and flow

Thirty-nine of the 58 sites with sufficient nutrient data showed a strong seasonal pattern in DIN; at seven further sites, the seasonal pattern was weak (Table 4-1). The 12 remaining sites did not show a seasonal pattern in DIN (makakahi_doc, makotuku_sh49, makuri_tuscan, mangatainoka_huk, mangatera_ds_dan, mangawhero_doc, moawhango_waiouru, oroua_awahuri, oroua_ds_fei, tiraumea_nga, waitangi_ds_wai, waitangi_us_wai).

At 29 of the 47 sites with flow records and sufficient nutrient data for analysis, DIN was significantly and positively correlated with flow, with highest concentrations of DIN associated with high flows. Periphyton accrual generally occurs during periods of stable flows, when therefore high DIN at high flows is unlikely to have consequences for instream growth (although it may have consequences for downstream receiving waters). DIN was negatively correlated with flow at three sites (makotuku_sh49, tiraumea_nga, whakapapa_ds_gen).

In contrast to the strong seasonal patterns seen in DIN, DRP showed a seasonal pattern at only one site, and that site was downstream of a waste-water treatment plant (WWTP, mangatera_ds_dan). DRP was positively correlated with flow at nine sites, including most sites on the Manawatu River mainstem. DRP had negative relationships with flow at three sites (makotuku_sh49, makotuku_us_rae, mangawhero_doc).

4.3.2 Nutrient limitation based on DIN : DRP

Across all flows, and based purely on observed DIN : DRP concentrations, half of the sites (29 of 58) fell into the P-limited category. These sites were mostly in the wider Manawatu catchment (down to tokomaru_hb in Table 4-1). Other P-limited sites were in the Makotuku, Mangawhero and Ohau rivers. Nineteen sites were theoretically N-limited. These included headwater sites that had low DIN concentrations, as well as all sites in the Rangitikei catchment. The remaining 10 sites were co-limited, according the DIN : DRP ratio, although their mean concentrations varied widely.

Filtering out flows greater than median flow (for the 47 available sites with flow records) shifted limitation status at eight sites. Three sites (manawatu_hop, mangapapa_troup, and mangawhero pakihi) shifted from P-limitation to co-limitation. At all three sites, DIN was significantly correlated with flow, and mean DIN at was at least 25% lower at flows less than median than across all flows. At five sites (mangatainoka_lars, oroua_apiti, tokomaru_hb, mangawhero_us_oha, mangawhero_ds_oha) co-limitation shifted to N-limitation, again reflecting a lower mean DIN at low flows compared to an overall mean DIN that was already relatively low (Table 4-1).

Table 4-1: Nutrient limitation status assigned to each of the periphyton monitoring sites using ratios. The first six columns show variability in DIN and DRP with season, flow and over time. Trends were obtained from Kilroy et al. (2016). Blank cells mean no trend was detected from the data. The next four columns (A) show limitation status based on all of the data. DIN and DRP are average concentrations (geometric means, in mg/m^3) at each site. DIN/DRP was calculated as the mean of DIN : DRP from each individual survey. Status was assessed from DIN : DRP < 7 = N-limited, > 15 = P-limited, and 7 – 15 = co-limited. The final four columns (B) are the same as A. except that data collected at flows > median flow were omitted. Sites are arranged in order of their HMW site number (from upstream to downstream in successive catchments).

	Variability in DIN			Variability in DRP			A. Limitation status, all flows				B. Limitation status, flows < median			
Site abbreviation	Season	Flow	Trend	Season	Flow	Trend	DIN	DRP	DIN/DRP	Status	DIN	DRP	DIN/DRP	Status
makakahi_doc	no	no		no	no		23	6	3.8	N	24	6	3.9	N
mangatainoka_putara	weak	no		no	no		13	5	2.7	N	12	5	2.8	N
mangatainoka_lars	yes	pos		no	no		39	5	7.7	co	29	5	5.9	N
tamaki_res	yes			no			48	9	5.2	N				
mangatera_us_dan	yes			weak			318	47	6.8	N				
mangatera_ds_dan	no			yes			1284	178	7.2	co				
mangatainoka_huk	no	no		no	no		596	6	105.6	P	610	5	117.8	P
kumeti_tr	yes	pos		no	pos	neg	557	10	58.1	P	407	9	49.7	P
manawatu_weber	yes	pos		no	pos		216	17	12.9	co	108	15	8.2	co
makakahi_ham	yes	pos		no	pos		300	6	50.6	P	179	5	33.8	P
oroua_apiti	yes	pos		no	no	neg	49	6	7.6	co	33	7	6.1	N
tamaki_ste	yes	pos		no	no		290	8	35.3	P	181	8	27.7	P
oruakeretaki_sh2	yes	pos		no	no	neg	758	14	55.2	P	637	15	44.8	P
makuri_tuscan	no	no		no	no		821	8	100.6	P	875	8	120.4	P
pohangina_pir	yes			no			33	6	6.1	N				
mangatainoka_scarb	yes	no		no	no		971	6	171.7	P	945	5	187.6	P
tiraumea_nga	no	neg		no	weak		575	9	65.3	P	586	9	67.0	P
mangatainoka_pahiatua	yes	no		no	no		914	6	148.7	P	895	6	151.9	P
mangatainoka_sh2	yes	weak		no	weak		831	7	126.4	P	777	6	131.0	P
mangatainoka_ds_db	weak	pos		no	no		809	8	107.1	P	709	7	111.4	P
mangatainoka_us_pah	yes	no		no	no	neg	843	9	91.9	P	781	10	85.8	P
mangatainoka_ds_pah	weak	no		no	no	neg	893	11	78.0	P	858	12	75.2	P

Site abbreviation	Variability in DIN			Variability in DRP			A. Limitation status, all flows				B. Limitation status, flows < median			
	Season	Flow	Trend	Season	Flow	Trend	DIN	DRP	DIN/DRP	Status	DIN	DRP	DIN/DRP	Status
manawatu_hop	yes	pos		no	no		326	21	15.9	P	171	19	10.2	co
mangatainoka_us_tir	yes	pos		no	no		741	7	104.8	P	631	7	95.7	P
mangapapa_troup	yes	pos	neg	no	no	neg	230	13	18.0	P	112	12	11.0	co
pohangina_mais	yes	pos		no	weak		38	12	3.2	N	19	13	1.5	N
manawatu_ug	yes	pos		no	pos		456	10	46.8	P	342	8	43.4	P
oroua_almadale	yes	pos		no	no		60	9	6.8	N	29	9	3.3	N
oroua_us_fei	yes		neg	no		neg	164	17	9.8	co				
oroua_ds_fei	no			no		neg	1362	18	77.8	P				
oroua_awahuri	no			no		neg	735	21	35.1	P				
manawatu_tc	yes	pos		no	pos		257	10	26.8	P	162	9	18.6	P
manawatu_us_pncc	yes	pos	neg	no	pos	neg	312	12	25.7	P	195	10	18.7	P
manawatu_ds_pncc	yes	pos		no	pos		602	17	34.7	P	518	15	35.4	P
manawatu_opik	yes	no		no	pos	neg	519	14	37.1	P	475	13	36.8	P
tokomaru_hb	weak	pos		no	no	pos	49	6	7.8	co	34	7	5.7	N
rangitikei_puk	yes	pos		no	no	pos	20	6	3.5	N	16	6	2.8	N
moawhango_waiouru	no	no		no	no		8	8	1.0	N	8	10	1.0	N
rangitikei_man	yes	pos		no	pos	pos	39	7	5.8	N	22	6	3.6	N
porewa_us_hun	yes			no			54	14	3.9	N				
porewa_ds_hun	yes			no			102	16	6.5	N				
rangitikei_one	yes	pos		no	no	pos	46	8	5.5	N	25	8	2.8	N
rangitikei_mk	yes	pos		no	no		48	12	4.1	N	28	12	2.4	N
mangawhero_doc	no	pos		no	neg	pos	10	14	0.7	N	7	18	0.4	N
makotuku_sh49	no	neg	neg	weak	neg	pos	191	9	20.2	P	269	14	19.4	P
mangawhero_us_oha	weak	pos	neg	no	no		153	15	10.1	co	92	14	6.1	N
mangawhero_ds_oha	yes	pos	neg	no	no		187	20	9.1	co	116	20	5.6	N
makotuku_rae	yes	pos		no	no	pos	295	7	41.3	P	238	8	29.8	P

Site abbreviation	Variability in DIN			Variability in DRP			A. Limitation status, all flows				B. Limitation status, flows < median			
	Season	Flow	Trend	Season	Flow	Trend	DIN	DRP	DIN/DRP	Status	DIN	DRP	DIN/DRP	Status
mangawhero_pakihi	yes	pos		no	no		200	13	15.5	P	152	13	11.7	co
whanganui_ds_gen	weak	no		no	no		11	27	0.4	N	12	26	0.5	N
whakapapa_ds_gen	weak	neg		no	no		25	24	1.0	N	33	22	1.5	N
waitangi_us_wai	no			no			275	30	9.0	co				
waitangi_ds_wai	no			no		neg	430	51	8.5	co				
makotuku_us_rae	yes	pos		no	neg		318	9	35.4	P	229	10	22.4	P
waikawa_nmr	yes	weak		no	no		48	10	4.8	N	35	9	3.6	N
ohau_gladstone	yes	pos		no	no		43	8	5.2	N	32	8	4.1	N
ohau_sh1	yes	no		no	no		200	9	22.1	P	180	8	23.9	P
ohau_haines	yes	no		no	weak		302	8	38.6	P	278	7	38.2	P

4.3.3 Nutrient limitation based on saturating concentrations

Using the McArthur et al. (2010) approach altered the outcome further. Using the complete dataset, including those sites with no linked flow record, 24 of the 58 sites were designated as co-limited, implying that addition of either N or P or both could potentially stimulate periphyton growth (Table 4-2, part A). Ten sites were assigned to a non-limited category because both DIN and DRP exceeded the saturation threshold in most surveys (>55% of all surveys). At flows lower than median, more sites were designated as co-limited (or a combination) (26 of the 47 sites with a flow record) (Table 4-2, part B). The difference between all data and low-flow data again reflected the generally lower DIN at lower flows. Only four sites were designated as having no nutrient limitation under low flows (oruakeretaki_sh2, manawatu_hop, manawatu_ds_pncc and manawatu_opik). Note that six sites with no limitation across all flows did not have an associated flow record.

Table 4-2: Summary of nutrient limitation status at each site based on occurrences of saturating concentrations of DIN and DRP. The first two columns show nutrient limitation based on DIN : DRP, copied from Table 4-1. Saturation thresholds were assumed to be 350 mg/m^3 for DIN and 15 mg/m^3 for DRP. Nutrient limitation status was assigned from the percentages of surveys in which DIN, DRP, both or neither exceeded the thresholds. See text for more details. Sites are arranged in order of their HMW site number (from upstream to downstream in successive catchments).

	Limitation based on ratios		A. Status based on all flows					B. Status based on flows < median				
			% surveys under saturation threshold for:					% surveys under saturation threshold for:				
Site	All	<med	neither	N only	P only	N, P	Status	neither	N only	P only	N, P	Status
makakahi_doc	N	N	0	3	0	97	co	0	3	0	97	co
mangatainoka_putara	N	N	0	3	0	97	co	0	3	0	97	co
mangatainoka_lars	co	N	0	3	0	97	co	0	6	0	94	co
tamaki_res	N		0	7	2	91	co					
mangatera_us_dan	N		63	35	2	0	none					
mangatera_ds_dan	co		94	6	0	0	none					
mangatainoka_huk	P	P	2	3	78	17	P	0	6	75	19	P
kumeti_tr	P	P	7	2	76	15	P	2	2	66	30	P
manawatu_weber	co	co	51	15	12	22	none-co	33	19	10	38	co - none
makakahi_ham	P	P	3	0	61	36	P	3	0	44	53	co-P
oroua_apiti	co	N	0	1	1	98	co	0	2	2	96	co
tamaki_ste	P	P	3	1	54	42	P-co	4	2	40	54	co-P
oruakeretaki_sh2	P	P	32	4	55	9	P	42	4	40	14	none-P
makuri_tuscan	P	P	9	0	90	1	P	11	0	89	0	P
pohangina_pir	N		0	3	0	97	co					
mangatainoka_scarb	P	P	2	0	92	6	P	0	0	92	8	P
tiraumea_nga	P	P	17	1	78	3	P	15	1	81	3	P
mangatainoka_pahiatua	P	P	6	0	94	0	P	5	0	95	0	P
mangatainoka_sh2	P	P	2	0	92	6	P	0	0	92	8	P
mangatainoka_ds_db	P	P	8	0	86	6	P	7	0	85	9	P
mangatainoka_us_pah	P	P	16	0	80	5	P	15	0	78	7	P

Site	Limitation based on ratios		A. Status based on all flows — % surveys under saturation threshold for:					B. Status based on flows < median — % surveys under saturation threshold for:				
	All	<med	neither	N only	P only	N, P	Status	neither	N only	P only	N, P	Status
mangatainoka_ds_pah	P	P	20	0	76	5	P	23	0	71	6	P
manawatu_hop	P	co	59	18	10	14	none	42	28	10	20	none-N
mangatainoka_us_tir	P	P	3	0	89	8	P	2	0	83	14	P
mangapapa_troup	P	co	13	20	39	28	P-co	13	22	22	42	co-P
pohangina_mais	N	N	1	31	2	66	co	2	36	2	60	co
manawatu_ug	P	P	23	2	46	29	P-co	9	4	46	42	P-co
oroua_almadale	N	N	2	11	11	76	co	4	11	2	84	co
oroua_us_fei	co		42	9	13	36	none-co					
oroua_ds_fei	P		53	1	45	1	none-P					
oroua_awahuri	P		56	5	32	7	none					
manawatu_tc	P	P	20	4	36	40	co-P	14	3	27	56	co
manawatu_us_pncc	P	P	31	4	31	33	co-none-P	20	7	25	48	co-P
manawatu_ds_pncc	P	P	59	2	32	8	none	50	2	36	11	none-P
manawatu_opik	P	P	46	7	34	14	none-P	41	7	37	15	none-P
tokomaru_hb	co	N	0	1	0	99	co	0	2	0	98	co
rangitikei_puk	N	N	0	4	0	96	co	0	5	0	95	co
moawhango_waiouru	N	N	0	22	2	76	co	0	21	2	76	co
rangitikei_man	N	N	1	3	0	96	co	0	4	0	96	co
porewa_us_hun	N		10	43	12	35	N-co					
porewa_ds_hun	N		16	37	14	33	N-co					
rangitikei_one	N	N	2	10	1	87	co	2	11	2	85	co
rangitikei_mk	N	N	11	20	2	67	co	12	25	2	61	co
mangawhero_doc	N	N	0	46	0	54	co-N	0	65	0	35	N
makotuku_sh49	P	P	10	18	10	61	co	21	28	14	37	co-N
mangawhero_us_oha	co	N	10	25	5	59	co	2	39	2	57	co
mangawhero_ds_oha	co	N	23	39	7	31	N-co	12	52	2	35	N-co
makotuku_rae	P	P	2	10	49	39	P-co	0	16	36	48	co-P
mangawhero_pakihi	P	co	13	23	16	47	co-N	7	31	10	53	co-N
whanganui_ds_gen	N	N	0	86	0	14	N	0	85	0	15	N
whakapapa_ds_gen	N	N	0	82	0	18	N	0	80	0	20	N
waitangi_us_wai	co		22	66	0	13	N					
waitangi_ds_wai	co		66	34	0	0	none					
makotuku_us_rae	P	P	5	14	54	26	P-co	6	21	36	36	P-co
waikawa_nmr	N	N	0	6	0	94	co	0	5	0	95	co
ohau_gladstone	N	N	0	3	1	96	co	0	2	0	98	co
ohau_sh1	P	P	0	5	25	70	co	0	6	21	72	co
ohau_haines	P	P	0	2	38	60	co	0	2	31	67	co

4.4 Discussion

This exercise demonstrated the difficulty in classifying sites based on nutrient limitation of periphyton growth because (a) the designation depends on whether the data are filtered to exclude high flows; (b) using ratios only without taking account of concentrations may give misleading results (i.e., DIN or DRP may be limiting in theory according to the N:P ratio but in fact both are available in excess of requirements); and (c) even if concentrations are taken into account, the concentration thresholds above which no effect on periphyton is expected are uncertain.

A further issue is that in the above analysis we did not take season into account. While we can assign nutrient limitation status to a site based on numbers of occasions of N-, P-, co- and no limitation, those occasions will almost always have a seasonal pattern. The 7-year length of the Horizons dataset allows these seasonal patterns to be seen clearly, and an example is shown in Table 4-3 for manawatu_hop. An informative analysis (which is currently beyond the scope of this report) would be to examine nutrient-limitation temporal patterns at multiple sites in relation to chlorophyll *a* and preceding flows. The patterns of nutrient limitation over time are not explicitly accounted for in the within-site analysis except as concentrations of DIN and DRP as predictor variables (see Section 7).

Table 4-3: Seasonal variation in nutrient limitation status in the Manawatu at Hopelands. Numbers of occurrences of N-, P-, co_ and no limitation are shown for each calendar month, for all occasions when the flow was less than median flow.

	Number of surveys											
Status	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
N-limited	1	6	6	3	3							
P-limited									1	3	2	1
co-limited	5	1	2	1	1						1	3
no limitation				5	4	4	4	3	2		2	5

With regard to point (c) above (thresholds for nutrient saturation), the threshold of 350 mg/m^3 used for DIN was based on wider literature. A recent analysis of data from Canterbury rivers suggests that this threshold is within a realistic range. Across 17 hill-fed rivers in Canterbury, annual maximum chlorophyll *a* in a dry year (i.e., potential for maximum biomass to be achieved at all sites in terms of accrual period) was highest (up to 600 mg/m^2) at sites with DIN < 300 mg/m^3. Sites with higher DIN had similar or lower maximum chlorophyll *a*, even though DRP was also generally higher at sites with high DIN (>40% of the variance in DRP explained by DIN) (Kilroy et al. 2017).

It is not usually possible to define levels of chlorophyll *a* that correspond to nutrient saturation. One reason is that "nutrient saturation occurs when the availability of a nutrient increases to a point at which another factor critical to growth becomes limiting" (Earl et al. 2006), which implies that nutrient saturation can occur under a range of conditions. Thus, how much biomass is supported at a site where both DIN and DRP are at saturating concentrations (assuming no other limiting factors) will depend on, for example, light, temperature and rates of herbivory, and whether any micronutrients are limiting growth. Using a very large dataset, Dodds et al. (2002, 2006) estimated levels of TN and TP corresponding to saturation by idenitifying a breakpoint in the relationships between TN or TP and chlorophyll *a*. The range of mean annual chlorophyll *a* at the breakpoint for TP, for example, spanned orders of magnitude.

A second reason for inability to define chlorophyll *a* levels corresponding to nutrient-saturated periphyton growth is that saturating nutrient concentrations are related to rates of removal of DIN or DRP by uptake by organisms or other processes such as denitrification. When the rate of uptake no longer responds to additions of DIN or DRP, then saturation is assumed. Thus, much of the literature on nutrient saturation in rivers does not consider periphyton standing crop (measured as chlorophyll *a*) but focuses on uptake metrics that indicate saturation of uptake (Bernot et al. 2006, Earl et al. 2006).

A more useful concept than a chlorophyll *a* level that suggests nutrient saturation may be the chlorophyll *a* level representing the carrying capacity at a site. The 600 mg/m^2 maximum chlorophyll *a* observed in Canterbury in a dry year (see above) could be close to an absolute maximum carrying capacity based on data from an international dataset in which about 95% of maximum chlorophyll *a* values were less than 600 mg/m^3 (Dodds et al. 2002)

Regarding specifying nutrient limitation, the final version of the four assessments (method 4 in Section 4.2) is probably the most pragmatic because it captures nutrient conditions at times when periphyton is most likely to be accruing (in low flows). We propose to use that assessment as the site classification in subsequent analyses. We simplified the nutrient limitation groups by assigning sites with a combined status to the most common status (e.g., co–N-limitation became co-limitation). Classes are indicated by the colour coding in Table 4-2.

McArthur et al. (2010) suggested that a much lower flow threshold could be used to define low flows under which periphyton to accrue (i.e., the 20th percentile rather than the median or 50th percentile). In view of the results of the within-site flow – periphyton analysis (Section 3) site-specific flow thresholds for accrual might provide a more accurate picture. Because the flow magnitudes that remove periphyton exceed 1.5 x median flow at most sites, a threshold of median flow is a conservative alternative and corresponds to periods when flows are unlikely to be suppressing periphyton biomass.

5 Test of the Biggs (2000a) relationships

Key messages

- We used the Horizons periphyton dataset as independent data to test published equations linking annual maximum chlorophyll *a* to DIN or DRP concentrations and accrual period calculated using the frequency of flows exceeding 3 x median flow (Biggs 2000a). Our aim was to determine whether prediction and ultimately management of periphyton chlorophyll *a* simply requires knowledge of DIN and DRP concentrations and accrual period. A further question was whether using accrual period based on ***effective flow*** would improve predictions.
- Across all sites, predictions of maximum chlorophyll *a* from the Biggs (2000a) equations were only weakly or not correlated with observed chlorophyll *a*. Relationships between nutrient concentrations and periphyton standing crop are unlikely to be accurately characterised in the Horizons region by the Biggs (2000a) equations.
- Using the effective flow to calculate accrual period did not improve the predictions.
- The Biggs (2000a) equations were expected to perform weakly given that almost half the annual mean DIN values in the Horizons dataset exceeded the range underpinning the relationship. Predicting beyond the range of the original data is unlikely to be accurate.
- The Biggs (2000a) equations were derived using data from a smaller range of river type than is found in the Horizons region, where hydro-physical characteristics are variable.
- The weak performance of the Biggs (2000a) equations in predicting annual maximum chlorophyll *a* across the Horizons region indicated the need for new predictive relationships, with additional variables considered for inclusion.

5.1 Introduction

Biggs (2000a) developed relationships between maximum annual chlorophyll *a* and mean DIN or DRP and accrual time using data from 30 sites in gravel bed rivers throughout New Zealand. Relationships using DIN and DRP were recommended for use at DIN-limited and DRP-limited sites, respectively. The relationships explained about 70% of the variance in maximum chlorophyll *a* extracted from time series of data up to 14 months long. The relationships are:

$$\mathrm{Log_{10}(maximum\ chlorophyll\ } a) = 4.285 * (\log_{10} D_a) - 0.929 *(\log_{10} D_a)^2 + (0.504 * \log_{10} \mathrm{DIN}) - 2.946$$

$$\mathrm{Log_{10}(maximum\ chlorophyll\ } a) = 4.716 * (\log_{10} D_a) - 1.076 *(\log_{10} D_a)^2 + (0.494 * \log_{10} \mathrm{DRP}) - 2.714$$

In Biggs (2000a), accrual time (D_a) was defined as 365/FRE3, where FRE3 was the annual frequency of high flow events exceeding three times the median flow. FRE3 can be calculated in different ways (Booker 2013). For example, a common practice is to count successive high flow events as a single event if the 3 x median thresholds are less than 5 days apart. Biggs (2000a) did not use such a filter. Therefore, all flow events greater than 3 x median flow were counted for the initial comparison. All flow data were extracted as daily mean flows.

5.2 Methods

Up to 48 periphyton sites with both a flow record and water quality data were available for the analysis in each year. The data were divided into hydrological years (from July to June). Accrual time, as defined above, was calculated for each year. Seven complete hydrological years were available for the analysis (2009-10 to 2015-16) (i.e., we omitted data from part-years starting in December 2008 and ending in December 2016 or April 2017).

We first ran separate tests in each year to predict $\log_{10}$maximum chlorophyll *a* from nutrients and flows measured in the same year using the original equations in Biggs (2000a). Second, we re-ran the tests, substituting accrual period based on "effective flow" (as defined in Table 3-1) for accrual period based on 3 x median flow.

For each scenario and in each year, we compared the relationship between observed and predicted $\log_{10}$ (maximum chlorophyll *a*). The performance of the models was assessed using both the R^2 value and the root mean squared deviation (RMSD) (Pineiro et al. 2008). R^2 quantifies the proportion of the variance of the observed value explained by the predicted value). Root-mean-squared-deviation (RMSD) is calculated as the square-root of the average squared deviations (i.e., predicted – observed) at all sites in each year and is an absolute representation of the accuracy of the predictions in the same units as the observations and predictions.

We also examined the performance of the equations within the flow groups and nutrient limitation groups defined in Sections 3 and 4.

5.3 Results

Using all the available data, predictions of annual maximum chlorophyll *a* using both the Biggs (2000a) DIN and DRP equations were weakly related to observed values (maximum R^2 of 0.33 in Table 5-1). Correspondence between observed and predicted was especially weak using the DRP equations (maximum R^2 of 0.10). Correspondence generally did not change when effective flow was substituted for 3 x median flow in the calculation of accrual period. For example, for the DIN equation, the mean R^2 over the seven years was 0.19 using 3 x median and 0.18 using effective flow. Plots of the relationships are shown in Figure 5-1.

RMSD values were high compared to the range of observed $\log_{10}$chlorophyll *a* (0 to 2.70) and were about 40% higher in the effective flow relationships than in the 3 x median relationships. The discrepancy between observed and predicted was > 1 (i.e., more than an order of magnitude when back-transformed from a log scale) in 31 of 42 regressions using effective flow (Table 5-1).

Across the flow-sensitive sites, correspondence between observed and predicted maximum chlorophyll *a* was similar to that for all sites (Table 5 1). The relationships were stronger for the flow-insensitive sites, particularly using the DIN equation. However, numbers of sites in the flow-insensitive group were small (n = 12 – 14 for 3 x median and n = 6 – 7 using effective flow). Therefore, the relationships are less reliable than those for the larger groups of sites.

Table 5-1: Summary results of linear regression on observed chlorophyll *a* and predictions from the Biggs (2000a) equations, with sites partitioned by flow response. RMSD is the root mean squared deviance and is an absolute measure (in the same units as $\log_{10}$chlorophyll *a*) of the average discrepancy between observed and predicted values. R^2 is adjusted R^2.

Model	Accrual period from	Year	All sites in dataset			Flow-sensitive sites			Flow-insensitive sites		
			n	R^2	RMSD	*n*	R^2	RMSD	*n*	R^2	RMSD
DIN	3 x median	2009-10	35	0.10	0.76	23	0	0.67	11	0.16	0.91
		2010-11	40	0.09	1.00	25	0.05	1.03	12	0.51	0.74
		2011-12	41	0.14	0.84	26	0.08	0.86	12	0.36	0.78
		2012-13	43	0.20	0.78	27	0.09	0.82	12	0.54	0.68
		2013-14	48	0.26	0.72	30	0	0.77	14	0.68	0.64
		2014-15	48	0.24	0.81	30	0.15	0.87	14	0.53	0.73
		2015-16	47	0.33	0.81	30	0.24	0.87	14	0.63	0.76
	Effective	2009-10	30	0.10	1.04	23	0.07	0.94	6	0.36	1.57
		2010-11	34	0.11	1.30	25	0.13	1.20	6	0.28	1.47
		2011-12	34	0.20	1.15	26	0.13	1.07	6	0.84	1.51
		2012-13	36	0.12	1.06	27	0.21	1.00	6	0.55	1.53
		2013-14	40	0.21	1.02	30	0.16	0.98	7	0.44	1.36
		2014-15	40	0.28	1.16	30	0.25	1.13	7	0.55	1.46
		2015-16	39	0.22	1.20	30	0.17	1.17	7	0.72	1.47
DRP	3 x median	2009-10	35	0	0.77	23	0.10	0.63	11	0	0.92
		2010-11	40	0	0.91	25	0	0.90	12	0.17	0.64
		2011-12	41	0	0.84	26	0.05	0.81	12	0	0.68
		2012-13	43	0	0.81	27	0.05	0.71	12	0.08	0.70
		2013-14	48	0	0.78	30	0	0.69	14	0.12	0.79
		2014-15	48	0	0.85	30	0	0.75	14	0	0.83
		2015-16	47	0.10	0.75	30	0.07	0.69	14	0.39	0.77
	Effective	2009-10	30	0	0.94	23	0	0.79	6	0.17	1.53
		2010-11	34	0	1.15	25	0	1.02	6	0	1.36
		2011-12	34	0	1.02	26	0	0.93	6	0.87	1.33
		2012-13	36	0	0.97	27	0	0.83	6	0	1.49
		2013-14	40	0.04	0.96	30	0.05	0.81	7	0	1.44
		2014-15	40	0.08	1.06	30	0.12	0.92	7	0.30	1.48
		2015-16	39	0.07	1.04	30	0.07	0.92	7	0.36	1.46

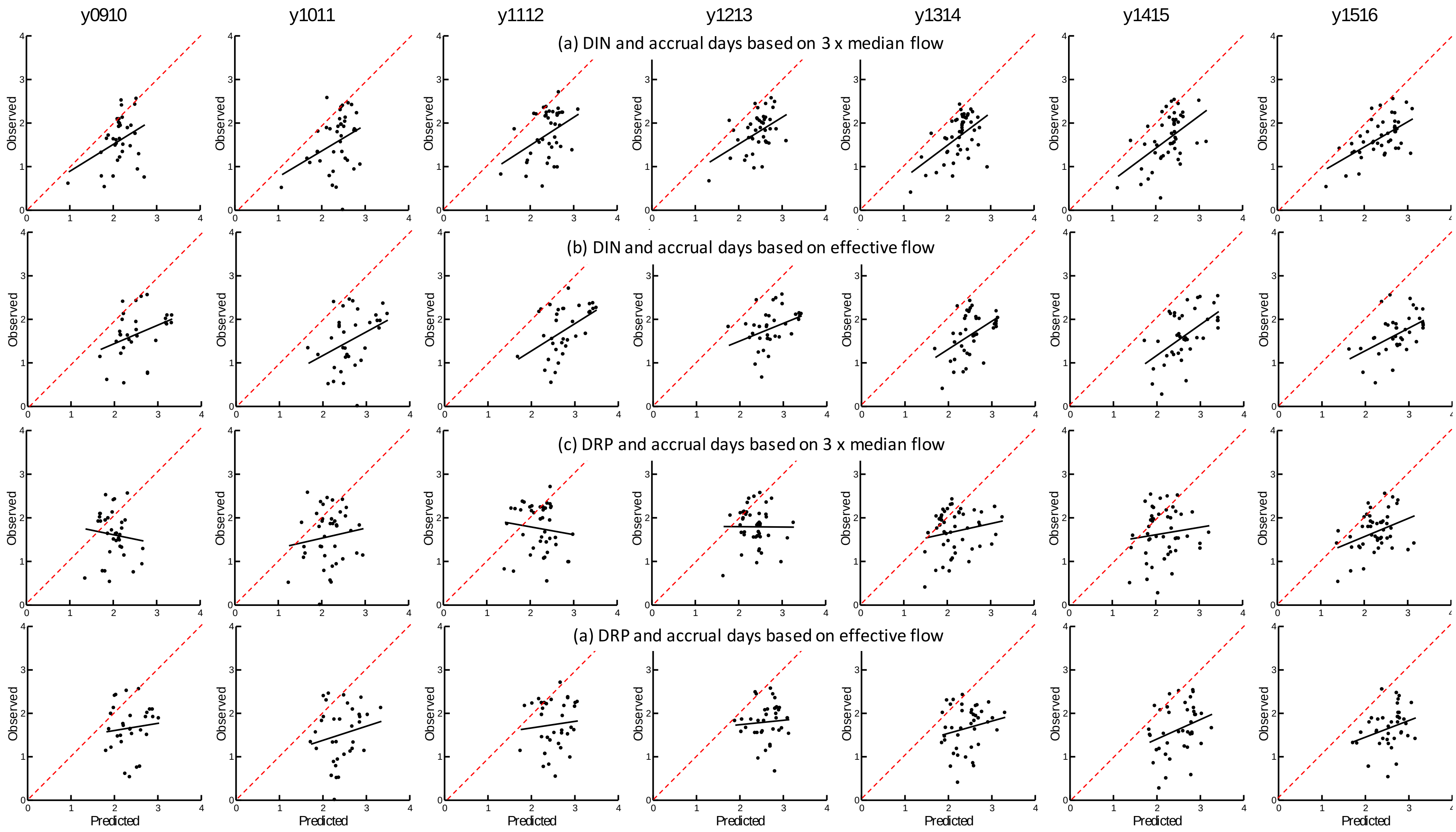


Figure 5-1: Observed annual maximum chlorophyll *a* plotted against annual maxima predicted using the Biggs (2000a) equations. Predictions were made for each complete year of the Horizons dataset. All data were log_{10}-transformed (in mg/m^2). Black lines are best linear fits through the data; red dashed lines are 1 : 1.

Most sites in the dataset were classified as either P-limited or co-limited, using the simplified classification based on Table 4-2 (see Section 4). Groups of sites that were N-limited or not limited were too small for regression analyses. Given the weak performance of models using accrual time based on effective flow, we looked only at prediction using the original Biggs (2000a) models.

For the equations based on DIN, the subset of sites assessed as co-limited had slightly higher mean R^2 in the DIN predictions (mean R^2 = 0.19 and 0.25 for all sites and co-limited sites, respectively). RMSD was similar in the two groups. There was large year-to-year variability R^2 in the small group of P-limited sites.

For the equations based on DRP, mean R^2 across the co-limited sites was higher than across all sites (mean R^2 = 0.01 and 0.11 for all sites and co-limited sites, respectively), though R^2 in the co-limited site group varied across years. RMSD was similar in the two groups. In the P-limited group, R^2 values for the first three years were relatively high (0.3 to 0.56) (Table 5-2). However, examination of plots showed that these relationships were negative, driven mainly by low observed maximum chlorophyll *a* at kumeti_tr compared to the predictions of much higher maxima because of high DRP concentrations. RMSD was lowest in the P-limited sites using the DRP equation because all the predictions were those clustered around the 1 : 1 line in Figure 5-1c because of their lower DRP values. At sites with higher DRP, chlorophyll *a* was generally overpredicted.

Table 5-2: Summary of linear regression on observed chlorophyll *a* and predictions from the Biggs (2000a) equations, with sites partitioned by nutrient limitation. Co- and P-limited sites were those defined in the right-hand column of Table 4-2, simplified as noted in the text. RMSD as in Table 6-4. R^2 is adjusted R^2. The original Biggs (2000) models were used (i.e., with days of accrual calculated using 3 x median flow). N-limited sites were not considered because the number of sites was too small (n = 4).

		All sites in dataset			Co-limited sites			P-limited sites		
Model	**Period**	**n**	**R^2**	**RMSD**	**n**	**R^2**	**RMSD**	**n**	**R^2**	**RMSD**
DIN	2009-10	35	0.10	0.76	21	0.20	0.79	8	0.18	0.79
	2010-11	40	0.09	1.00	22	0.22	1.03	10	0.38	0.73
	2011-12	41	0.14	0.84	23	0.18	0.91	10	0.47	0.63
	2012-13	43	0.20	0.78	24	0.17	0.76	10	0	0.77
	2013-14	48	0.26	0.72	26	0.40	0.72	13	0	0.71
	2014-15	48	0.24	0.81	26	0.16	0.86	13	0	0.64
	2015-16	47	0.33	0.81	26	0.43	0.75	13	0	0.91
DRP	2009-10	35	0	0.77	21	0	0.80	8	0.41	0.71
	2010-11	40	0	0.91	22	0.08	0.93	10	0.30	0.62
	2011-12	41	0	0.84	23	0	0.88	10	0.56	0.52
	2012-13	43	0	0.81	24	0.05	0.76	10	0.13	0.48
	2013-14	48	0	0.78	26	0.22	0.78	13	0	0.51
	2014-15	48	0	0.85	26	0.09	0.87	13	0	0.42
	2015-16	47	0.10	0.75	26	0.35	0.73	13	0	0.58

5.4 Discussion

The results of this test of the Biggs (2000a) equations both agreed and contrasted with the outcome of a similar test across 17 river sites in the Canterbury region. Across the Canterbury sites, as in the present test, the original equations overpredicted maximum chlorophyll *a* at most sites and there was weak correspondence between observed and predicted maximum chlorophyll *a* (Kilroy et al. 2017). However, the present results contrasted with those on the Canterbury dataset because substituting the effective flow for 3 x median improved the correlation between observed and predicted annual maximum chlorophyll *a* in the Canterbury dataset, particularly for the DRP equation. Nevertheless, in the Canterbury dataset, the observed and predicted values still did not correspond closely at most sites, and new relationships were developed for the Canterbury region instead of the Biggs (2000) equations.

5.4.1 Effect of river type

The original Biggs (2000a) relationships were derived from a set of sites with similar hydro-physical characteristics, which were described as follows: "… all sites were in streams and rivers flowing from hill-country watersheds where snowmelt affected flow regimes for 3 months per year, and lakes or large springs did not dominate flow regimes."

Discussions with Biggs indicate that these sites were interpreted as those where periphyton biomass was strongly influenced by high flows. The analysis in Section 3 suggests that the sites in the Horizons dataset cover a wider range of hydro-physical characteristics. Within the Horizons dataset, sites belonging to the flow-sensitive group (and especially group A, with low flood removal thresholds) may fall into the same category at the sites used by Biggs (2000a). At flow-sensitive sites, accrual period explained 20% to 53% of the variance in chlorophyll *a* over time (Table 3-1), even though the flow removal threshold varied among sites. Accordingly, we might expect to see better correspondence between observed and predicted maximum chlorophyll *a* using the flow-sensitive sites.

However, the results using the DIN equation showed the opposite pattern, with roughly equivalent correspondence between observed and predicted in the flow-sensitive group compared to the whole dataset, but stronger relationships within the flow-insensitive group. One explanation for this is that at sites where periphyton is not regularly removed by high flows, it more likely that chlorophyll *a* can accrue to its maximum value (or "carrying capacity", Biggs and Close 1989), which is strongly influenced by nutrient availability. In other words, strong correspondence between observed and predicted values in the DIN equation simply reflect a stronger relationship with DIN than with accrual period defined from 3 x median flow. Relationships with nutrient concentrations across sites are explored further in Section 6.

5.4.2 Range of nutrient values in the original Biggs (2000a) dataset

One of the limitations of the Biggs (2000a) equations is that they were developed along a relatively small gradient of DIN compared to that in the Horizons dataset (Table 5-3). Across the 302 annual datasets (from 2009-10 to 2015-16, up to 50 sites with flow records), DIN exceeded the maximum value in the Biggs (2000a) dataset (232 mg/m^3) in 143 cases (47%). There were eight corresponding exceedances for DRP (2.6%). Therefore, the problem of predicting outside the range of the original data is particularly severe for DIN. In theory, we should obtain better predictions from the equations by restricting data to that in the original range. We therefore plotted observed versus expected $\log_{10}$maximum chlorophyll *a* at sites with DIN < 235 mg/m^3 and with DIN > 235 mg/m^3 (Figure 5-2).

Table 5-3: Ranges mean annual DIN, DRP and days of accrual in the Biggs (2000a) dataset compared to the Horizons dataset.

	Biggs (2000a)			Horizons dataset		
Statistic	**DIN (mg/m³)**	**DRP (mg/m³)**	**Days of accrual**	**DIN (mg/m³)**	**DRP (mg/m³)**	**Days of accrual**
Minimum	6	1	10	1	3	10
Maximum	232	32	183	4467	498	396
Median	54	4	21	241	10	22
Mean	81	8	30	384	17	37
Standard deviation	70	9	33	430	31	63

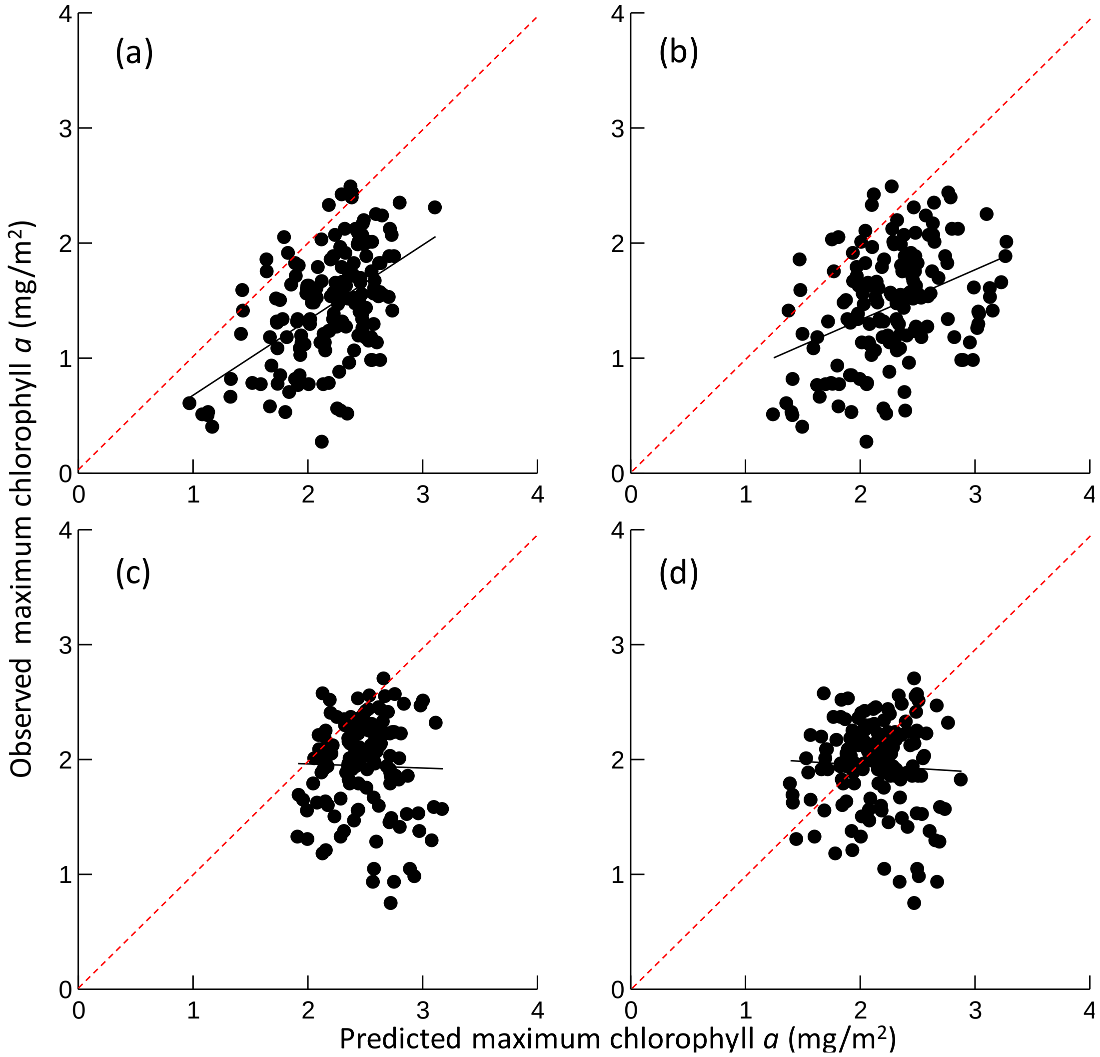


Figure 5-2: Observed annual maximum chlorophyll *a* plotted against predicted values using the Biggs (2000) equations on two datasets. Plots (a) and (c) show results for the DIN equation, (b) and (d) for the DRP equation. Plots (a) and (b) used a dataset in which all values of mean DIN > 235 mg/m^3 were excluded; in plots (c) and (d) all values of DIN were > 235 mg/m^3. In the dataset used by Biggs (2000a) to develop the relationships the range of DIN was 6 to 232 mg/m^3.

The plots confirmed that the observed and predicted values corresponded more closely in the low range of DIN than in the high range. At DIN < 235 mg/m^3, R^2 for the DIN equation was 0.27 and for the DRP equation 0.14 (Figure 5-2 a, b). With DIN > 235 mg/m^2, there was no correspondence between observed and predicted values (R^2 = 0 in both the DIN and DRP equations) (Figure 5-2 c, d).

This exercise illustrated that using the Biggs (2000a) equations indiscriminately represents mis-use of the original models. One of the reasons that the relationships worked well in the first place was likely that all DIN values were within the generally accepted range of non-saturating DIN concentrations and that there was limited variation in stream geomorphology across the 30 sites. Taking the original range of the data into account, Figure 5-2a, b also shows that the equations do not work particularly well for predicting actual chlorophyll *a* maxima across all sites in the Horizons region because the equations over-predicted by up to more than an order of magnitude. However, the predictions were accurate in some cases, and there was some correspondence between observed and predicted. The generally weak performance of the models at a regional scale suggests that additional factors important in driving periphyton biomass need to be included. Alternatively, nutrient availability and accrual time are not appropriately captured by mean annual DIN or DRP and accrual period as defined by Biggs (2000a).

Note that this test of the Biggs (2000a) equations was a true test of the model on independent data. In the original relationships, about 70% of the variance in log^{10}maximum chlorophyll *a* was explained by a combination DIN or DRP and accrual period. No cross-validation tests were performed at the time of the analysis, and, had they been, a lower R^2 would have been expected than the reported 70%.

6 Between-site relationships

Key messages

- Aims were to explore (a) correlations between peak chlorophyll *a* and averaged DIN or DRP across sites and years; and (b) relationships between peak chlorophyll *a* and a combination of environmental variables, in both cases applying a space-for-time approach using linear regression. We tested relationships in annual and 3-year datasets and a 7-year dataset.
- "Peak chlorophyll *a*" was annual maximum chlorophyll *a* for annual datasets, and the 92nd percentile of chlorophyll *a*, the for multi-year datasets. The latter is the metric used in the periphyton attribute of the National Policy Statement for Freshwater Management (NPS-FM).
- For aim (a):
  - DIN was significantly and positively related to peak chlorophyll *a* in most time periods. Relationships were especially strong across sites classed as flow-insensitive (i.e., where a threshold for flows effective for removing periphyton could not be identified);
  - despite significant relationships, separate tests of predictive ability (cross-validation tests) showed that mean DIN in isolation from other variables was not a good predictor of peak chlorophyll *a* across sites within the Horizons region;
  - mean DRP was weakly or not correlated with peak chlorophyll *a* in all periods.
- For aim (b):
  - in addition to mean DIN and DRP, potential predictor variables included water conductivity, river bed sediment composition, mean water temperature, and mean accrual period (based on both 3 x median and effective flow);
  - generally, the strongest models in each time period included **DIN**, **conductivity** and **accrual period** as predictors. The initial models explained at least 50% of the variance in peak chlorophyll *a* across sites in all time periods;
  - accrual period calculated from the effective flow always produced stronger relationships than accrual period from 3 x median flow;
  - some models also included terms for substrate, water temperature and DRP;

Key messages (continued)

- leave-one-out cross-validation (a robust method for evaluating the predictive ability of models) produced encouraging results, with high proportions of variance in observed chlorophyll *a* explained by predicted chlorophyll *a* for some periods (e.g., 75% in 2012 - 2015).
- Models for the 3-year datasets performed better than the annual datasets with means of 63% vs. 55% explained respectively across all 3-year and annual periods, and 67% for the 7-year dataset;
- substituting total nitrogen (TN – all N in a sample including organic particles) for DIN produced slightly stronger relationships, which, again, were optimised if accrual period based on effective flow was included;
- substituting land-cover variables such as percentage of the catchment under intensive farmland (which is correlated with mean DIN) for DIN did not improve the models;
- the best models included all of the available sites (not smaller subsets).

▪ The models may be useful for (a) predicting likely chlorophyll *a* at new sites or at the same sites under different scenarios, such as reduced flood frequency or increased nutrient concentrations; and (b) setting nutrient limits. The error in each model was determined.

▪ Conductivity was highlighted in all strongly performing models as having a positive effect on chlorophyll *a*. This points to either a direct effect on periphyton chlorophyll *a* (e.g., via algal community composition) and/or a positive feedback into other factors such as nutrient availability. Conductivity was weakly associated with DIN across the region, suggesting that the cause of conductivity variation was not strongly linked to catchment DIN losses (e.g., from land use practices) but was more likely a function of underlying catchment hydrogeology.

6.1 Introduction

The Horizons periphyton monitoring programme was initially designed to provide data suitable for developing a regional model to supersede the Biggs (2000a) relationships. The monitoring programme included sites that encompassed a range of DIN, DRP and river flow variability (summarised as flood frequency) across the landscape. In a heterogeneous landscape, a range of other variables that could potentially affect periphyton also vary from site to site, and were measured as part of the monitoring programme. These include bed substrate composition (a product of catchment geology and topography), water conductivity (also influenced by geology), degree of shade (e.g., site aspect and riparian shading), and temperature (which can reflect altitude, shading and source of water).

In this part of the analysis we first looked for simple linear regression relationships between chlorophyll *a* and both DIN and DRP across multiple sites. The seven-year dataset enabled tests within individual years (representing different hydrological conditions) and across multiple years. In the second part of the analysis, we used a regression approach to develop new relationships that included additional potentially explanatory variables such as substrate composition.

6.2 Between-site relationships with DIN and DRP

6.2.1 Methods

Before running this section of the analysis we removed the fortnightly sampling data (see Section 2.1.1). To do this we took the first sample in a month for any site–year–month combination that had more than one sample in it. The periphyton data were divided into hydrological years (from July to June). Hydrological statistics for each year were calculated from June to June so that the metrics included the month preceding the start of the periphyton data. Seven complete hydrological years were available for the analysis (2009-10 to 2015-16) (i.e., we omitted data from part-years starting in December 2008 and ending December or April 2017). We also divided the data into three-year blocks (e.g., June 2009 to June 2012, June 2010 to June 2013, and so on). Finally, we compiled a dataset of all of the data (means over seven years).

Maximum chlorophyll *a* was identified in each hydrological year at each site, and the 92nd percentile of chlorophyll *a* was calculated for the multi-year datasets. The 92nd percentile is the metric used in the periphyton attribute of the NPS-FM (NZ Government 2017) and is therefore relevant to river management at a national level. Relationships between maximum chlorophyll *a* or the 92nd percentile of chlorophyll *a* and the geometric mean values of DIN or DRP over the period of interest were explored using simple linear regression. Data were log-transformed prior to analysis.

The regressions were performed first using all of the available data in each time period and then on subsets of sites based on chlorophyll *a* responses to flow and on nutrient limitation status. For flow responses, one subset comprised all sites at which we identified a definite threshold for periphyton removal (Groups A and B combined, in Table 3-1). The second subset included the remaining sites with a flow record at which no threshold for periphyton removal was evident. For nutrient limitation status we used a simplified version of the nutrient limitation categories based on saturating concetrations at low flows (see Section 4.3.3) (four categories, N, P, co and none).

The fit of each model in each dataset was assessed using **leave-one-out cross-validation**. In this procedure, the independent variables (DIN or DRP in this case) are used to generate a series of models omitting one datapoint each time. Each model is used to predict the value of the dependent variable (chlorophyll *a*) for the omitted datapoint. Observed values are plotted against predicted values and several statistics can be computed to allow assessment of the model fit (i.e., accuracy and precision). Useful statistics are:

- the coefficient of determination, R^2, which is a measure of the proportion of variance in the observed values explained by the predicted values;
- the root mean square deviation (RMSD), which is an absolute measure of the difference between predicted and observed values, in the same units as the dependent variable (i.e., $\log_{10}$chlorophyll *a*). The lower the value the better;
- Nash Sutcliffe Efficiency (NSE) is commonly used to assess predictive power in hydrological models (Nash and Sutcliffe 1970). NSE ranges from $-\infty$ to 1, where the

closer the number is to 1, the better model fit. NSE = 1 indicates perfect model fit, 0 indicates that model predictions are as accurate as the mean of the observed data and negative values indicate that the mean is a better predictor than the model. NSE is generally proportional to R^2, but is specifically used to quantify how well a model simulation predicts the outcome variable. As well as testing the correlation between observed and predicted values, NSE accounts for correspondence of values (i.e., the slope and intercept in the relationship). Unlike R^2, NSE can take negative values.

6.2.2 Results

Plots of log-transformed mean (geometric) DIN and DRP against log-transformed annual maximum chlorophyll *a* or the 92nd percentile of chlorophyll *a* (for multiple years) indicated variability in the data. The plots of DIN suggested generally linear relationships with chlorophyll *a* with a few outlying points at lower DIN values (Figure 6-1). Patterns in DRP were less distinct. In some years, the fitted distance-weighted least-squares smoothing line suggested increasing chlorophyll *a* as DRP increased, with a flattening of the relationship at higher values (e.g., y1114 in Figure 6-2), but there were few datapoints in the higher range of DRP.

Linear regression confirmed statistically significant ($P < 0.001$) linear and positive relationships between in all of the time periods between annual maximum chlorophyll *a* or the 92nd percentile of chlorophyll *a* vs. geometric mean DIN (Table 6-1). In contrast, and as expected from the plots, linear relationships between chlorophyll *a* and DRP were weak and not statistically significant in all periods tested (Table 6-1).

The proportion of variance in chlorophyll *a* explained by DIN (i.e., R^2) ranged from 17% in the three-year period from July 2010 to June 2013, to 46% from July 2011 to June 2014. In the three-year datasets, the proportion of variance explained was higher in relationships based on the two subsets of sites defined by responses to high flows (flow-sensitive and flow-insensitive sites as defined in Section 3.2.1) than over the whole dataset (mean R^2 of 0.28, 0.41, and 0.59 in the whole dataset, flow sensitive and flow-insensitive sites, respectively; Table 6-1).

Although chlorophyll *a* vs. DIN produced many statistically significant relationships, their predictive ability was generally poor. The NSE statistic from the cross-validation tests was negative for all periods when all the data were used (Table 6-1). Negative NSE means that the regression equation cannot predict at individual sites any better than just using the mean of all of the data (vs. NSE of 1, which is a perfect fit).

For the flow-sensitive sites NSE was greater than zero only in 2011-12 (NSE = 0.48). The higher R^2 values in regressions across flow-insensitive sites, especially in three-year periods, produced better predictive ability (NSE range up to 0.37). The relationship between seven-year mean DIN and the 92nd percentile of chlorophyll *a* flow was strong, with good predictive ability (NSE = 0.58). However, the number of sites was small ($n = 14$) (Table 6-1). Cross-validation plots for DIN versus maximum or 92nd percentile of chlorophyll *a* in all time periods are shown in Appendix D.

For the nutrient limitation groups, the numbers of sites in N-limited or not limited (none) groups were too small for meaningful regression analysis ($n = 4$). Regressions on the co-limited group (n = up to 26) and P-limited group (n = up to 13) produced weak relationships with both DIN and DRP (maximum R^2 of 0.3) and negative NSE values. In other words, the relationships were no stronger than those across the all the sites, or within the flow-sensitive sites (data not shown).

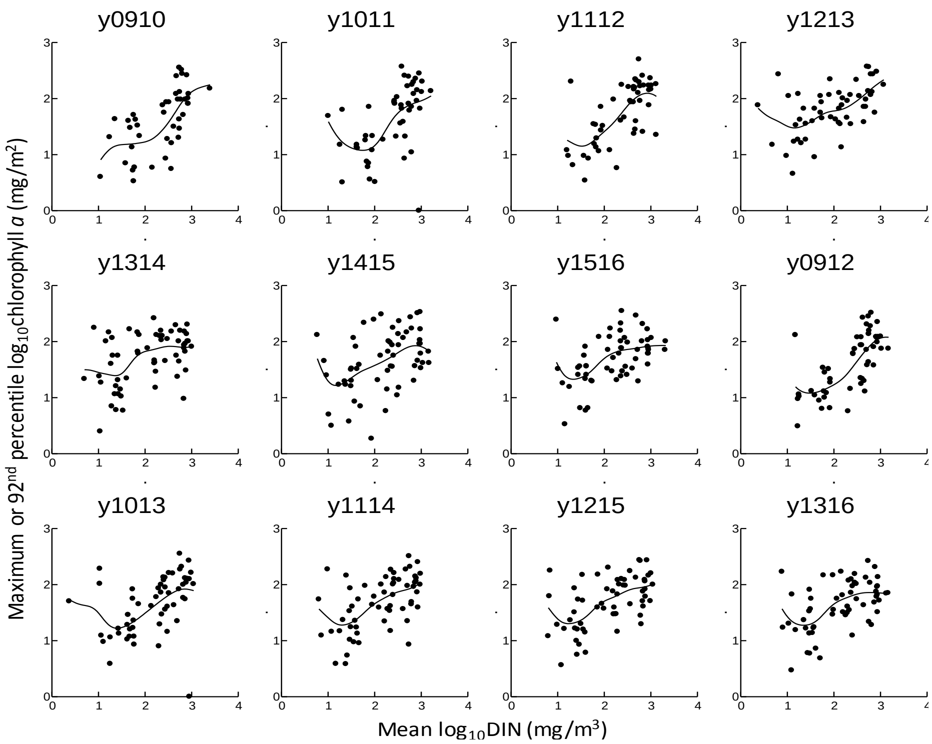


Figure 6-1: Scatter plots of log-transformed chlorophyll *a* against DIN in annual and three-year periods. Chlorophyll *a* was the maximum value for annual plots and the 92nd percentile for three-year periods. Distance-weighted least-squares smoothing lines (tension 0.5) show the general pattern of the relationships. We used linear regression in data analysis and the DWLS lines suggest that linear regression was generally appropriate.

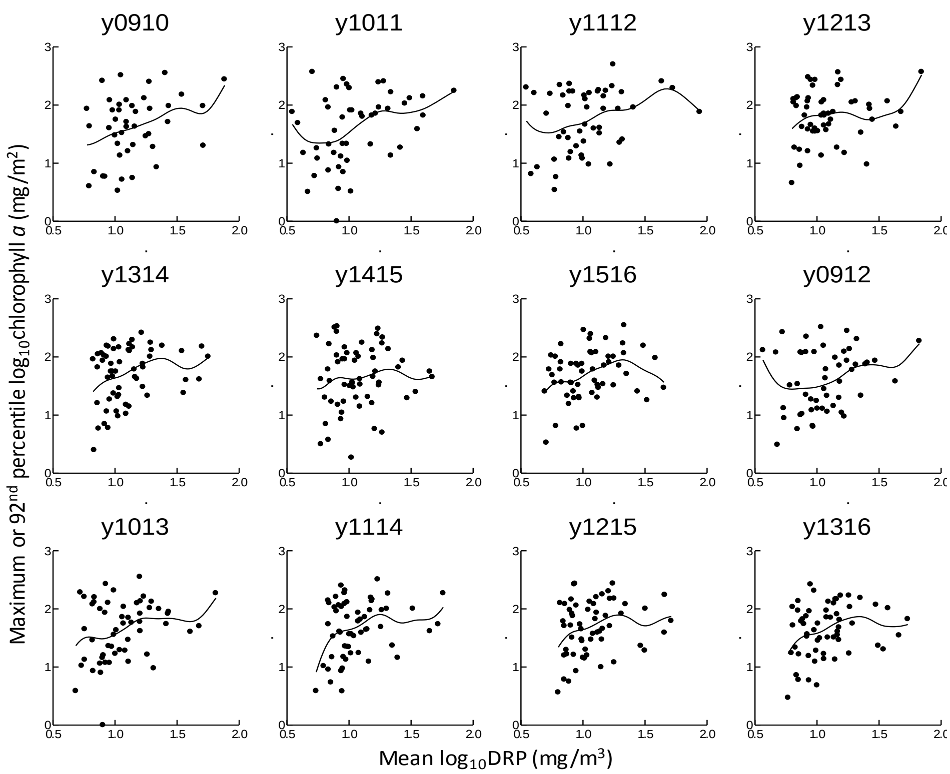


Figure 6-2: Scatter plots of log-transformed chlorophyll *a* against DRP in annual and three-year periods. For notes see Figure 6-1.

Table 6-1: Summary results of linear regressions of chlorophyll *a* versus DIN and DRP across various periods. Chlorophyll *a* is the annual maximum or 92nd percentile for multi-year periods. R^2 is the adjusted R^2 of the initial linear regression. Adjusted R^2 takes into account the number of samples. Bold numbers indicate P < 0.001. Leave-one-out cross- validation tests were run for each period and the NSE statistic is shown. An NSE value of 1 indicates perfect predictive ability. Negative NSE means essentially no predictive ability. Positive NSE values are shown in red. All NSE values for the DRP relationships were high negative values (not reported).

		All sites in dataset			Flow-sensitive sites			Flow-insensitive sites		
Dataset	**Period**	**n**	**R^2**	**NSE**	**n**	**R^2**	**NSE**	**n**	**R^2**	**NSE**
Chlorophyll *a* versus DIN										
Annual	2009-10	43	**0.39**	-0.60	23	0.19	-3.07	11	0.33	-0.96
	2010-11	49	**0.26**	-1.71	25	**0.44**	-0.35	12	0.32	-0.53
	2011-12	50	**0.44**	-0.31	26	**0.70**	0.48	12	0.44	0.03
	2012-13	54	**0.27**	-1.70	27	**0.37**	-0.78	12	0.40	-0.49
	2013-14	59	**0.18**	-3.31	30	0.15	-4.16	14	0.52	0.01
	2014-15	59	**0.18**	-3.44	30	0.20	-2.68	14	**0.62**	0.4
	2015-16	58	**0.22**	-2.47	30	0.20	-3.01	14	0.49	0.07
Three-Y	2009-12	52	**0.46**	-0.23	27	**0.55**	0.06	12	0.51	0.29
	2010-13	56	**0.17**	-3.48	28	**0.56**	0.17	12	0.54	0.21
	2011-14	59	**0.28**	-1.57	30	**0.31**	-1.2	14	0.60	0.25
	2012-15	59	**0.28**	-1.57	30	**0.32**	-1.08	14	**0.64**	0.37
	2013-16	59	**0.23**	-2.24	30	**0.28**	-1.56	14	**0.64**	0.37
Seven-Y	2009-16	61	**0.34**	-1.0	31	**0.42**	-0.43	14	**0.75**	0.63
Chlorophyll *a* versus DRP										
Annual	2009-10	43	0.09		23	0		11	0	
	2010-11	49	0.10		25	0.21		12	0	
	2011-12	50	0.07		26	0.27		12	0	
	2012-13	54	0.03		27	0		12	0	
	2013-14	59	0.05		30	0		14	0.25	
	2014-15	59	0.00		30	0		14	0	
	2015-16	58	0.04		30	0.04		14	0.16	
Three-Y	2009-12	52	0.05		27	0.16		12	0	
	2010-13	56	0.07		28	0.16		12	0	
	2011-14	59	0.06		30	0.10		14	0	
	2012-15	59	0.04		30	0		14	0	
	2013-16	59	0.04		30	0		14	0.07	
Seven-Y	2009-16	61	0.05		31	0.13		14	0	

The above results all used either maximum chlorophyll *a* (annual datasets) or the 92nd percentile of chlorophyll *a* (multi-year datasets) as the dependent variable (i.e., peak chlorophyll *a*). As a comparison, we repeated the analysis for DIN using mean chlorophyll *a* (arithmetic mean, log-transformed) as the dependent variable. That analysis (detailed data not shown) indicated that the relationships between mean chlorophyll *a* and DIN using all of the data or data from flow-sensitive sites were generally weaker than those with maximum or 92nd percentile data. At flow-insensitive sites the relationships were similar, on average, in terms of R^2 (Table 6-2). Finally, we also ran the analyses with weighted composite cover (WCC) as the dependent variable (see Section 9). The outcome was that none of the relationships explained as much variance as those with chlorophyll *a*.

Table 6-2: Comparison of adjusted R^2 values in relationships between chlorophyll *a* and DIN using two chlorophyll *a* metrics. All data were log-transformed. Mean R^2 values are shown.

	R^2 using Maximum or 92nd percentile			R^2 using Mean		
Period	ALL	Flow-sensitive	Flow-insensitive	ALL	Flow-sensitive	Flow-insensitive
Annual (n = 7)	0.28	0.32	0.45	0.16	0.31	0.50
Three-year (n = 5)	0.28	0.41	0.59	0.13	0.34	0.61
Seven-year (n = 1)	0.34	0.42	0.75	0.12	0.37	0.68

6.3 Development of between-site relationships using multiple variables

6.3.1 Introduction

In this part of the analysis, we used a regression approach to develop new relationships analgous to those developed by Biggs (2000). These new relationships were expected to include environmental factors additional to nutrients and a flow variable. To be useful in the context of the periphyton attribute in the National Policy Statement for Freshwater Management (NPS-FM) (NZ Government 2017) and for nutrient targets already set in the One Plan for the Horizons region, any nutrient variables would need to be DIN and DRP, or linked to DIN and DRP.

6.3.2 Data selection and preparation

As described in Section 6.2.1, we used data averaged over annual and three-year periods (seven one-year and five three-year datasets, respectively, with three-year periods overlapping, and a composite dataset comprising data averaged over up to seven years). Years ran from July to June. The dependent variables were either maximum annual chlorophyll *a* (for the annual datasets) or the 92nd percentile of chlorophyll *a* (for multi-year datasets). The number of sites in each dataset varied because more sites were added to the monitoring programme over the years. Refer to Table 2-1 for the start dates at each site. The datasets included all of the variables described in Table 2-2 (i.e., hydrological, water quality (nutrients and conductivity), substrate and catchment variables).

Prior to analysis, we generated Pearson correlation matrices for the candidate predictor variables for each time period and used the results to guide subsequent variable selection for the models. The correlation analysis, and preliminary consideration of the main factors known to influence periphyton growth (i.e., chlorophyll *a* accrual) identified that only a small core of explanatory variables might be required to produce useful relationships (i.e., many explanatory variables were correlated with each other, so that multiple variables could be represented by a single variable). The "core" set of variables that was available at all sites included: DIN, DRP, water conductivity, bed substrate composition, and water temperature. Additional potentially useful nutrient variables included TN, TP, TDP and TSS, which were available at most sites.

Flow variables were available at the subset of 51 sites with a flow record. We used mean accrual period as the main flow-based variable, calculated as:

Accrual period (Da) = (365 – mean annual no. days flow > *n* x median flow) / FRE*n*

where FRE*n* is the mean annual frequency of events exceeding n x median flow, with a 5-day window (i.e., events occurring 5 days or less apart were counted as a single event). This approach of using the average of the accrual period in each year was used by Biggs (2000a), but the calculation in the present analysis was slightly different.

We considered that the slightly revised method of calculating accrual period (compared to Biggs 2000a) was a realistic representation of the actual time available for periphyton accrual when using the accrual period calculated from the effective flow as a predictor variable. Effective flow varies from 1.5 to 15 x median flow, and the proportion of time this flow threshold is exceeded varies considerably across sites (Table 3-1). Therefore, excluding the time under high flows from the calculation of accrual period can make a substantial difference, compared to incuding that time. Refer to Section 3 for the derivation of effective flows at each site. For each site and time period, accrual period at each site was calculated based on the effective flow (Da_EFF) and on 3 x median flow (Da_3med). At sites where an effective flow could not be determined (flow-insensitive sites) only Da_3med was available.

Catchment land cover variables were available as fixed variables at all sites (i.e., constant from year to year) from the LCDB3 database for each upstream catchment. We used three variables that were uncorrelated: percentage of catchment under indigenous forest (%indig_forest), percentage of catchment under high-producing grassland, horticulture and orchards/vinyards (%farm) and percentage of catchment under low-productivity grassland (%lo_grass).

6.3.3 Methods

Following the contract, we used a stepwise multiple regression approach. Models were run using the GLM package in R. Cross-correlations between predictor variables were checked prior to running the models for each period (see Appendix B). Pearson correlation matrices were also generated for the main predictor variables in each multi-year time period and were checked prior to selecting potential predictor variables (see Appendix E). Variance inflation factors (VIF) were calculated for final models to assess collinearity.

The fit of each model was assessed using leave-one-out cross-validation, as described in Section 6.2.1. In summary, once the stepwise procedure has selected a model, the variables included in the selected model are used to generate a series of models omitting one datapoint each time, each of which is used to predict the value of the dependent variable in the omitted datapoint.

In view of issues raised in the literature about the utility and validity of stepwise linear regression (see Appendix B), we also ran, for selected models, a procedure that identifies all the best subsets of models given a selection of predictor variables and ranks them on the basis of a range of model evaluators including R^2, the Akaike Information Criterion (AIC), and Mallows C_p (refer to Geyer 2003 for information on each). The ranking procedure was performed using the "best subsets" routine in

SYSTAT v. 13. The models identified in the stepwise procedure were compared with those identified by best subsets to ensure that no good alternative models were ignored.

Thirty-one analyses were run using the core variables (DIN, DRP, conductivity, water temperature, % coarse, %sand, and either Da_EFF or Da_3med), including on sub-groups (Table 6-3, see below). Models 24 to 31 in Table 6-3 used alternate nutrient variables and catchment variables (as noted above). Each was run on the seven annual datasets, five three-year datasets and the compiled seven-year dataset. In total, the results of 537 model runs were evaluated.

Table 6-3: List of between-site regression models run on the Horizons dataset. Models were run using different combinations of the same set of nine explanatory variables. Max. n is the maximum number of sites available in the datasets for the different annual, three-year and six-year periods.

Model no.	Model abbreviation	Sites included	Flow variable	Temp. incl.	Data gr.	Max. n
Models runs including DIN and DRP to determine best models						
1	All_data_no_flow	All available	None	Yes	All	61
2	All_data_no_flow_no_temp	All available	None		All	61
3	All_data_3med	All with flow	Da_3med	Yes	All	50
4	Flow_only_3med	Effective flow	Da_3med	Yes	All	42
5	Flow_only_EFF	Effective flow	Da_EFF	Yes	All	42
6	Flow_only_3med_no_temp	Effective flow	Da_3med		All	42
7	Flow_only_EFF_no_temp	Effective flow	Da_EFF		All	42
8	Flow_only_no_flow	Effective flow	None	Yes	All	42
9	Flow_only_no_flow_temp	Effective flow	None		All	42
Subsets						
10	All_3med_byFLOW	All with flow	Da_3med	Yes	Flow	31,14
11	Flow_only_EFF_byFLOW	Effective flow	Da_EFF	Yes	Flow	31
12	All_3med_no_temp_byFLOW	All with flow	Da_3med		Flow	31, 14
13	Flow_only_EFF_no_temp_byFLOW	Effective flow	Da_EFF		Flow	31
14	Flow_only_no_flow_noFLOW	Effective flow	None	Yes	Flow	31
15	Flow_only_no_flow_temp_noFLOW	Effective flow	None		Flow	31
16	Flow_only_3med_byNL	Effective flow	Da_3med	Yes	NL	23, 11
17	Flow_only_EFF_byNL	Effective flow	Da_EFF	Yes	NL	23, 11
18	Flow_only_3med_no_temp_byNL	Effective flow	Da_3med		NL	23, 11
19	Flow_only_EFF_no_temp_byNL	Effective flow	Da_EFF		NL	23, 11
20	Flow_only_3med_bySHADE	Effective flow	Da_3med	Yes	Shade	19, 11
21	Flow_only_EFF_bySHADE	Effective flow	Da_EFF	Yes	Shade	19, 11

Model no.	Model abbreviation	Sites included	Flow variable	Temp. incl.	Data gr.	Max. n
22	Flow_only_3med_no_temp_bySHADE	Effective flow	Da_3med		Shade	19, 11
23	Flow_only_EFF_no_temp_bySHADE	Effective flow	Da_EFF		Shade	19, 11
Additional / alternate nutrient variables						
24	All_data_TN_TP_TDP_no_flow	All available	None	Yes	All	61
25	Flow_only_TN_TP_TDP_no_flow	Effective flow	None	Yes	All	42
26	Flow_only_TN_TP_TDP_3_med	Effective flow	Da_3med	Yes	All	42
27	Flow_only_TN_TP_TDP_EFF	Effective flow	Da_EFF	Yes	All	42
Land cover variables						
28	All_data_landcover_DRP_noflow	All available	None	Yes	All	61
29	Flow_only_landcover_DRP_nflow	Effective flow	None	Yes	All	42
30	Flow_only_landcover_DRP_da3	Effective flow	Da_3med	Yes	All	42
31	Flow_only_landcover_DRP_EFF	Effective flow	Da_EFF	Yes	All	42

The choice of models was designed to:

1. compare, using equivalent datasets, the effect on predictability of including or omitting a flow variable;
2. where a flow variable was available for the model, evaluate, using equivalent datasets, the effect of using accrual period based on 3 x median flow compared to accrual period based on effective flow;
3. assess the effect on predictability of subdividing sites on the basis of their physico-chemical characteristics;
4. assess the effect on predictability of including nutrient variables other than those thought to be directly taken up by periphyton (i.e., including TN, TP, TDP);
5. determine whether including catchment land cover variables improved predictive performance.

We ran all of the models across all available sites, except that, when comparing the effect of using different flow metrics (item 2 above), we restricted the dataset to the sites at which an effective flow had been identified. This allowed a direct comparison of the performances of models run on exactly the same dataset, with changes to one predictor variable.

Subdivision of data (item 5) was based on flow group (as defined in Table 3-1) and nutrient limitation group (as defined in Section 4.3.3). We also subdivided sites into shaded or unshaded sites. This assessment used observations at the sites as documented by Kilroy et al. (2012). Subdivision of the dataset produced small groups, which means that there is lower confidence that the data are a true representation of all sites with those characteristics. Even with this limitation, the comparisons of model performance (especially between 3 x median flow and effective flow) may still be informative.

Quantile regressions for the 10^{th}, 20^{th}, 30^{th}, 40^{th}, 50^{th}, 60^{th}, 70^{th}, 80^{th} and 90^{th} percentiles were run on the best model identified using model runs 1 to 9 in Table 6-3 for each time period dataset using:

a) the predictor variables retained in the relevant stepped linear regression;

b) all predictor variables entered into the relevant stepped linear regression.

See Appendix B for results from quantile regressions and an explanation of why results are not included in this section.

6.3.4 Results

Correlations among variables

Pearson correlation matrices for the main predictor variables are shown in Appendix E. In some years, water temperature was moderately correlated with DIN and with % coarse sediment. We therefore ran models both including and excluding water temperature. Across sites, TN was highly correlated with DIN, TP with DRP, and TDP with DRP. TN, TP and TDP were therefore excluded from the main analysis. Additional model runs were performed to assess the effect on the results of including TN, TP and TDP instead of DIN and DRP. VIF for parameters in the final models indicated low collinearity (all <2.5; Table 7-2).

Model selection

The stepwise procedure identified models identical or close to those ranked as the best models in the best subsets procedures. Model ranking varied slightly depending on the ranking criterion used (R^2, C_p, AIC) but all differences were in variables that explained only a small proportion of the variance in chlorophyll *a*.

Best models for predicting across all data

All models derived from all of the data (i.e., the maximum number of sites having data on the variables included for selection) included DIN and conductivity, and models in 12 of the 13 periods included Da_EFF (Table 6-4). Other variables featured in some periods, explaining a minor proportion of the variance in chlorophyll *a*. Except for 2009-10, all of the models were generated from model runs 5 or 7 (see Table 6-3).

Except for data from 2009-10, 2010-11 and 2009-12, all of the models performed strongly in the cross-validation tests (e.g., cross-validated observed vs. predicted $R^2 > 0.5$). RMSD ranged from 9% to 17% of the range of $\log_{10}$chlorophyll *a* (annual maximum or multi-year 92^{nd} percentile) and R^2 in the observed vs. predicted relationships ranged from 0.53 to 0.75. Plots of observed vs. predicted confirmed close correspondence (Figure 6-3). There was some bias in many models (i.e., the slope of the best fit line differed significantly from the 1:1 line). However, the discrepancies were small (Figure 6-3).

In general the models generated from the multi-year datasets were stronger than those from the annual datasets. In other words, annual maxima appeared to be more difficult to predict than the multi-year 92^{nd} percentile). This was especially the case in the later years of the programme (2011 onwards).

Table 6-4: Summary of the best models identified across the maximum number of sites in each period tested. The models predicted $\log_{10}$maximum chlorophyll *a* in each of seven annual datasets, and $\log_{10}92^{nd}$ percentile chlorophyll *a* in each of five overlapping three-year datasets and one seven-year dataset. For model numbers, see Table 6-3. Nutrient variables were DIN and DRP. The adjusted R^2 (R^2) and model degrees of freedom (df) of the initial regressions are shown (df = number of predictor variables + 1). All models were statistically significant (P <0.00001). Cross validation statistics are R^2 (of the regression of observed vs. predicted), Nash-Sutcliffe efficiency (NSE) and root-mean squared deviation (RMSD). Variables included are shown on the right. Variance Inflation Factor (VIF) values for each variable in the model are in brackets after the model name. VIF <5 is generally accepted as lack of collinearity.

				Regression		Cross-validation			
Model	**Model_name**	**Period**	***n***	**R^2**	**df**	**R^2**	**NSE**	**RMSD**	**Variables included (VIF)**
Annual periods									
2	All_data_no_flow_no_temp	2009-10	43	0.50	3	0.46	0	0.41	DIN (1.4), cond (1.4)
7	Flow_only_EFF_no_temp	2010-11	33	0.51	4	0.43	0.02	0.45	DIN (1.2), cond (1.1), daEFF (1.1)
5	Flow_only_EFF	2011-12	35	0.74	6	0.65	0.57	0.34	DIN (1.5), cond (1.3), daEFF (1.6), %coarse (1.9), %sand (1.2)
5	Flow_only_EFF	2012-13	37	0.67	7	0.56	0.38	0.30	DIN (1.8), DRP (1.6), cond (1.3), daEFF (1.5), %coarse (1.4), temp (1.8)
7	Flow_only_EFF_no_temp	2013-14	40	0.64	4	0.59	0.41	0.32	DIN (1.2), cond (1.1), daEFF (1.2)
5	Flow_only_EFF	2014-15	41	0.66	5	0.61	0.45	0.35	DIN (1.1), cond (1.1), daEFF (1.3), %coarse (1.2)
7	Flow_only_EFF_no_temp	2015-16	39	0.63	4	0.55	0.34	0.30	DIN (1.2), cond (1.1), daEFF (1.1)
Three-year periods									
5	Flow_only_EFF	2009-12	36	0.61	5	0.45	0.2	0.42	DIN (2.1), cond (1.1), daEFF (1.2), temp (2.0)
7	Flow_only_EFF_no_temp	2010-13	37	0.64	5	0.53	0.32	0.34	DIN (1.3), cond (1.1), daEFF (1.4), %coarse (1.3)
5	Flow_only_EFF	2011-14	40	0.74	5	0.71	0.63	0.28	DIN (2.0), cond (1.1), daEFF (1.3), temp (2)
5	Flow_only_EFF	2012-15	40	0.78	6	0.75	0.69	0.24	DIN (1.8), DRP (1.3), cond (1.3), daEFF (1.2), temp (1.8)
7	Flow_only_EFF_no_temp	2013-16	40	0.76	5	0.72	0.65	0.25	DIN (1.2), cond (1.1), daEFF (1.3), %coarse (1.2)
Seven-year period									
7	Flow_only_EFF_no_temp	2009-16	42	0.73	5	0.67	0.56	0.28	DIN (1.4), cond (1.1), daEFF (1.4), %coarse (1.3)

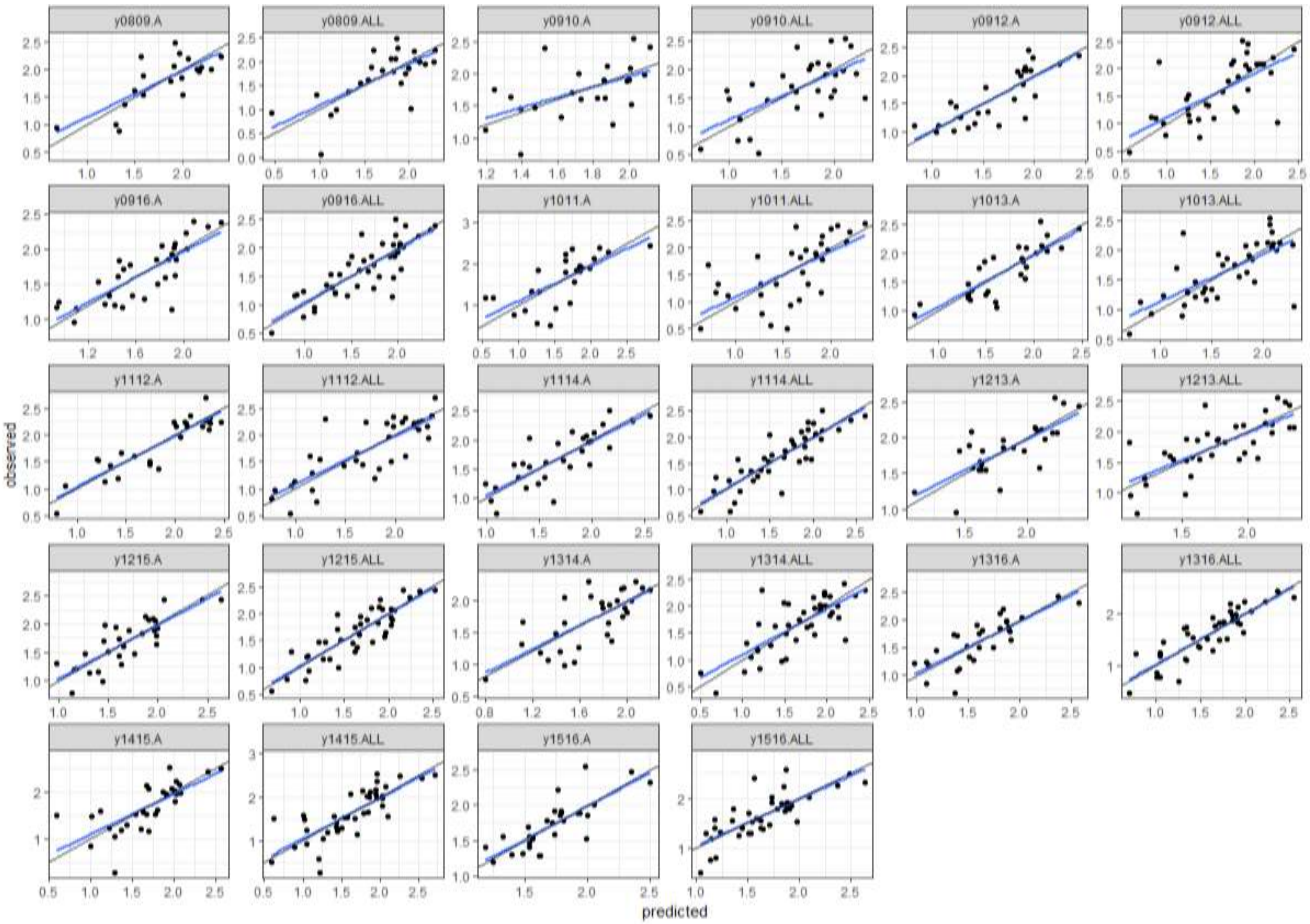


Figure 6-3: Plots of observed versus predicted log$_{10}$chlorophyll *a* generated from the leave-one-out cross validation procedure for model 7 in Table 6-3. Blue lines are best linear fit through the data; black lines are 1: 1. Plots are shown for all annual and three-year periods and the seven-year period 2009-16. Maximum annual chlorophyll *a* was predicted for the annual periods and the 92nd percentile of chlorophyll *a* for multi-year periods. A refers to the group of sites classed as flow-sensitive. ALL refers to all of the available data.

Comparison of performance using accrual period based on effective flow vs. 3 x median flow vs. omitting a flow variable

Accrual period based on 3 x median flow (Da_3med) did not feature in any of the best models. We compared equivalent models in which either Da_3med or Da_EFF was the only available flow variable, to determine the difference in performance. In all periods, the cross-validation statistics indicated more successful models using Da_EFF rather than equivalent models (compare Table 6-4 with **Error! Not a valid bookmark self-reference.** a). Omitting the flow variable led to a further reduction in performance (**Error! Not a valid bookmark self-reference.**) with mean values of the regression R^2, cross-validation R^2 and NSE declining and RMSD slightly increasing (Table 6-6).

Table 6-5: Summary of the best models identified in each period, with (a) Da_3med as the available flow variable and (b) no flow variable. For model numbers, see Table 6-3. Nutrient variables were DIN and DRP. Refer to Section 6.2.1 and Table 6-4 for explanations of statistics.

	(a) Models including Da_3med							(b) Models with no flow variables						
			Regression		Cross-validation					Regression		Cross-validation		
Period	Model	*n*	R^2	df	R^2	NSE	RMSD	Model	*n*	R^2	df	R^2	NSE	RMSD
Annual periods														
2009-10	4	30	0.45	3	0.39	-0.23	0.42	1	43	0.50	3	0.46	0.00	0.41
2010-11	4	33	0.47	3	0.41	-0.12	0.45	8	33	0.47	3	0.41	-0.12	0.45
2011-12	4	34	0.70	5	0.53	0.40	0.40	8	34	0.68	6	0.55	0.40	0.38
2012-13	4	36	0.65	7	0.57	0.36	0.28	8	36	0.56	6	0.46	0.12	0.32
2013-14	4	40	0.60	7	0.42	0.16	0.40	1	59	0.49	7	0.40	-0.03	0.37
2014-15	4	40	0.51	4	0.46	0.05	0.42	8	40	0.48	3	0.43	-0.05	0.43
2015-16	4	39	0.64	6	0.49	0.25	0.33	8	39	0.58	3	0.51	0.22	0.32
Three-year periods														
2009-12	4	36	0.61	6	0.44	0.21	0.42	8	36	0.56	4	0.47	0.14	0.40
2010-13	4	37	0.61	5	0.52	0.27	0.34	1	55	0.50	3	0.45	-0.03	0.35
2011-14	4	40	0.64	4	0.58	0.39	0.33	8	40	0.56	4	0.50	0.19	0.36
2012-15	4	40	0.66	4	0.6	0.43	0.31	8	40	0.60	4	0.55	0.31	0.33
2013-16	4	40	0.65	4	0.59	0.41	0.31	8	40	0.60	4	0.55	0.30	0.32
Seven-year period														
2009-16	4	42	0.64	4	0.54	0.34	0.33	8	42	0.60	4	0.54	0.28	0.33

Table 6-6: Mean statistics across models in all time periods for groups of models including Da_Eff, Da_3med and no flow variable, and across data subsets. For model numbers, see Table 6-3. Nutrient variables were DIN and DRP. df = model degrees of freedom (number of predictor variables + 1).

		Regression		Cross-validation		
Model group	n	R^2	range of df	R^2	NSE	RMSD
All available data with different flow variables						
Models with Da_EFF as flow variable	39	0.66	3 – 7	0.59	0.40	0.33
Models with Da_3med as flow variable	37	0.60	3 – 7	0.50	0.22	0.36
Models with no available flow variable	41	0.55	3 – 7	0.48	0.13	0.37
Best models across sub-groups of sites						
All models, flow-sensitive sites	28	0.67	4 – 7	0.60	0.43	0.29
All models, flow-insensitive sites	13	0.84	3 – 8	0.52	0.44	0.53
All models, co-limited sites	22	0.65	4 – 7	0.53	0.31	0.35
All models, P-limited sites	10	0.91	3 – 8	0.79	0.77	0.18
All models, unshaded sites	19	0.69	2 – 6	0.58	0.41	0.33
All models, shaded sites	10	0.88	5 – 7	0.72	0.70	0.29

Performance of models on data subsets

The purpose of running the models on data subsets was to determine whether smaller groups of sites showed stronger relationships with the environmental variables than all of the available sites. Stronger relationships could occur if the subsets represented additional physico-chemical factors that influence periphyton biomass. For example, heavy shade may cause periphyton to be lower than its potential maximum at an unshaded site.

The largest subset comprised sites at which periphyton was classed as sensitive to flow (Table 3-1). Model performance at these sites was generally no better than across all sites. On average the models included more variables than those derived across all available sites (mean of 4.4 compared to 3.9) (n = df – 1, Table 6-6).

Flow-insensitive sites made up a group of 13 sites, on average. Initial regression relationships were strong across these sites (mean R^2 of 0.84). However, model performance was no better than for the larger group of flow-sensitive sites (Table 6-6), suggesting that a small number of sites drove the strong regressions.

At a group of 22 sites periphyton chlorophyll *a* was classed as co-limited by both DIN and DRP. On average predictive models for maximum chlorophyll *a* across these sites were weak. A smaller group of P-limited sites (n = 10) had strong models. Small *n* compared to the number of variables in the models restricts model reliability.

Across 19 unshaded sites model strength was similar to that across all sites. In general, the variables included were similar (DIN, conductivity and Da_EFF in most periods). The models for the small groups of shaded sites (n = 10) were stronger, but again reliability was limited by high numbers of predictor variables (4.8 averaged across all the periods compared to an average of 3.6 variables for the larger group of unshaded sites (Table 6-6)).

Effect of including alternative nutrient variables (TN, TP, TDP)

We ran four models in which DIN and DRP were replaced by TN, TP or TDP (models 24 to 27 in Table 6-3). TP and TDP were strongly correlated in most periods (Appendix E). Therefore, either TP or TDP was included, not both. Predictive ability was compared between models when no flow variable, Da_3med or Da_EFF were available in the suite of predictors, using only data from sites at which an effective flow had been identified. We also looked at the performance of models run across all available data, with no flow variable included. For brevity, we report on the three-year and seven-year periods only. The summary results in Table 6-7 provide a comparison of four models run in each three-year period and the seven-year period, the best model of which can then be compared with the best model for the same period, which are shown in Table 6-4. The differences were small, except in 2010-13 when NSE using DIN was 0.32 and NSE using TN was 0.77 (although the TN model included one extra term, temp) (Table 6-7). All models included both TN and conductivity.

Effect of including land cover variables (%farm, %indig_for, %grass_lo)

The land cover variable %farm was strongly correlated with TN and moderately strongly correlated with DIN. The variable %indig_for was moderately negatively correlated with TP in some years, but not with DRP (Appendix E). In the final set of model runs we therefore included the three uncorrelated landcover variables (%farm, %indig_for and %grass_lo) and retained DRP, but dropped TN and DIN. We ran four models, equivalent to the four runs for TN, TP and TDP (see above).

All models with land cover variables were weak (Table 6-8). Maximum NSE of 0.56 was for the model including Da_EFF as the flow variable, in 2011-14, and no landcover variables was selected by the stepwise procedure in this model (Table 6-8). Three models (that did not include a flow variable) had negative NSE, indicating no predictive ability. Eighteen of the 24 models included Farm as a predictor. Only one did not include conductivity.

Table 6-7: Summary results of model runs in which DIN and DRP were replaced by TN and TP or TDP. Model numbers as in Table 6-3. Refer to Section 6.2.1 and Table 6-4 for explanations of statistics.

			Regression		Cross validation			
Model	**Model_name**	**n**	**R^2**	**df**	**R^2**	**NSE**	**RMSD**	**Variables included**
Three years, 2009 - 12								
24	All_data_TN_TP_TDP_no_flow	38	0.69	5	0.65	0.53	0.32	TN, cond, %coarse, temp
25	Flow_only_TN_TP_TDP_no_flow	30	0.68	5	0.64	0.52	0.34	TN, cond, %coarse, temp
26	Flow_only_TN_TP_TDP_3_med	30	0.68	5	0.64	0.52	0.34	TN, cond, %coarse, temp
27	Flow_only_TN_TP_TDP_EFF	30	0.68	4	0.65	0.51	0.33	TN, cond, Da_EFF
Three years, 2010_13								
24	All_data_TN_TP_TDP_no_flow	51	0.72	4	0.70	0.59	0.26	TN, cond, %coarse
25	Flow_only_TN_TP_TDP_no_flow	33	0.81	6	0.77	0.74	0.24	TN, TDP, cond, %coarse, temp
26	Flow_only_TN_TP_TDP_3_med	33	0.81	6	0.77	0.74	0.24	TN, TDP, cond, %coarse, temp
27	Flow_only_TN_TP_TDP_EFF	33	0.83	6	0.80	0.77	0.22	TN, cond, Da_EFF, %coarse, temp
Three years, 2011_14								
24	All_data_TN_TP_TDP_no_flow	56	0.69	4	0.66	0.53	0.28	TN, cond, %coarse
25	Flow_only_TN_TP_TDP_no_flow	37	0.75	5	0.72	0.65	0.27	TN, TP, cond, %coarse
26	Flow_only_TN_TP_TDP_3_med	37	0.75	5	0.72	0.65	0.27	TN, TP, cond, %coarse
27	Flow_only_TN_TP_TDP_EFF	37	0.78	4	0.76	0.70	0.25	TN, cond, Da_EFF
Three years, 2012_15								
24	All_data_TN_TP_TDP_no_flow	56	0.71	4	0.69	0.58	0.26	TN, cond, %coarse
25	Flow_only_TN_TP_TDP_no_flow	37	0.75	5	0.72	0.65	0.27	TN, TP, cond, %coarse
26	Flow_only_TN_TP_TDP_3_med	37	0.75	5	0.72	0.65	0.27	TN, TP, cond, %coarse
27	Flow_only_TN_TP_TDP_EFF	37	0.80	4	0.78	0.73	0.23	TN, cond, Da_EFF
Three years, 2013_16								
24	All_data_TN_TP_TDP_no_flow	56	0.68	5	0.65	0.51	0.26	TN, cond, %coarse, %sand
25	Flow_only_TN_TP_TDP_no_flow	38	0.73	5	0.69	0.60	0.27	TN, cond, %coarse, %sand
26	Flow_only_TN_TP_TDP_3_med	38	0.73	5	0.69	0.60	0.27	TN, cond, %coarse, %sand
27	Flow_only_TN_TP_TDP_EFF	38	0.78	5	0.74	0.68	0.25	TN, cond, %sand, Da_EFF
Seven years, 2009_16								
24	All_data_TN_TP_TDP_no_flow	58	0.74	4	0.72	0.64	0.24	TN, cond, %coarse
25	Flow_only_TN_TP_TDP_no_flow	39	0.80	5	0.78	0.74	0.23	TN, TP, cond, %coarse
26	Flow_only_TN_TP_TDP_3_med	39	0.80	5	0.78	0.74	0.23	TN, TP, cond, %coarse
27	Flow_only_TN_TP_TDP_EFF	39	0.81	4	0.79	0.74	0.22	TN, cond, Da_EFF

Table 6-8: Summary results of model runs in which DIN was replaced by land cover variables (Farm, Indig_for, Grass). DRP was uncorrelated with the land cover variables and was retained in the model runs. Model numbers as in Table 6-3. Refer to Section 6.2.1 and Table 6-4 for explanations of statistics.

			Regression		Cross validation			
Model	Model_name	n	R^2	df	R^2	NSE	RMSD	Variables included
Three years, 2009 - 12								
28	All_data_landcover_DRP_noflow	52	0.45	3	0.40	-0.28	0.41	Farm, cond
29	Flow_only_landcover_DRP_nflow	36	0.57	3	0.51	0.19	0.38	Farm, cond
30	Flow_only_landcover_DRP_da3	36	0.67	4	0.59	0.44	0.35	Farm, indig_for, Da_3med
31	Flow_only_landcover_DRP_EFF	36	0.63	4	0.49	0.26	0.39	Farm, cond, Da_EFF
Three years, 2010_13								
28	All_data_landcover_DRP_noflow	55	0.43	4	0.38	-0.34	0.37	Farm, grass, cond
29	Flow_only_landcover_DRP_nflow	37	0.52	6	0.37	-0.06	0.40	DRP, cond, %coarse, %sand, temp
30	Flow_only_landcover_DRP_da3	37	0.66	7	0.52	0.33	0.35	Farm, indig_for, cond, Da_3med, %coarse, temp
31	Flow_only_landcover_DRP_EFF	37	0.63	5	0.45	0.21	0.37	Farm, cond, Da_EFF, temp
Three years, 2011_14								
28	All_data_landcover_DRP_noflow	59	0.51	6	0.45	0.01	0.36	Farm, indig for, grass, cond, %coarse
29	Flow_only_landcover_DRP_nflow	40	0.57	3	0.53	0.21	0.35	Cond, temp
30	Flow_only_landcover_DRP_da3	40	0.61	4	0.55	0.28	0.34	Farm, cond, Da_3_med
31	Flow_only_landcover_DRP_EFF	40	0.71	4	0.68	0.56	0.29	Cond, Da_EFF, temp
Three years, 2012_15								
28	All_data_landcover_DRP_noflow	59	0.51	4	0.48	0.01	0.33	Farm, grass, cond
29	Flow_only_landcover_DRP_nflow	40	0.55	3	0.51	0.14	0.34	Cond, temp
30	Flow_only_landcover_DRP_da3	40	0.61	5	0.54	0.30	0.33	Farm, indig_for, cond, Da_3med
31	Flow_only_landcover_DRP_EFF	40	0.68	4	0.65	0.49	0.29	Cond, Da_EFF, temp
Three years, 2013_16								
28	All_data_landcover_DRP_noflow	59	0.51	5	0.46	0.03	0.33	Farm, cond, %coarse, %sand
29	Flow_only_landcover_DRP_nflow	40	0.55	3	0.49	0.12	0.34	Farm, cond
30	Flow_only_landcover_DRP_da3	40	0.61	5	0.55	0.31	0.32	Farm, indig_for, cond, Da_3med
31	Flow_only_landcover_DRP_EFF	40	0.67	5	0.62	0.46	0.30	Cond, %sand, Da_EFF, temp
Seven years, 2009_16								
28	All_data_landcover_DRP_noflow	61	0.54	4	0.50	0.12	0.32	Farm, cond, grass
29	Flow_only_landcover_DRP_nflow	42	0.58	3	0.52	0.21	0.34	Farm, cond
30	Flow_only_landcover_DRP_da3	42	0.66	5	0.57	0.39	0.32	Farm, indig_for, cond, Da_3med
31	Flow_only_landcover_DRP_EFF	42	0.68	6	0.61	0.45	0.31	Farm, indig_for, cond, %coarse, Da_EFF

6.4 Discussion

6.4.1 Statistical approach

Biggs (2000a) used simple regression techniques and was able to summarise the relationships in regression equations that were easy to understand and apply in a management context. Use of the Horizons seven-year dataset has confirmed that the regression approach used by Biggs (2000a) can still be relevant despite availability of new modelling approaches capable of handling complex datasets. Simple linear regression has been assessed as having advantages over more complex modelling techniques in other fields (Aertsen et al. 2010). In future analyses, random forests (Prasad et al. 2006) could be an alternative to linear regressions. Random forests include non-linear relationships and interactions, and can predict to new sites and conditions. A prerequisite is a "reasonably large" dataset, and the approximately 60 sites in the Horizons dataset may be sufficient. Nevertheless, the straightforward multiple linear regression approach above yielded strong results, when a range of environmental factors (additional to those used by Biggs 2000) were taken into account. Regression relationships have the advantage for policy makers and resource managers that predictions for any combination of conditions can be made easily.

6.4.2 Relationships between chlorophyll *a* and DIN or DRP across sites

In general, nutrient concentrations in isolation from other variables were not good predictors of peak chlorophyll *a* at sites across the Horizons region. This finding is consistent with the results of spatial analyses of periphyton (usually as chlorophyll *a*) in other regions and countries (e.g., Niyogi et al. 2007, Lewis and McCutchan 2010). DRP concentrations were largely unrelated to periphyton biomass despite many sites being P-limited according to classical nutrient limitation theory (see Section 4). DIN concentrations were correlated with chlorophyll *a* but with explanatory power too low to be useful. Lack of association does not imply that DIN or DRP are not affecting periphyton in the region. The weak relationships more likely indicate that peak chlorophyll *a* is simultaneously affected by other factors, or that DIN and DRP supply are not adequately represented by concentrations of DIN or DRP in the water column. For example, for DRP in particular, recycling processes within sediments allow periphyton to access P from bound forms that are not included in DRP (Dodds 2003b).

The exception to this result was that across the small group of sites defined as flow-insensitive (groups C and D combined in Table 3-1). DIN explained relatively high proportions of the between-site variance in the 92nd percentile of chlorophyll *a* in the multi-year periods. As noted in Section 5.3, a strong relationship with DIN explained the better performance of the Biggs (2000a) relationships for predicting maximum chlorophyll *a* across this group of sites.

Characteristics of the 14 flow-insensitive sites (groups C and D in Table 3-1) are summarised in Table 6-9. The flow insensitive sites tended to have smaller flows and, on average, higher proportions of coarse bed material than the flow-sensitive sites. The flow-insensitive sites also tended to have low frequencies of effective floods (seven or fewer events per year, on average). However, there was overlap between the two groups and no one factor or set of factors separated the two groups clearly (see Figure 3-2). Nevertheless, long accrual periods at these sites may ensure that periphyton consistently attains maximum values, as determined by the carrying capacity of the site (Biggs et al. 1998, Biggs and Stockseth 1996). Carrying capacity is likely to be determined, at least in part, by nutrient availability, leading to stronger DIN – chlorophyll *a* relationships between sites, for sites where flow does not exert predictable control over variability in chlorophyll *a* over time.

Table 6-9: Characteristics of sites classified as flow-insensitive in the analysis of chlorophyll *a* versus accrual times. Means calculated over the seven-year dataset. LSC class is the Horizons designated "life supporting capacity" code designated for the site. NL is nutrient limitation status from Table 4-2. Chlorophyll *a* (Chla) is the 92nd percentile, DIN and DRP are geometric means. % coarse is the mean percentage of observed bedrock + boulders + large cobbles.

HRC n	Site abbreviation	LSC class	Mean flow (m_3/s)	NL	Chla (mg/m^2)	DIN (mg/m^3)	DRP (mg/m^3)	% coarse	% sand	Cond. (µS/cm)
1	makakahi_doc	HM	6.3	co	5	27	6.1	52	1	56
2	mangatainoka_putara	UHS	5.0	co	2	13	4.4	52	3	50
9	manawatu_weber	HM	14.0	co-none	162	202	16.3	28	2	269
10	makakahi_ham	HM	6.3	co-P	117	291	5.8	35	2	106
18	mangatainoka_pahiatua	HM	18.0	P	135	889	6.1	36	2	108
26	mangapapa_troup	HM	0.7	co-P	30	213	12.3	16	8	122
37	tokomaru_hb	LM	2.2	co	32	49	6.0	44	5	78
45	mangawhero_doc	UVA	4.7	N	11	10	14.1	62	3	61
46	makotuku_sh49	UVA	0.9	co-N	34	190	9.3	39	11	77
47	mangawhero_us_oha	UVA	4.7	co	49	146	15.1	44	5	86
48	mangawhero_ds_oha	UVA	4.7	N-co	70	174	19.8	50	5	92
49	makotuku_rae	UVA	1.7	co-P	96	284	7.1	54	4	92
57	makotuku_us_rae	UVA	1.7	P-co	132	304	8.8	37	4	98
60	ohau_gladstone	UHS	6.5	co	7	39	8.1	36	7	69

6.4.3 Relationships between chlorophyll *a* and multiple variables

The outcome of the analysis with multiple predictor variables was that we identified consistent and relatively strong relationships between peak chlorophyll *a* and environmental variables calculated across a range of time periods. Links between periphyton abundance and nutrient concentrations across multiple catchments and within catchments have been identified in the past, especially along river continua (Biggs and Gerbeaux 1993, Harding et al. 1999, Klose et al. 2012, Suplee et al. 2012). However, it has rarely been possible to derive relationships robust enough to allow prediction of periphyton at unmonitored sites, or prediction of the consequences for periphyton of altering nutrient supply or flow regime at monitored sites.

Possibly the most successful attempt to date to provide predictive between-site relationships for maximum chlorophyll *a* in rivers prior to this analysis was that of Biggs (2000a), who used chlorophyll *a* and environmental data averaged over time, and also included a flow variable as a predictor. Other attempts have been less successful (e.g., Munn at al. 2002, Lewis and McCutchan 2010, Liess et al. 2012), and have acknowledged that poor relationships can be expected because the environmental controllers of periphyton biomass are so variable.

Development of between-site relationships involves a space-for-time assumption that the environmental factors that lead to peak biomass are the same across sites. The main drivers of algal growth rates and biomass accrual and loss are well known (i.e., temperature and light, the major nutrients DIN and DRP, accrual time, and losses through the effects of high flows and invertebrate grazing, Biggs 2000a), and probably apply across most sites. However, the effects of each potential driver variable may be site-specific and dependent on variables that we do not or cannot measure. In that case, a time-series based (within-site approach) could be informative (see Section 7).

One of the main issues preventing development of robust empirically-based models for predicting periphyton using a space-for-time approach has been lack of datasets long enough to calculate accurate metrics that (a) average out seasonal and flow-driven variability in both the response variable (periphyton abundance, in this case as chlorophyll *a*) and the predictor variables (mainly stream physiochemistry), and (b) take account of variability in river flows. Biggs (2000a) accounted for flow variability using accrual period calculated from the annual frequency of events greater than 3 x median flow, applied to all sites in the dataset. Such use of a flow metric in periphyton – environment relationships is still unusual (Dodds 2007).

In this analysis, three environmental variables explained most of the variability in peak chlorophyll *a* between sites: conductivity, DIN and accrual period based on effective flow, with that order of explanatory power.[1] TN could be substituted for DIN and generally improved predictive power. A variable representing bed sediment composition or water temperature was also included in some relationships. DRP featured in only two of the strongest relationships (for the periods 2012–13, 2012–15) and this may be useful for evaluating the effect of different levels of DRP. Space-for-time vs. time series is discussed further in Section 7.

6.4.4 TN versus DIN as a predictor variable

Models in which nitrogen was represented by TN generally predicted chlorophyll *a* more accurately than models including DIN. Total nitrogen and also TP are routinely measured in streams in North America and are most often used to develop nutrient criteria (Dodds and Welch 2000, Dodds 2007, Stevenson et al. 2008). Dodds (2003a) argued strongly that DIN and DRP should not be used to develop algae – nutrient relationships or to assess nutrient limitation potential, and that TN and TP were always more appropriate. The view of Dodds (2003a) was that DIN and DRP cannot represent available N or P supplies because continual uptake affects concentrations. Furthermore, there are well-known difficulties in analysis of phosphate concentrations using standard methods (Jarvie et al. 2002).

TN and TP have long been the accepted measures for assessing eutrophication in still-water bodies (lakes and large, slow-moving rivers) (Carlson 1977). The use of TN and TP to assess productivity in lakes is logical because productivity in lakes comprises phytoplankton growth in the water column. In a static or slow-moving system, dissolved inorganic N or P are rapidly taken up into biomass in the water column when conditions are suitable. Therefore, the total mass of N or P (both dissolved and assimilated) reflects N and P content at a given time, whereas concentrations of dissolved N and P simply reflect what is left over after assimilation.

In most New Zealand rivers, most algal biomass is restricted to the benthos. At times of very prolific periphyton biomass, the periphyton itself may contribute to TN and TP through sloughing of cells into the water column. This could drive stronger relationships between TN or TP and benthic chlorophyll *a*, but does not provide information about potential drivers of high chlorophyll *a*. To measure stream productivity in the same way as in lakes, measures of TN and TP should <u>include</u> periphyton. This clearly presents difficulties for sample collection, and tells us nothing about the relationships between periphyton and nutrients.

[1] The order of explanatory power was determined by checking differences in R^2 after adding and removing variables (obtained from the best subsets analyses). The strongest predictors make the largest differences.

One way to circumvent the issue of depletion from uptake confounding relationships between chlorophyll *a* and dissolved nutrients is to restrict predictions to mean or maximum chlorophyll *a* over at least a year, and use mean DIN and DRP calculated over the same period to represent average N or P availability at a site. This was suggested by Biggs and Close (1989), applied successfully to develop models (Biggs 2000a), and used in this analysis.

The rapid flux of N and P through river systems means that, often, the proportion of DIN and DRP taken up by benthic algae is very low, especially when DIN is elevated (Mulholland et al. 2008). However, uptake of DIN can be measurable along a reach during summer low flows if there are no inputs from tributaries, groundwater or diffuse sources (e.g., Peterson et al. 2001). As Dodds (2003a) pointed out, measured DIN in the water column under those circumstances will reflect what is left over after uptake, as it does in lakes (see above). Some rivers in the Horizons dataset probably experience measurable declines in DIN through instream uptake during low flows when background DIN concentrations were also low and an example is described for the Rangitikei River in Section 7.4.

6.4.5 The role of conductivity

Conductivity was included as a predictor in all of the strongest regression relationships, with a positive correlation with chlorophyll *a*. Water conductivity is mostly a function of the geological setting of a catchment rather than catchment activities, although substantial changes in nutrient concentrations along a river continuum, or over time, may be detectable as changes in conductivity (Kim and Furumai 2013, Ballantine and Davies-Colley 2014). Conductivity may also be inversely correlated with distance from the sea because of increasing sodium deposition near to the coast. In the Manawatu-Whanganui region, volcanic geology of the northern part of the region leads to high sodium and calcium concentrations and therefore high conductivity in some headwater streams (Goldsmith et al. 2008).

Numerous studies have demonstrated links between conductivity and periphyton community composition (e.g., Potapova and Charles 2003; Vilches et al. 2013; Rott and Schneider 2014). In contrast there appear to be few reports in the literature linking conductivity to periphyton biomass at the scale of landscapes. Yet this pattern in the periphyton was strongly detected in the earliest broad-scale investigation into periphyton in New Zealand rivers, in 1983 (Biggs and Price 1987), and in a separate more detailed study a few years later (Biggs 1990, and see discussion in Biggs 2000a). In both NZ studies, conductivity appeared to be linked to species composition, with taxa forming the highest biomass occurring at higher conductivities. Conductivity data over time carries information about hydrology as well as catchment geology, because there is usually a strong negative relationship between the concentration of ions in the water and flow magnitude because of dilution during rainfall events. However, the 1983 survey was conducted during low flows; therefore, the conductivity – biomass relationship was unlikely to have been confounded by the effect of flows.

Taking conductivity into account in managing rivers is complex. Although there may a cause-effect relationship between conductivity and periphyton chlorophyll *a* (through the influence of ion chemistry on community composition), there is also a direct relationship between conductivity and DIN. A relationship with DIN is expected because when NO^{3-} leaches from soils, it carries cations with it, especially Ca^{2+} and Mg^{2+} (Likens et al. 1970). In the Likens et al. (1970) study, a 56-fold increase in NO^{3-} (from ~200 to 11000 mg N m^{-3}) corresponded to a 3- to 8- fold increase in conductivity (from 20 to 65 – 160 µS cm^{-1}). The relative effect of such leaching would vary depending on background conductivity but in general the effect of increasing DIN on conductivity may start to become noticeable at DIN concentrations > 1000 mg/m^3. A starting value of 1000 mg/m^3 is suggested because an increase of this concentration of DIN is expected to cause conductivity to increase by at least 10 µS/cm, which would be picked up by most conductivity meters. Added to pristine waters[2] such a change would be substantial. The expected increase is based on the changes noted by Likens et al. (1970), which suggested that an increase of 1000 mg/m^3 DIN was equivalent to an increase in conductivity of about 12 µS/cm. The Likens et al. (1970) value was confirmed in nutrient enrichment experiments carried out in streamside channels in Canterbury, in which background conducitivity of 74 µS/cm increased to 92 µS/cm after enrichment with nitrate-N from 55 to 1400 mg/m^3 (NIWA unpublished data).

An analysis of conductivity vs. DIN relationships in the Horizons dataset suggests that conductivity cannot be used as a surrogate for enrichment by DIN, either between sites or within most sites. Between-site relationships are summarised in (Table 6-10), which shows that the relationship between chlorophyll *a* and conductivity was stronger than that with DIN, across all sites and within catchments. Conductivity was weakly related to DIN across all sites and was unrelated in the Whangaehu catchment. Plots of within-site relationships between DIN and conductivity (see Appendix F) indicate that DIN and conductivity are positively correlated at only a few sites (e.g., makotuku_sh49, oroua_awahuri, oroua_ds_fei) and at most sites there is a negative relationship or no relationship.

High conductivity associated with calcium-dominated geology in a catchment may also affect periphyton through changing the availability of phosphorus. Phosphorus forms complexes with calcium carbonate, becoming biologically unavailable (Withers and Jarvie 2008). Removal of available phosphorus from the water column into calcium carbonate complexes could potentially drive systems into P-limitation (Corman et al., 2016). On the other hand, in conditions of anoxia or high pH sediment-bound P within algal mats can be reduced to an available form (Dodds 2003b, Wood et al. 2015). Furthermore, in P-limiting conditions most algae release phosphatases that facilitate the release of available phosphate from organic molecules (Ellison and Brett 2006). Overall, it is unclear whether calcium enhances or restricts the supply of P to algae. In the Horizons dataset, conductivity was more strongly correlated with TP than with either TDP, DRP, TN or DIN (for the 7-year dataset, r = 0.71, 0.48, 0.40, 0.55, 0.36, respectively, see Appendix D).

Table 6-10: Summary of linear regression relationships between chlorophyll *a*, DIN and conductivity at river sites in the Manawatu-Whanganui region. Averaged data from 2012 to 2015 was used for the analysis (log-transformed). Highly significant relationships ($P < 0.001$) are shown in bold and weaker but significant relationships ($P < 0.05$) in italics. Note the higher proportion of variance in chlorophyll *a* explained by conductivity than DIN and only moderately strong correlations between conductivity and DIN.

		DIN vs. chlorophyll *a*		Conductivity vs. chlorophyll *a*		Conductivity vs. DIN	
	n	R^2	P	R^2	P	R^2	P
All sites	61	**0.294**	**<0.001**	**0.407**	**<0.001**	*0.104*	*0.007*
Manawatu	36	**0.528**	**<0.001**	**0.582**	**<0.001**	**0.313**	**<0.001**
Whangaehu	11	0.426	0.018	0.298	0.048	0.000	0.417
Rangitikei	7	0.000	0.406	*0.641*	*0.019*	*0.578*	*0.029*

[2] As an example of "pristine waters", DIN concentrations in headwater sites in the Horizons dataset (e.g., mangatainoka_putara, makakahi_doc) are generally < 40 mg/m^3, with background conductivity of 40 to 60 µS/cm.

The Horizons dataset demonstrated that sites with both high conductivity (> 160 µS/cm) and high DIN (> 480 mg/m^3) tend to have higher peak chlorophyll *a* than sites with lower conductivity but equivalent DIN and DRP. The relationship with conductivity was seen mostly clearly across sites where periphyton was readily removed by small floods. Across these sites, the high conductivity sites may be more productive, or support more chlorophyll *a*-rich periphyton, than low conductivity sites, depending on DIN and DRP concentrations. In other words, sites with naturally high conductivity (e.g., from calcareous catchment geology) may be more vulnerable to developing high periphyton standing crop if DIN increases. An example is seen in a comparison of periphyton in two Canterbury Rivers, conductivity was taken to be the indicator of "enrichment" (Suren et al. 2003). Both rivers had extremely low DIN concentrations over an extended period of summer low flows (presumably due to instream uptake), but differed in conductivity (mean of 91 µS/cm in the Okuku River vs. 412 µS/cm in the Waipara River). Despite higher DRP in the Okuku River, the Waipara had consistently higher chlorophyll *a* and higher visible cover by filamentous algae and periphyton mats than the Okuku River. Note that at high DIN concentrations (e.g., > 1000 mg m^{-3}) DIN enrichment increases conductivity, and the effect of the two factors cannot be separated (see discussion above).

The main conclusion from the preceding discussion is that there is limited understanding of the mechanism behind the apparent positive effect of conductivity on periphyton standing crop. Nevertheless, data on this variable are almost universally collected as part of water quality monitoring programmes. In view of the strong relationships between conductivity and periphyton standing crop (as chlorophyll *a*) detected in the Horizons dataset and in previous studies (Biggs and Price 1987), and the results of earlier studies in Canterbury (Kilroy et al. 2017) it is suggested that conductivity data could be utilised more often as an indicator of background productivity at a site (as used by Suren et al. 2003). Further investigations into conductivity – biomass – community composition relationships using data from different regions and different geological settings could contribute to developing guidelines for applying a conductivity indicator, or at least a means to modify nutrient guidance in line with the effect on chlorophyll *a*. One possibility is that conductivity may be a better (or alternative) indicator of rivers that are "naturally productive" than the definition currently applied in the NPS-FM (NZ Government 2017)[3]. A conductivity indicator of naturally high productivity would need to take account of the effect of anthropogenic DIN in elevating conductivity, and some analysis is required before numerical values can be suggested.

6.4.6 Effective flow as a predictor variable

The choice of 3 x median flow as the periphyton removal threshold by Biggs (2000a) was based on an earlier analysis, which found that the metric FRE3 (annual frequency of events exceeding 3 x median flow) was the best of a wide range of hydrological variables for explaining a range of benthic community measures in New Zealand streams, including chlorophyll *a* (Clausen and Biggs 1997). FRE3 is a useful general metric and has been used to represent flows that typically remove periphyton in several studies (e.g., Booker 2013, Jellyman and Harding 2016). However, the threshold of 3 x median flow has different effects in different rivers (Hoyle et al. 2017). In this analysis, we applied a method for determining empirically, using a long dataset, the threshold for periphyton removal across a wide range of sites and found that the threshold, which we call the "effective flow", varies considerably (from 1.5 to at least 15 times median flow). Importantly, when we calculated

[3] The current definition of naturally productive waterways in the NPS-FM is as follows: "The Productive periphyton class is defined by the combination of REC "Dry" Climate categories (i.e., Warm-Dry (WD) and Cool-Dry (CD)) and REC Geology categories that have naturally high levels of nutrient enrichment due to their catchment geology (i.e., Soft-Sedimentary (SS), Volcanic Acidic (VA) and Volcanic Basic (VB)). Therefore the productive category is defined by the following REC defined types: WD/SS, WD/VB, WD/VA, CD/SS, CD/VB, CD/VA. The Default class includes all REC types not in the Productive class."

mean accrual period using the effective flow, the predictive performance of simple regression equations linking peak chlorophyll *a* with environmental variables improved compared with those in which accrual period was based on 3 x median flow.

We expected improvement in model performance when using accrual period based on effective flow rather than 3 x median partly because there is circularity in the procedure of developing the variable Da_EFF in that the effective flow at a site was derived from the same data that was used to derive the dependent variable (maximum or 92nd percentile of chlorophyll *a*). The circularity is minimised, however, because within-site data were used to derive Da_EFF, but the variable was then used to predict between sites. Thus, the variable Da_EFF is semi-independent of the dependent variable and, when averaged across years, Da_EFF and maximum or 92nd percentile of chlorophyll *a* were not strongly correlated in any of the annual or multi-year time periods tested in the analysis ($r < 0.25$ in all periods).

The statistical implications of potential circularity are unclear but probably not serious.[4] A practical implication is that the regression relationships are not classical predictive relationships, but more explanatory (i.e., trained on the data more strongly than usual). Nevertheless, ability to explain patterns of chlorophyll *a* over a region in a quantitative way is still useful for setting nutrient limits. If the regional dataset includes most sites of interest, then circularity in the regression relationships is likely unimportant, assuming the effective flow at each site remains stable. Effective flow is almost certainly driven mainly by the geomorphological characteristics at a site (Hoyle et al. 2017), with an interacting effect of climate (i.e., precipitation and river flows). River geomorphology is expected to remain stable unless there is a drastic change in flow regime such as that caused by installation of a dam (e.g., Opuha River; Lessard et al. 2013). At this stage, only empirical methods (such as those used in Section 3 and by Hoyle et al. 2017) or the field-based method in Hoyle et al. (2017) are available to determine the effective flow at a site. However, for predicting at completely new sites, then a straightforward means of estimating the effective flow is required. Using site characteristics may not be adequate, but a simple field procedure to allow estimation of the effective flow is under development (see Section 3.3.2).

[4] The problem of "spurious correlations" in regressions between non-independent variables is common in scientific literature (Brett 2004) and methods have been proposed to circumvent the problem (Gao & Zhang 2016). Examples of potentially spurious correlations involve variables that are much more clearly related than Da_EFF and maximum annual or 92nd percentile of chlorophyll *a* in the present study (e.g., regression of a ratio against the denominator of that ratio, Lake et al. 2016, Li et al. 2017). Therefore, the statistical problem here appears to be minimal.

7 Within-site analyses

Key messages

- A long periphyton dataset (>7 years) may enable an alternative approach to determining factors associated with variation in periphyton standing crop, by exploring relationships over time within sites.
- Explanatory variables were coeval, lagged and averaged DIN and DRP (over the previous 4 and 6 months). Lagged and averaged data were included because periphyton on a particular date has been influenced by preceding conditions.
- Regardless of the DIN metric used, most relationships between chlorophyll *a* and DIN were negative: high chlorophyll *a* was associated with low DIN.
- Negative correlations persisted even when the data were filtered to remove samples associated with high flows (i.e., when DIN tends to be high but chlorophyll *a* is low because of flood-removal).
- At some sites, low DIN at times of high periphyton could indicate high rates of uptake of DIN from the water. For example, at sites in the lower Rangitikei River, the negative correlation between chlorophyll *a* and DIN became stronger as data associated with high flows were removed from the dataset.
- Correlations between chlorophyll *a* and DRP were much weaker than for DIN and were positive or negative (but with low coefficient of determination).
- Reducing the dataset to annual peak chlorophyll *a* revealed shifts in the direction of the relationships between chlorophyll *a* and DIN or DRP from negative/neutral (using all data) to more positive. For DIN, 4% of sites with positive correlations increased to 33%; for DRP 22% increased to 41%.
- Adding in other variables (water temperature, accrual period) to predictive relationships for chlorophyll *a* (using all of the data) led to reasonably strong predictive models at some sites (e.g., cross-validated R^2 up to 0.6), although some models still included negative terms for DIN or DRP or both.
- Accrual period based on the effective flow was the only predictor that operated consistently across sites (using the between-site approach) and within sites, with a positive effect on chlorophyll *a*.
- There was no clear and simple linear relationship between periphyton standing crop and nutrient availability (as DIN or DRP) throughout the year, either across all flows, or in low flows only.

7.1 Introduction

In the between-site analysis, consistent predictors of maximum chlorophyll *a* across sites were the combination of conductivity, DIN and accrual time (generally as Da_EFF). From this result, the logical conclusion in relation to nutrients and periphyton would be that DIN is the primary nutrient controlling peak chlorophyll *a* in the Horizons Region. That conclusion carries at least two qualifications. First, the relationship is a correlation only, and cannot be interpreted as cause and effect without further evidence. Second, even with a stong regional pattern, periphyton at individual sites may be responding to nutrients in different ways. An analogy in freshwaters is that strong positive relationships between chlorophyll *a* and TN between sites in the Waikato Lakes could be interpreted as indicative of N control of phytoplankton biomass. However, detailed analysis of time-series data within-sites showed clearly that chlorophyll *a* was in fact controlled by P, and the between site pattern was an artefact caused by increasing DIN (contributing to TN) at all sites, coinciding with declining TP and chlorophyll *a*, as chlorophyll *a* production became P-limited over time (Verberg 2016). While controls on lake phytoplankton and river periphyton are very different, the example illustrates the important point that between-site analyses can obscure important patterns that occur only within sites. Consequently, an important aspect of this project was to examine relationships between chlorophyll *a* and nutrients and/or other variables within sites as well as between sites.

Conductivity (and its constituent concentrations of major ions) generally reflect the geological setting of a catchment and tend to remain characteristic at a site over time and therefore was not considered to be of primary importance in influencing variability in chlorophyll *a* over time. Long-term changes in conductivity can be driven by large changes in nutrient (mainly DIN) concentrations (Likens at al. 1970) and long-term climate trends (Lutz et al. 2012). Dissolved nutrient concentrations are more susceptible to rapid change over time through the effects of catchment modification caused by land-use change or extreme climatic events. Larned et al. (2016) detected significant trends in both DIN and DRP in many rivers over a 10-year period, including sites in the Horizons region, which the authors suggested were the result of changes in land cover or land use practices (for details see Snelder et al. 2014). Nutrient concentrations in rivers can also fluctuate widely over much shorter time scales, such as seasonally through biogeochemical processes. Dissolved inorganic nitrogen concentrations (but not DRP – see Table 4-1) in many New Zealand rivers are typically at their minimum in late summer (when both instream and terrestrial uptake are highest) and at their maximum in winter. Concentrations also vary with flow: high and low flows are usually associated with, respectively, high and low nitrate concentrations. Correlations with flow are seen less often for DRP.

Examples of demonstrated links between periphyton abundance and nutrient concentrations over time at the same site appear to be rare. In one example, trend analysis was used to link declining periphyton chlorophyll *a* coincided with declining nutrient (TP) concentrations following improvements to a waste-water treatment plant (Suplee et al. 2012). No examples of periphyton changes linked to diffuse nutrient inputs (either increasing or declining) within sites and over time were found in the literature in a recent review (Kilroy 2016). As Biggs and Close (1989) pointed out, straightforward within-site relationships between measures of biomass and nutrient concentrations cannot be expected because of the temporal autocorrelation of chlorophyll *a*, river flows and nutrient concentrations, including the effect of chlorophyll *a* on instream of DIN and DRP through uptake (Dodds 2007 and see discussion in Section 6.4.4). Thus, time-based lags rather than truly environmental effects that are independent of time (and of the dependent variable, chlorophyll *a*) can drive what appears to be an environmental relationship. In particular, uptake of DIN and DRP by periphyton and by other processes means that concentration estimates do not represent the true amount of bioavailable N and P at the time of the chlorophyll *a* measurement. The problem of the chlorophyll *a* – nutrient uptake process was avoided in the between-site analysis by using annual mean values of DIN and DRP to represent the average availability of N and P at a site, relative to the maximum biomass (or 92nd percentile when considering multiple years).

Relevant to the analysis below, relationships at each site between DIN and DRP and flow, and seasonality of DIN, DRP and chlorophyll *a* are summarised in Table 4-1. More details of the relationships (including plots of the relationships at each site are in Kilroy et al. (2016). These analyses highlighted that concentrations of nutrients (especially DIN) tend to be highest in high flows (during which periphyton is unlikely to be accruing) and these high concentrations generally occur in winter. Thus, nutrient availability is generally highest when periphyton growth is lowest. Another reason for the lack of published evidence for temporal changes in periphyton abundance that can be linked to increased nutrient concentrations from diffuse inputs may be a lack of long-term datasets that are robust enough to be able to separate the effects of flow variability on periphyton from the effects of nutrients. At over seven years long, the Horizons dataset is the longest consistent record amongst state-of-the-environment periphyton monitoring by regulatory agencies in New Zealand and represents the best regional dataset to employ this alternative within-site approach with which to determine environmental effects on periphyton production.

The analysis below is in two parts. First, we explored within-site relationships between chlorophyll *a* and DIN and DRP (Objective 1, item c2 in the contract, see Appendix A). Second, we explored relationships between multiple variables and chlorophyll *a* within sites, including checking for autocorrelation of variables (Objective 1, item 3).

Before running this section of the analysis we removed the fortnightly sampling data (see Section 2.1.1). To do this we took the first sample in a month for any site-year-month combination that had more than one sample in it.

7.2 Within-site relationships between chlorophyll *a* and nutrient concentrations

As discussed above, strong relationships between accrual period and chlorophyll *a* identified at some sites (Section 3.2.1), and seasonal and flow-driven fluctuations in DIN and DRP at many sites (Table 4-1), suggest that relationships with DIN and DRP concentrations will be difficult to identify unless flows are taken into account (Biggs and Close 1987). Furthermore, measurements of DIN or DRP made coincidentally with the periphyton sample collection are unlikely to reflect the nutrient concentrations that influenced biomass accumulation. Both prior nutrient conditions (over varying prior periods up to 6 months) and flows were accounted for in the following analyses.

7.2.1 Methods

We explored within-site relationships between chlorophyll *a* and DIN or DRP using simple correlation analyses rather than regression, to identify any broad patterns. Two approaches were used.

1. At each site, we generated Pearson correlation matrices of individual chlorophyll *a* datapoints ($\log_{10}$(n+1)-transformed) at each site, against a series of DIN and DRP variables. The variables were log-transformed DIN and DRP measured at the time of the chlorophyll *a* sample collection, DIN and DRP with lags of one and two months (e.g., chlorophyll *a*

collected in June compared with DIN or DRP measured in May and April), and DIN and DRP averaged over the preceding 4 or 6 months (up to and including the date of sample collection). In addition, the data were filtered in three steps:

1. removal of chlorophyll *a* samples collected within 21 days of a high flow (the effective flow) ("accrual" in results tables);
2. removal from the filtered dataset all chlorophyll *a* samples collected when flows were greater than median flow ("< median" in results tables); and
3. further removal of chlorophyll *a* samples collected when flows exceeded half the median flow ("<0.5 median" in results tables).

Including filter (1) meant that we used only data from sampling occasions when there had been at least 3 weeks during which no major periphyton removal occurred, which was expected to remove noise from the relationships. Filters (2) and (3) left only samples collected in low flows and very low flows after the 3-week accrual period, under which conditions periphyton accrual was expected. The analysis was restricted to 39 sites at which an effective flow had been identified and there was sufficient data. The total number of correlations run was 1560.

2. Correlation analyses of the annual maximum value of chlorophyll *a* against DIN and DRP variables averaged over the previous 4 and 6 months, including the date of chlorophyll a sample collection. Use of annual data limits the number of datapoints to seven, but this may be enough to identify patterns if they are strong. All sites with >5 years of data were included in this analysis (*n* = 53).

7.2.2 Results

Simple correlations: all data and flow-filtered data

Complete results are presented in Appendix G. Plots of chlorophyll *a* vs. 6-month averaged DIN and DRP are shown in Appendix H, to illustrate the general patterns. Across the whole analysis, 68% of the correlation coefficients were negative. More were negative for the DIN variables (80%) than the DRP variables (56%). The average negative correlation was stronger for the DIN variables than the DRP variables (mean r = -0.29 and -0.21 respectively). Positive correlations were weaker on average than the negative correlations and were similar for the two nutrients (r = 0.19 and 0.18 for the DIN and DRP variables, respectively).

On average, the 1-month and 2-month lagged DIN data yielded weaker correlations (both positive and negative) than the 4-month and 6-month averaged DIN (mean r = 0.22 and 0.24 for lagged DIN versus r = 0.31 and 0.32 for time-averaged DIN, absolute values). The mean correlation coefficient for chlorophyll *a* versus DIN at the time of sampling was r = 0.24 (absolute value, with most coefficients negative – see Appendix G). This pattern was the same for DRP (mean r = 0.21, 0.16, 0.18, 0.23 and 0.23, absolute values for, respectively, DRP at sampling, 1- and 2-month lagged DRP, and 4-month and 6-month averaged DRP).

The strengths of correlations varied across the different datasets (Table 7-1). Mean r was sometimes stronger using the datasets filtered to remove all but the lowest flows, but there was no consistent pattern. Note that the filtered datasets had low numbers of data points (8 to 29, mean 17, see Table 7-1 and Appendix G).

Table 7-1: Mean correlation coefficients between chlorophyll *a* and DIN and DRP variables from analyses on 39 sites using ten nutrient variables and four datasets. Negative and positive correlations are shown separately. Mean numbers of samples per site for each dataset were: all data, 68; accrual, 42; < median, 30; < 0.5median, 17.

Dataset	DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m
	Mean correlation coefficient (r) for nutrient variable:									
Average of negative correlation coefficients in each dataset										
all data	-0.33	-0.27	-0.29	-0.38	-0.35	-0.25	-0.13	-0.14	-0.25	-0.24
accrual	-0.23	-0.25	-0.26	-0.34	-0.35	-0.21	-0.14	-0.19	-0.24	-0.25
< median	-0.24	-0.24	-0.29	-0.35	-0.36	-0.23	-0.16	-0.20	-0.26	-0.26
<0.5median	-0.25	-0.23	-0.27	-0.35	-0.38	-0.26	-0.22	-0.21	-0.22	-0.22
Average of positive correlation coefficients in each dataset										
all data	0.12	0.07	0.08	0.18	0.15	0.13	0.15	0.13	0.16	0.18
accrual	0.16	0.15	0.14	0.23	0.18	0.15	0.17	0.16	0.21	0.20
< median	0.19	0.15	0.12	0.23	0.22	0.20	0.15	0.18	0.21	0.20
<0.5median	0.21	0.10	0.17	0.13	0.22	0.19	0.21	0.25	0.23	0.22

There was high variability across sites in the relationships between chlorophyll *a* and the different nutrient metrics and the responses within the four datasets. All 39 sites had at least eight negative relationships (out of the total of 40 possible correlation analyses per site). At 16 sites at least 30 of the 40 correlation results were negative (Table 7-2). Two sites (mangatainoka_us_tir, manawatu_ug) had negative coefficients in all 40 correlation tests, and a further 13 had negative coefficients for all 20 tests with DIN variables. The sites with mostly negative DIN – chlorophyll *a* relationships included most of the sites on the Mangatainoka and Manawatu Rivers, but did not include two headwater sites (makakahi_doc, maingatainoka_putara). Two further sites in the catchment also showed mostly positive (albeit weak) correlations with DIN (makuri_tuscan, tiraumea_nga). All four sites in the Rangitikei River had exclusively negative correlations between chlorophyll *a* and DIN (Table 7-2).

To highlight the primary reason for the negative correlations (i.e., highest nutrient availability at times of lowest potential for periphyton growth and accrual) we plotted DIN (value on the day of sampling) against water temperature. These plots (Appendix H) reinforce the pattern seen at almost all sites of high DIN in the coldest months (see Discussion below for more on the drivers of negative correlations). Positive correlations between chlorophyll *a* and DIN were usually weak ($r < 0.3$). The main exception to this was at ohau_sh1 where 14 positive correlations with DIN had a mean r of 0.38 (Table 7-2), and ranged up to $r = 0.73$ (DIN_6 m, in the dataset of <0.5 x median flows, Appendix G).

At many sites there were large differences in the correlations from the four different (and progressively smaller) datasets. In some cases the analyses on all of the data produced the strongest correlations between chlorophyll *a* and DIN, and these correlations were always negative (e.g., mangapapa_troup, manawatu_ug, manawatu_tc). At these sites, the strength of the negative correlation declined as the dataset was filtered to remove samples taken at higher flows (Table 7-1).

Table 7-2: Summary results of correlation analyses between chlorophyll *a* and DIN or DRP variables at each site. Mean chlorophyll *a* (mg/m^2), DIN and DRP (mg/m^3) are shown for each site. Forty correlations were run for each site (10 nutrient variables, four datasets). Datasets ranged from all of the data to data restricted to times following an accrual period of at least 21 days, and when flow was less than half the median flow. Numbers of negative correlations are shown, with mean correlation coefficients. Blank cells indicate no data. *These sites had 30 correlations only because there was no data less than half the median flow. Sites with highest mean coefficients are highlighted (bold red) for further discussion. See Appendix G for full results.

HRC no.	Site abbreviation	Means at site			No. of -ve correlations			Mean -ve coeff.		Mean +ve coeff.	
		Chl	DIN	DRP	DIN	DRP	All	DIN	DRP	DIN	DRP
1	makakahi_doc	2	37	7	3	5	8	-0.05	-0.11	0.15	0.11
2	mangatainoka_putara	1	17	5	15	6	21	-0.13	-0.04	0.09	0.11
3	mangatainoka_lars	6	60	6	20	12	32	-0.37	-0.28		0.10
7	mangatainoka_huk	7	667	7	19	11	30	**-0.54**	-0.20	0.18	0.15
8	kumeti_tr	5	667	10	18	20	38	-0.17	-0.39	0.03	
9	manawatu_weber	43	429	18	20	17	37	**-0.50**	-0.14		0.15
10	makakahi_ham	48	519	7	19	4	23	-0.15	-0.10	0.04	0.13
11	oroua_apiti	3	80	7	16	2	18	-0.24	-0.08	0.06	0.19
12	tamaki_ste	5	518	9	20	13	33	-0.35	-0.22		0.12
13	oruakeretaki_sh2	12	1020	14	19	15	34	-0.36	-0.20	0.11	0.04
14	makuri_tuscan*	89	862	9	4	11	15	-0.12	-0.30	0.19	0.04
16	mangatainoka_scarb	16	1061	6	20	20	40	**-0.42**	-0.32		
17	tiraumea_nga	97	604	11	2	16	18	-0.10	-0.34	0.19	0.13
19	mangatainoka_sh2	42	887	7	20	17	37	-0.27	-0.25		0.12
20	mangatainoka_ds_db	33	924	10	20	17	37	-0.26	-0.15		0.09
21	mangatainoka_us_pah	29	896	13	20	3	23	-0.25	-0.07		0.14
22	mangatainoka_ds_pah	45	1018	20	20	5	25	**-0.47**	-0.11		0.20
23	manawatu_hop	58	637	23	20	8	28	-0.25	-0.15		0.16
24	mangatainoka_us_tir	34	854	8	20	20	40	-0.28	-0.27		
26	mangapapa_troup	8	447	14	20	17	37	**-0.40**	-0.10		0.05
27	pohangina_mais	5	87	13	11	4	15	-0.14	-0.09	0.12	0.17
28	manawatu_ug	13	596	12	20	20	40	-0.29	-0.21		
29	oroua_almadale	4	160	9	10	11	21	-0.13	-0.24	0.09	0.17
33	manawatu_tc	9	439	11	19	14	33	-0.22	-0.13	0.18	0.14
34	manawatu_us_pncc	20	697	15	19	15	34	-0.27	-0.17	0.09	0.19
35	manawatu_ds_pncc	66	635	20	17	20	37	-0.30	-0.27	0.13	
36	manawatu_opik	32	577	17	13	20	33	-0.18	-0.36	0.10	
38	rangitikei_puk	5	30	7	20	1	21	-0.30	-0.01		0.13
39	moawhango_waiouru*	88	49	11	15	0	15	-0.23			0.27
40	rangitikei_man	11	67	7	20	12	32	-0.37	-0.08		0.15
43	rangitikei_one	10	88	10	20	8	28	**-0.49**	-0.11		0.13
44	rangitikei_mk	16	137	15	20	3	23	**-0.56**	-0.10		0.23

HRC no.	Site abbreviation	Means at site			No. of -ve correlations			Mean -ve coeff.		Mean +ve coeff.	
		Chl	DIN	DRP	DIN	DRP	All	DIN	DRP	DIN	DRP
46	makotuku_SH49	10	236	13	10	0	10	-0.15		0.08	**0.49**
50	mangawhero_pakihi	21	262	14	18	17	35	-0.28	-0.06	0.10	0.15
53	whakapapa_ds_gen	7	36	26	13	0	13	-0.31		0.15	0.23
59	waikawa_nmr	5	67	11	6	6	12	-0.12	-0.09	0.19	0.38
60	ohau_gladstone	3	57	9	17	16	33	-0.21	-0.17	0.05	0.14
61	ohau_sh1	6	242	116	6	11	17	-0.14	-0.19	0.38	0.18
62	ohau_haines	19	341	7	14	20	34	-0.29	-0.33	0.15	

Examples of patterns of flow in relation to DIN vs. chlorophyll *a* relationships at different sites are shown as contour plots in Figure 7-1 and Figure 7-2.

The plots highlight:

A. the wide range of patterns of flow in relation to DIN and chlorophyll *a*;

B. generally low chlorophyll *a* at high flows (yellow to red contours);

C. restriction of high chlorophyll *a* to generally low flows (green to blue contours), regardless of DIN.

Annual maximum chlorophyll *a* vs. preceding DIN and DRP

At most sites there was high temporal variability in mean DIN. The small number of datapoints meant that few sites showed a significant correlation between chlorophyll *a* DIN averaged over the 4 and 6 months preceding each survey. Plots of the data (see Appendix I) reflected those in approach 1 (see Section 7.2.1) in that many relationships were negative. At the few sites where there was a possible relationship (uncorrected $P < 0.05$), it was usually negative. The exception was at waitangi_ds_wai (positive, R^2=0.454, P=0.028). The results for DRP also reflected those of approach 1. No relationships were statistically significant (data not shown).

However, from Appendix I it appeared that the there were more positive (albeit non-significant) correlations than for the relationships between 6-month mean DIN and DRP and all chlorophyll *a* observations. Correlation corefficients from the plots in Appendix H and Appendix I are shown in Table 7-3. Summarising the results in Table 7-3, we see that for DIN, 48 sites (83%) had negative chlorophyll *a* vs. DIN relationships when all of the data were used, but this reduced to 29 sites (50%) when annual maxima were used. The change in positive relationships was from 4 (7%) to 19 (33%) sites. Corresponding numbers for DRP were 71% negative relationships reducing to 38%, and 22% positive relationships increasing to 41%.

Thus, while the relationships using annual maximum chlorophyll *a* were generally not significant, the slopes of the relationships suggested that at many sites, filtering out all data except for annual peak chlorophyll *a* may start to allow patterns of nutrient effects to appear in the data.

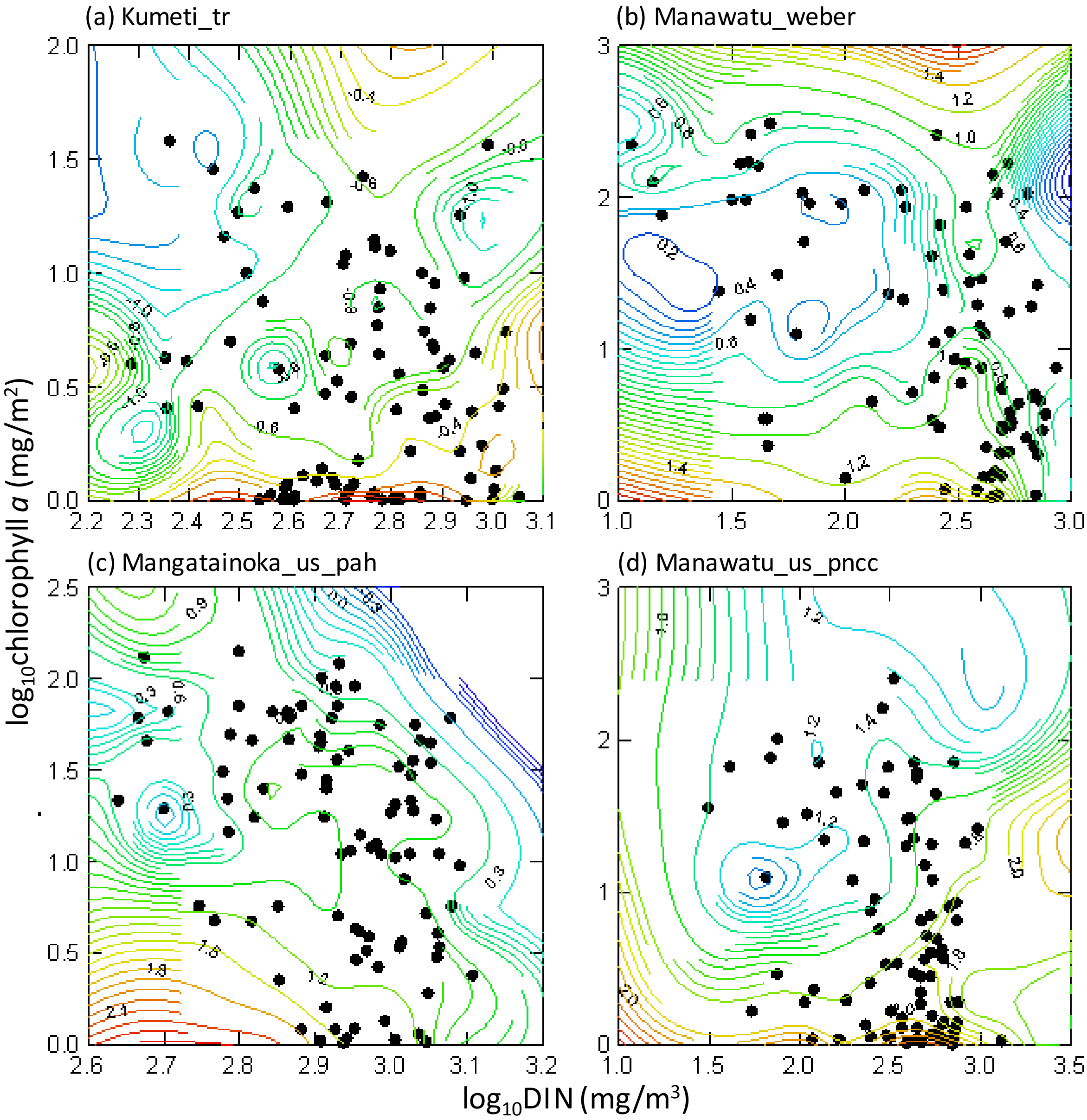


Figure 7-1: Contour plots of mean flow on the day of sampling on a matrix of chlorophyll *a* plotted against DIN, part 1. All data were log-transformed. The DIN metric is the 6-month mean up to and including the day of sampling. Red contours are high flows, blue contours are low flows.

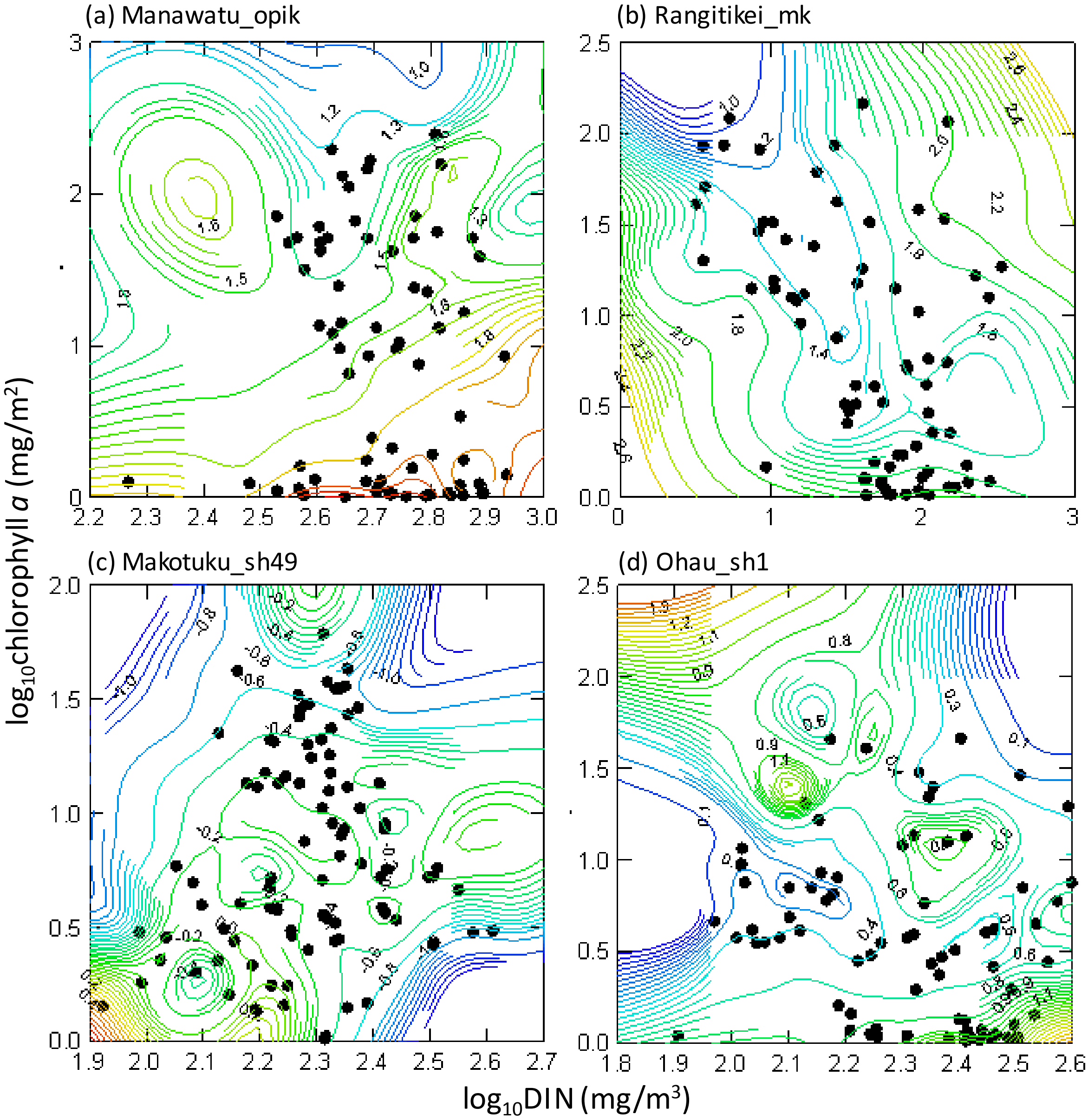


Figure 7-2: Contour plots of mean flow on the day of sampling on a matrix of chlorophyll *a* plotted against DIN, part 2. All data were log-transformed. The DIN metric is the 6-month mean up to and including the day of sampling. Red contours are high flows, blue contours are low flows.

Table 7-3: Within-site correlations (Pearson r values) between chlorophyll *a* and DIN or DRP at each site. In all all cases, the nutrient variables were the average over the past 6 months, including the date of the survey. The last two columns show nutrient limitation calculated from all of the data (i.e., no flow filters) so that all sites could be compared (taken from Table 4-1). Negative relationships (r < -0.2) are shaded in blue, positive (r > 0.2) in pink and no relationship in grey. Note transitions at many sites from negative or no relationship using all of the data to positive relationships using peak annual chlorophyll *a* only.

		Relationship with DIN		Relationship with DRP		Nutrient limitation	
HRCn	Site	All data	Ann. Max.	All data	Ann. Max.	Ratios	Saturating
1	makakahi_doc	0.19	0.04	-0.25	-0.65	N	co
2	mangatainoka_putara	-0.37	-0.01	-0.32	-0.41	N	co
3	mangatainoka_lars	-0.09	0.38	-0.04	-0.13	co	co
4	tamaki_res	-0.46	-0.26	-0.16	0.54	N	co
5	mangatera_us_dan	-0.49	0.42	0.19	-0.19	N	none
6	mangatera_ds_dan	0.01	-0.54	0.31	0.58	co	none
7	mangatainoka_huk	-0.42	0.96	-0.12	0.96	P	P
8	kumeti_tr	-0.15	-0.44	-0.42	-0.59	P	P
9	manawatu_weber	-0.30	0.07	-0.30	0.22	co	none-co
10	makakahi_ham	-0.27	0.26	0.06	-0.25	P	P
11	oroua_apiti	-0.50	-0.40	0.09	0.71	co	co
12	tamaki_ste	-0.34	0.16	-0.12	0.04	P	P-co
13	oruakeretaki_sh2	-0.45	-0.27	-0.08	-0.17	P	P
14	makuri_tuscan	0.13	0.48	-0.36	-0.55	P	P
15	pohangina_pir	-0.54	-0.28	0.12	-0.27	N	co
16	mangatainoka_scarb	-0.20	-0.87	-0.23	0.35	P	P
17	tiraumea_nga	0.90	0.74	-0.32	-0.45	P	P
18	mangatainoka_pahiatua	-0.41	-0.90	-0.50	0.38	P	P
19	mangatainoka_sh2	-0.33	0.17	0.02	0.74	P	P
20	mangatainoka_ds_db	-0.57	0.43	-0.15	0.06	P	P
21	mangatainoka_us_pah	-0.33	-0.54	-0.19	0.38	P	P
22	mangatainoka_ds_pah	-0.38	-0.41	-0.28	-0.33	P	P
23	manawatu_hop	-0.11	0.38	-0.37	-0.30	P	none
24	mangatainoka_us_tir	-0.44	0.19	-0.30	-0.25	P	P
26	mangapapa_troup	-0.52	-0.74	-0.22	-0.45	P	P-co
27	pohangina_mais	-0.19	0.43	0.20	-0.02	N	co
28	manawatu_ug	-0.52	-0.51	-0.21	-0.16	P	P-co
29	oroua_almadale	0.01	0.25	-0.19	-0.29	N	co
30	oroua_us_fei	-0.24	0.34	-0.22	0.36	co	none-co
31	oroua_ds_fei	-0.16	0.10	-0.29	-0.02	P	none-P
32	oroua_awahuri	-0.15	0.39	-0.22	0.41	P	none
33	manawatu_tc	-0.22	0.12	-0.29	-0.43	P	co-P
34	manawatu_us_pncc	-0.32	0.07	-0.34	-0.34	P	co-none-P

HRCn	Site	Relationship with DIN		Relationship with DRP		Nutrient limitation	
		All data	Ann. Max.	All data	Ann. Max.	Ratios	Saturating
35	manawatu_ds_pncc	-0.44	-0.76	-0.29	-0.28	P	none
36	manawatu_opik	-0.29	0.07	-0.16	-0.50	P	none-P
37	tokomaru_hb	-0.30	-0.50	0.18	-0.09	co	co
38	rangitikei_puk	-0.43	-0.65	0.20	0.03	N	co
39	moawhango_waiouru	-0.12	-0.38	0.35	0.42	N	co
40	rangitikei_man	-0.41	-0.64	-0.04	-0.13	N	co
41	porewa_us_hun	-0.07	-0.96	0.06	0.41	N	N-co
42	porewa_ds_hun	-0.07	-1.00	-0.04	0.50	N	N-co
43	rangitikei_one	-0.56	-0.56	-0.10	0.17	N	co
44	rangitikei_mk	-0.53	-0.03	0.16	0.51	N	co
45	mangawhero_doc	-0.17	-0.05	0.06	-0.34	N	co-N
46	makotuku_sh49	0.15	-0.01	0.50	0.45	P	co
47	mangawhero_us_oha	-0.32	0.26	-0.16	0.30	co	co
48	mangawhero_ds_oha	-0.36	-0.81	-0.14	0.93	co	N-co
49	makotuku_rae	0.00	0.08	-0.24	0.02	P	P-co
50	mangawhero_pakihi	-0.28	-0.51	-0.01	0.25	P	co-N
52	whanganui_ds_gen	-0.07	0.25	0.25	0.60	N	N
53	whakapapa_ds_gen	-0.34	-0.81	0.31	0.47	N	N
54	waitangi_us_wai	0.02	-0.20	-0.05	0.21	co	N
55	waitangi_ds_wai	0.29	0.28	0.33	0.52	co	none
57	makotuku_us_rae	-0.10	0.26	-0.13	-0.35	P	P-co
59	waikawa_nmr	-0.20	-0.65	-0.21	0.07	N	co
60	ohau_gladstone	-0.14	0.44	-0.27	0.44	N	co
61	ohau_sh1	-0.15	0.47	-0.10	0.44	P	co
62	ohau_haines	-0.16	0.01	-0.37	-0.14	P	co

7.3 Within-site relationships between chlorophyll *a* and multiple variables

The results of the regression analyses between chlorophyll *a* and accrual times (Section 3) indicated that, within sites, removal by high flows and subsequent biomass accrual are strong drivers of periphyton biomass. In contrast, the correlation analyses between chlorophyll *a* and nutrient metrics (Section 7.2) showed that direct stimulatory responses of chlorophyll *a* to nutrient concentrations within a site are difficult to detect in isolation from other environmental variables, even with a dataset spanning seven years. However, reducing the chlorophyll *a* dataset to annual peak biomass peak biomass and using that small dataset only (maximum of nine data points) suggested that year-to-year differences in chlorophyll *a* might by directly linked to DIN or DRP concentration in at some sites.

In this section, we investigated whether combinations of days of accrual (as discussed in Section 3) and other variables (including DIN and DRP) would improve ability to predict periphyton at a particular site under given nutrient concentrations and accrual times. We used multiple regression techniques as a starting point, in view of the ease with which the results can be disseminated, should they perform well.

7.3.1 Methods

We restricted variables included in the models to those that had a known effect on periphyton growth over time (i.e., based on the literature and on analyses earlier in this report). Conductivity was excluded because conductivity tends to be characteristic at a site and variability over time is strongly linked to flow (see Appendix J). Included variables were days of accrual based on the effective flow (Section 3), the dissolved nutrient variables used in Section 6, and water temperature. The only water temperature variable available were spot temperatures. Values of spot temperatures depend on time of day as well as time of year. However, in most regular monitoring programmes there is some consistency in the measurement time. Water temperatures also vary widely seasonally. Therefore, within a site, even monthly spot water temperatures can be a useful measure of annual and interannual water temperature variability.

Cross-correlations between predictor variables were assessed prior to running the models for each site (see Section 6.3 and Appendix B). We used generalised least squares (GLS) regression in R. An assumption of linear regression is that residuals are independently normally distributed. With repeat measurements from the same site over time there is potential for residuals of a model to be correlated over time (autocorrelation). To account for this the following procedure was performed for each site:

- stepwise linear multiple models were fitted on temporal data;
- the final stepwise model was run both with and without a residual autocorrelation structure that accounts for covariation in the residuals. This was done using the 'correlation' option in the GLS function in R with AR-1 autocorrelation as the autocorrelation structure;
- likelihood ratio tests (α = 0.05) were used to test whether addition of the autocorrelation structure improved model fit. If yes, the structure was included in the final model, otherwise the model was run without the autocorrelation structure.

The fit of each model was assessed using leave-one-out cross-validation, using the variables selected it in the stepwise procedure. We relied on NSE to assess model fit in these analyses. NSE is specifically used to quantify how well a model simulation can predict the outcome variable. NSE is proportional to R^2 except that R^2 cannot take on negative values (see Section 7.3.1 for details).

Three sets of models were run. DIN and DRP lagged by one and two months were relatively weakly correlated (mean Pearson r < 0.45 across all sites) and were therefore included in the same model. Mean values of DIN and DRP over the past 4 and 6 months were strongly correlated (average Pearson r = 0.87 across all sites). Therefore, separate models were run for the 4-m and 6-m means. The basic models tested were:

$\log_{10}$Chla as a function of $\log_{10}$DIN_lag1 + $\log_{10}$DRP_lag1 + $\log_{10}$DIN_lag2 + $\log_{10}$DRP_lag2 + $\log_{10}$Da_EFF + Temperature

$\log_{10}$Chla as a function of $\log_{10}$DIN_4m + $\log_{10}$DRP_4m + $\log_{10}$Da_EFF + Temperature

log_{10}Chla as a function of log_{10}DIN_6m + log_{10}DRP_6m + log_{10}Da_EFF + Temperature

In addition, we ran models using log_{10}Da_EFF as the only predictor variable, to provide a direct comparison (using leave-one-out cross-validation) with the complex models.[5] The analysis was run on data from 40 sites at which we identified an effective flow, and at which there was sufficient data.

7.3.2 Results

Addition of the autocorrelation structure did not improve the fit in any of the models and we therefore excluded the autocorrelation structure from final models. On average, the models using 6-month mean DIN and DRP yielded slightly better fits in the cross-validation tests than the models using the lagged DIN and DRP data or the 4-month means (Table 7-4). The difference was small and choice of one metric over the other is essentially arbitrary. We focused the models using 6-month means for further discussion.

DIN was included in the models at 25 sites. The DIN coefficient was negative in all but four cases (makuri_tuscan, oroua_almadale, manawatu_opik, makotuku_SH49). DRP was included in the models at 15 sites, with negative coefficients at seven sites, and positive at eight. Both nutrient variables were included in the models at 10 sites. Both DIN and DRP had negative coefficients at three sites (tamaki_ste, oruakeretaki_sh2, mangawhero_pakihi) and both had positive coefficients at one site (makotuku_sh49) (Table 7-4).

Da_EFF was included as a variable in all of the models, and was the sole variable selected at the following seven sites: mangatainoka_putara; manawatu_hop; manawatu_ug; manawatu_ds_pncc; moawhango_waiouru; ohau_sh1; ohau_haines. A comparison of leave-one-out cross-validation with only Da_EFF available as a variable with the full models including the nutrient variables and water temperature showed that in most other cases NSE improved (i.e., predictive ability was better) when other variables were included (Table 7-4).

All models at the headwater sites, makakahi_doc and mangatainoka_putara, were weak (NSE < 0.12). These sites had generally low chlorophyll *a* that was not strongly correlated with accrual time. Across other sites, NSE ranged from <0.2 to 0.61 at oruakeretaki_sh2.

Water temperature was initially selected for the models at 12 sites. Omitting water temperature as an available variable reduced NSE at nine of these sites, but made little difference at tamaki_ste, rangitikei_one and whakapapa_ds_gen. Therefore, the models not including water temperature were preferred at the latter three sites (Table 7-4).

Plots of the observed versus predicted values highlighted high scatter in the data, even at the higher values of NSE (e.g., NSE > 0.4) but reduced scatter at the site with highest NSE (0.61, at oruakeretaki_sh2) (Figure 7-3). There was bias in some of the predictions. For example, at rangitikei_mk chlorophyll *a* was generally underpredicted at high values and over predicted at low values (Figure 7-3).

[5] The results from leave-one-out cross validation are expected to differ from the linear regression results shown in Table 5-1 and provide a better test of predictive ability when applied to new data. Cross validation results from of all the models can be compared directly.

Table 7-4: Summary of performance of regression models for predicting chlorophyll *a* within sites. NSE is Nash-Sutcliffe Efficiency, which summarises the performance of leave-one-out cross-validation tests (1 is a perfect fit, 0 is no better than predicting the mean value of the dataset). In this case NSE and R^2 were almost identical (except that R^2 does not have negative values). Models were run using DIN and DRP lagged by 1 and 2 months (lag12), and DIN and DRP averaged over the previous 4 and 6 months (4m, 6m). y = variable included in best model. Shaded cells indicate negative coefficients. NSE, no temp shows the effect on NSE of omitting temperature from the available variables. NL = nutrient limitation status from Table 4.2.

HRC code	Site abbreviation	n	NSE in different models				Coefficients in best 6m model				NSE, no temp.	NL
			Da_EFF only	lag12	4m	6m	DIN	DRP	Da_EFF	Temp		
1	makakahi_doc	34	0.00	-0.01	0.03	0.03			y	y	-0.01	co
2	mangatainoka_putara	87	0.09	0.14	0.11	0.11			y			co
3	mangatainoka_lars	35	0.41	0.45	0.48	0.47	y		y	y	0.41	co
7	mangatainoka_huk	34	0.21	0.43	0.44	0.47	y		y	y	0.43	P
8	kumeti_tr	90	0.44	0.50	0.51	0.49		y	y			P
9	manawatu_weber	83	0.35	0.39	0.41	0.43	y		y			co-none
10	makakahi_ham	82	0.12	0.17	0.15	0.20	y	y	y			co-P
11	oroua_apiti	85	0.36	0.45	0.43	0.46	y		y			co
12	tamaki_ste	81	0.41	0.48	0.51	0.48	y	y	y	y	0.48	co-P
13	oruakeretaki_sh2	85	0.45	0.54	0.58	0.61	y	y	y			none-P
14	makuri_tuscan	83	0.28	0.30	0.34	0.34	y	y	y			P
16	mangatainoka_scarb	33	0.10	0.15	0.16	0.18	y		y			P
17	tiraumea_nga	62	0.50	0.53	0.54	0.58		y	y			P
19	mangatainoka_sh2	87	0.47	0.50	0.49	0.50	y		y			P
20	mangatainoka_ds_db	81	0.30	0.37	0.38	0.39	y		y			P
21	mangatainoka_us_pah	85	0.43	0.42	0.49	0.47	y		y			P
22	mangatainoka_ds_pah	76	0.35	0.41	0.43	0.43	y	y	y			P
23	manawatu_hop	83	0.43	0.46	0.45	0.45			y			none-N
24	mangatainoka_us_tir	60	0.33	0.35	0.37	0.38	y		y			P
26	mangapapa_troup	91	0.26	0.41	0.43	0.42	y		y	y	0.37	co-P
27	pohangina_mais	85	0.25	0.30	0.29	0.28		y	y	y	0.26	co
28	manawatu_ug	81	0.38	0.47	0.46	0.46			y			P-co
29	oroua_almadale	82	0.17	0.21	0.19	0.19	y		y			co
33	manawatu_tc	79	0.41	0.47	0.47	0.47			y	y	0.40	co
34	manawatu_us_pncc	83	0.49	0.57	0.58	0.58		y	y			co-P
35	manawatu_ds_pncc	50	0.41	0.44	0.45	0.45			y			none-P
36	manawatu_opik	69	0.48	0.60	0.60	0.57	y	y	y	y	0.53	none-P
38	rangitikei_puk	76	0.28	0.27	0.26	0.27	y		y			co
39	moawhango_waiouru	53	0.14	-0.02	0.24	0.17			y			co
40	rangitikei_man	78	0.26	0.22	0.22	0.23	y		y			co
43	rangitikei_one	75	0.29	0.41	0.39	0.41	y	y	y	y	0.42	co
44	rangitikei_mk	74	0.35	0.37	0.35	0.41	y	y	y			co

HRC code	Site abbreviation	n	NSE in different models: Da_EF F only	lag12	4m	6m	Coefficients in best 6m model: DIN	DRP	Da_EFF	Temp	NSE, no temp.	NL
46	makotuku_sh49	87	0.08	0.22	0.23	0.28	y	y	y			co-N
50	mangawhero_pakihi	78	0.28	0.33	0.31	0.31	y	y	y			co-N
53	whakapapa_ds_gen	61	0.29	0.34	0.28	0.30		y	y	y	0.29	N
59	waikawa_nmr	86	0.17	0.15	0.17	0.21	y		y	y	0.16	co
60	ohau_gladstone	90	0.18	0.20	0.23	0.20	y		y	y	0.18	co
61	ohau_sh1	58	0.33	0.40	0.37	0.38			y			co
62	ohau_haines	37	0.27	0.30	0.32	0.28			y			co

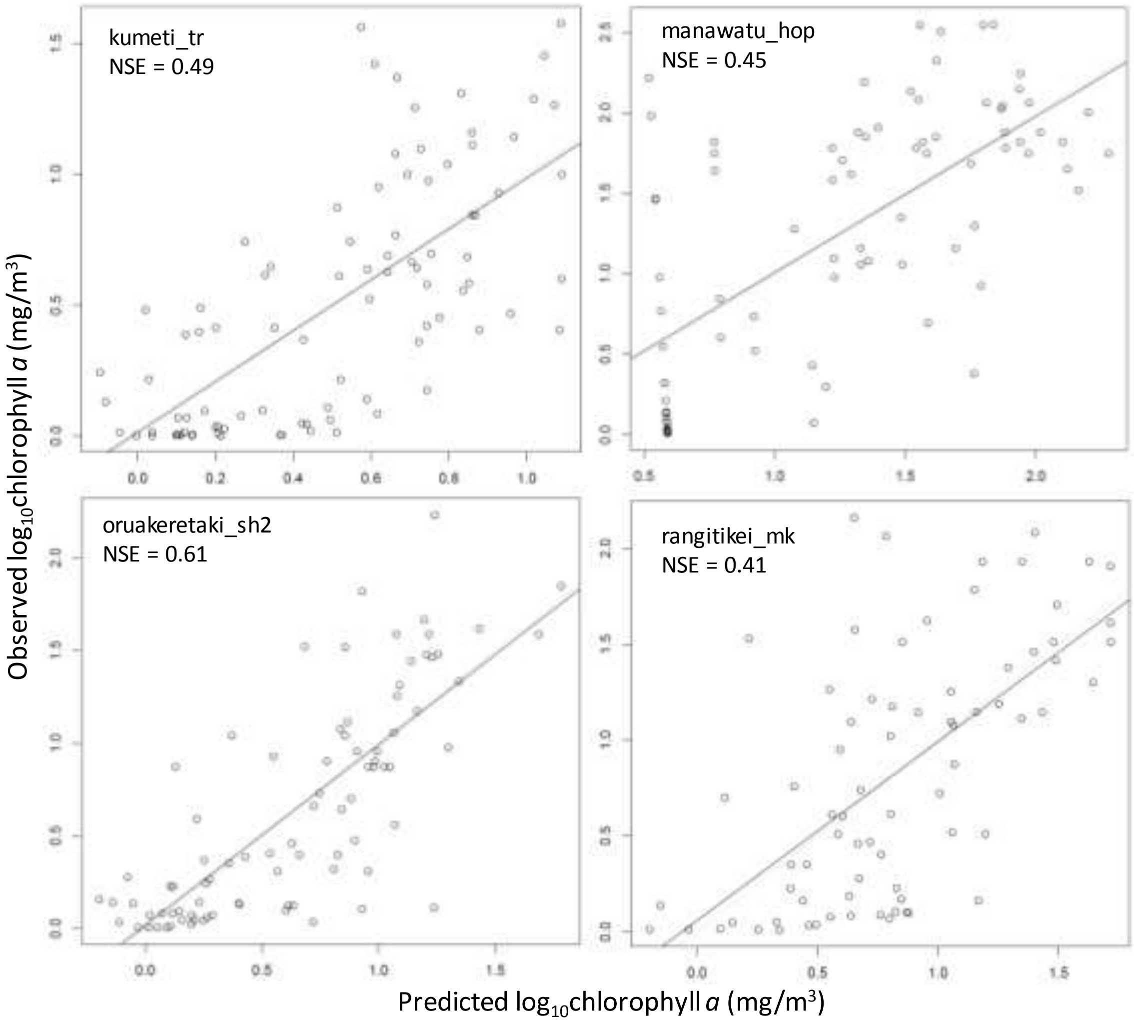


Figure 7-3: Example plots of observed versus predicted chlorophyll *a* from the leave-one-out cross-validation tests. The solid line is the best fit line through the points. The 1 : 1 relationships are shown by the dashed lines (very faint, and almost coincident with the 1 : 1 lines). At kumeti_tr, manawatu-hop and oruakeretaki_sh2, the predictions are relatively unbiased, but there is variable scatter in the data (highest at manawatu_hop). At rangitikei_mk there is scatter an also bias, with under predicted high values and over-predicted low values.

7.4 Discussion

7.4.1 Within-site relationships: flow vs. nutrients

The above series of analyses showed that, at many sites, periphyton chlorophyll *a* variability over time (between 2008 and 2016) was, as expected from the analysis in Section 3, influenced more by variability in flow rather than variability in nutrients. In the single-variable analyses with nutrient concentrations (Section 7.2), most correlations between chlorophyll *a* and DIN, using all of the data, were negative, reflecting the pattern at most sites of a positive relationship between flow and DIN. Chlorophyll *a* was not strongly correlated with DRP concentrations in the water column at most sites, even though almost half of the sites were assessed using classical nutrient-limitation theory as P-limited (based on DIN and DRP concentrations at flows less than median flow) (see Table 4-1). Biggs and Close (1989) pointed out that "high correlations nutrient concentrations and flow indicated that the nutrient data were also carrying hydrological information and that simple causal relationships between nutrients and biomass are difficult to establish in rivers". That observation appears to hold even with a time-series of eight years

In view of the potential for shifts in nutrient limitation at low flows at some sites (see Section 4.3.3), and the likelihood that most periphyton accrual occurs under low flows, we also ran correlations between chlorophyll *a* and the nutrient variables using samples collected at successively lower flows. Under the combination of a filter of at least 21 days since an effective flow and a filter for low flows (< half median flow) we expected that correlations between periphyton biomass and any underlying temporal changes in DIN or DRP over time (long-term, annual or seasonal) should more detectable.

For DIN, while there was no general pattern of negative correlations with DIN becoming weaker or reversing to positive along the gradient of all flows → flows > 21 days since a flood → flows < median → flows < half median, a few sites showed this pattern. These sites included: mangatainoka_us_tir, mangapapa_troup, manawatu_tc, manawatu_ug, manawatu_us_pncc, manawatu_ds_pncc, manawatu_opik, ohau_sh1. At several sites the opposite occurred, with negative correlations becoming stronger over the flow gradient, especially with DIN_4m and DIN_6m. These sites included: mangatainoka_lars, mangatainoka_huk, rangitikei_man, rangitikei_mk, rangitikei_one, mangawhero_pakihi, ohau_gladstone (see Appendix E for details).

For DRP, fewer sites than for DIN showed a trend across the gradient of successively filtering out samples collected within 21 days of an effective flow and then flows greater than the median flow and half median flow. The pattern varied depending on which DRP metric was used. For DRP-6m, increasingly positive correlations were seen at makakahi_doc, makakahi_ham, manawatu_hop. No sites showed a clear pattern of increasingly negative correlations across the gradient. Presumably the difference between DIN and DRP reflected the fact that there were strong correlations between DIN and flow at over 65% of sites (most of which were positive), but strong correlations between DRP and flow at only 25% of sites (again mostly positive) (for details see Section 4.3.1).

The second approach to the analysis of relationships between chlorophyll a DIN or DRP alone (i.e., focusing on relationships between annual peak chlorophyll *a* and averaged DIN or DRP values in the preceding months) required working with very small datasets (maximum of 9 datapoints, if incomplete years were included). Using this approach indicated that within-site positive relationships between nutrients and chlorophyll *a* (that were otherwise obscured by the effects of flows) could be revealed. The shift to more frequent within-site positive relationships (for both DIN and DRP) (from results summarised in Table 7-3; see also Appendices F and G) suggests that even longer time series of data will be helpful. The second approach effectively created a sub-set of data that represented optimal conditions (in each year) for periphyton standing crop. Optimal conditions are more likely to occur at times when climatic conditions (light, temperature) are less likely to constrain periphyton growth.

In summary, the results of the within-site analyses showed that, using all of the data, low DIN was rarely associated with low chlorophyll *a* at a particular site, even at sites identified as N-limiting, especially when all of the data were used. In fact, low DIN – at sites with variable DIN – was a better predictor of high chlorophyll *a*. Overall, the variables included in the models at each site showed no clear pattern in relation to nutrient limitation of periphyton growth as assessed at low flows (see Section 4.3.3). For example, models included DIN at sites that were assessed as co-limited, P-limited and not limited by either N or P.

The analysis overall demonstrates a key finding: there is no clear and simple linear relationship between nutrient availability throughout the year and corresponding algal standing crop. GLS regression model selection confirmed that flow was the key determinant of changes to algal standing crop over time within stations, throughout the Horizons region. It should be stressed that negative associations between DIN (and less so, DRP) with chlorophyll *a* over time at many sites do not indicate that DIN is having some inhibitory effect on algal production. This is a classic case of "correlation is not causation". The pattern simply corresponds to the dominance of flow on controlling algal production and in turn, the positive association between flow with DIN (and less so, DRP).

Terms included in each model were used to develop a site classification, which is discussed further in Section 10.

7.4.2 Demostration of instream uptake of DIN

Patterns seen across the four sites on the Rangitikei River may illustrate the effect of instream uptake of DIN down a river continuum. At all four sites, DIN was correlated with flow, and also varied across seasons (with low DIN in summer, see Appendix C in Kilroy et al. 2016). The negative DIN – chlorophyll *a* correlation at low flows in the Rangitikei River increased in strength in a downstream direction and as the data were filtered to include only low flows (Table 7-5).

Such a downstream gradient could indicate that at low flows, N concentrations may be directly affected by instream uptake (from periphyton growth) (Rode et al. 2016). The Rangitikei is a large river (minimum flow usually >10 m^3/s and median flow of >45 m^3/s at the three downstream sites). Rates of N uptake relative to supply, even at low concentrations, would therefore account for only a small proportion of the available N load at a particular point in the river. During low flows, the effect of uptake on concentrations would become more pronounced over time and with distance downstream – leading to increasing strength of the negative DIN – chlorophyll *a* correlation in a downstream direction, as noted above. All four sites in the Rangitikei had relatively strong relationships between chlorophyll *a* and accrual period (following flows > 4 x median, Table 3-1), but including DIN (and also DRP, with a positive effect, Table 7-4) in within-site relationships at the two downstream sites strengthened the relationships further.

Table 7-5: Pattern of correlation between chlorophyll *a* and DIN at four sites in the Rangitikei River. Sites are in order from upstream to downstream. Note increasing strength of negative correlation from upstream to downstream, and along the datasets that grade from all flows to very low flows only, with the strongest gradient at the most downstream site.

		Correlation between chlorophyll *a* and DIN in different datasets			
HRCn	**Site**	**all data**	**after accrual**	**after accrual, <median**	**after accrual, <half_median**
38	rangitikei_puk	-0.39	-0.42	-0.45	-0.34
40	rangitikei_man	-0.42	-0.39	-0.61	-0.64
43	rangitikei_one	-0.56	-0.55	-0.65	-0.71
44	rangitikei_mk	-0.52	-0.51	-0.76	-0.90
No. of samples (range):		67 - 75	48 - 55	34 - 38	16 - 20

7.4.3 Positive nutrient – chlorophyll *a* correlations

The 11 sites in Table 7-4 at which relationships in which DIN or DRP had a positive effect on chlorophyll *a* may be of particular interest. To aid discussion on these sites further, they are listed in Table 7-6, along with site characteristics.

Table 7-6: List of sites for which within-site predictive models included DIN and/or DRP with a positive coefficient. NSE is shown for Da_EFF only at each site, and then for the best model, including DIN_6m and DRP_6m. Variables included in the best model (in addition to Da_EFF are indicated by "y". Shaded cells indicate negative coefficients. Average concentrations of DIN and DRP are shown (geometric means, in mg/m^3). Trends from Kilroy et al. (2016). Sites discussed below are highlighted in red and blue.

			Model NSE		Variable including in model			Mean conc. (mg/m³)		Trends, 2008 - 2016		Nutr. liimit.
HRC code	**Site abbreviation**	**n**	**Da_EFF**	**Best comb**	**DIN**	**DRP**	**Temp**	**DIN**	**DRP**	**DIN**	**DRP**	
10	makakahi_ham	82	0.12	0.20	y	y		300	6	no	no	co-P
14	makuri_tuscan	83	0.28	0.34	y	y		821	8	no	no	P
22	mangatainoka_ds_pah	76	0.35	0.43	y	y		893	11	no	neg	P
27	pohangina_mais	85	0.25	0.28		y	y	38	12	no	no	co
29	oroua_almadale	82	0.17	0.19	y			60	9	no	no	co
34	manawatu_us_pncc	83	0.49	0.58		y		312	12	neg	neg	co-P
36	manawatu_opik	69	0.48	0.57	y	y	y	519	14	no	neg	none-P
43	rangitikei_one	75	0.29	0.42	y	y		46	8	no	pos	co
44	rangitikei_mk	74	0.35	0.41	y	y		48	12	no	no	co
46	makotuku_sh49	87	0.08	0.28	y	y		191	9	neg	pos	co-N
53	whakapapa_ds_gen	61	0.29	0.30		y	y	25	24	no	no	N

The first thing to note from Table 7-6 is that, at most sites, chlorophyll *a* can be predicted using accrual time (Da_EFF) almost as accurately as using accrual time combined with DIN, DRP and /or temperature. That is, NSE shown in the "Best comb" column is only slightly higher than NSE in the "Da_EFF" column. We considered that a difference in NSE of ≥0.1 would represent a reasonable difference in model performance through including DIN or DRP. Two sites met this criterion, indicating that adding a nutrient variable (DIN or DRP) to the flow relationships improved ability to predict chlorophyll *a* over time. The sites (marked in red on Table 7-6) are:

- rangitikei_one: including DIN (negative coefficient) and DRP (positive coefficient) improved the model NSE from 0.26 (with Da_EFF alone) to 0.42. DIN concentrations were generally low at this site, with significant negative correlation with flow and seasonal variability (Table 4-1). DRP concentrations were moderate and between 2008 and 2016 significant upward trend in DRP concentrations and in chlorophyll *a* were recorded (Kilroy et al. 2016);

- makotuku_sh49: including DIN and DRP (both positive coefficients) improved the model NSE from 0.08 (with Da_EFF alone) to 0.28. Both DIN and DRP were low to moderate at this site and both had negative correlations with flow, and varied across seasons. Between 2008 and 2016, DRP increased at this site, DIN declined, and chlorophyll *a* increased (Kilroy et al. 2016).

Thus, the only two sites at which the within-site models for predicting chlorophyll *a* included DRP with a positive coefficient also showed an upward trend in DRP (from 2008 to 2016) along with an upward trend in chlorophyll *a*.

Two further sites highlighted in Table 7-6 (in blue) are mangatainoka_ds_pah, and manawatu_us_pncc. Both showed a positive effect of DRP on periphyton over the 8-year period, with addition of DRP increasing model NSE by 0.08 and 0.09 over the effect of flow (as Da_EFF). Between 2008 and 2016 DRP declined at both sites (Kilroy et al. 2016). While no corresponding declines in chlorophyll *a* were recorded, at both sites there were significant declines in percentage cover by mats and filaments (see Table 4-11 in Kilroy et al. 2016), suggesting a change in periphyton driven by the decline in DRP.

To establish that a trend over time in DIN or DRP is potentially stimulating or reducing periphyton (as chlorophyll *a* or percentage cover) the simplest approach may be to run trend analyses on both the nutrient concentrations and periphyton chlorophyll *a* and look for corresponding trends. This was the approach taken by Suplee et al. (2012). Kilroy et al. (2016) noted coincident trends in the Horizons region, which substantiate the present analysis that within-site relationships between chlorophyll *a* and DRP concentrations can indicate that recent changes in DRP may be the driver of corresponding changes in periphyton abundance.

Another approach is to apply the data to mechanistic models of periphyton abundance in streams, such as those developed for the Tukituki River (Rutherford et al. 2000, Rutherford 2013). Such models attempt to describe periphyton accrual taking into account the major biomass growth and loss process and interactions with nutrient concentrations (including the effects of instream uptake). Currently coordinated efforts are underway in two research programmes at NIWA to develop a generalised model for predicting chlorophyll *a* in streams. The starting point is to model processes at a small scale in experimental channels, and then test the models using river data such as the fortnightly data collected by Horizons Regional Council (see Section 8 below).

7.4.4 Time series versus space-for-time

The between-site analysis of chlorophyll *a* vs. environmental variables described in Section 6 took a space-for-time approach to identifying predictors of chlorophyll *a* across the Horizons region. The space-for-time approach is commonly applied in ecology (both terrestrial and aquatic systems) for modelling environmental drivers, which are then used to infer into the past or future (e.g., Blois et al. 2013). A classic application of the approach is in palaeolimnology, in which past environmental conditions lakes are inferred based on relationships between biota (e.g., diatoms) and environmental conditions in present-day lakes (e.g., Augustinus et al. 2006, Boeff et al. 2016). Other applications of the space-for-time approach include validating biomass predictions in forest landscapes (Ma et al. 2017) and predicting effects of drying periods in a wetland (Banet and Trexler 2013). As far as we are aware, space-for-time substitution has been the only approach used for assessing the effects of catchment and hydrological changes on periphyton, and even the space-for-time approach has had limited success (see discussion, Section 6.4.2). Therefore, our ability to develop relatively strong simple regression equations (e.g., cross-validated $R^2 > 0.7$) for predicting between sites (i.e., Section 6) was unusual.

The time-series analysis (within sites) described in this section is a new approach for periphton in rivers. The analysis demonstrated the issues that arise in developing periphyton – environment relationships within sites. As discussed above, the issues are well known (Biggs and Close 1989). In a more recent attempt at identifying factors influencing periphyton using a space-for-time approach, Munn et al. (2011) suggested a conceptual model for understanding the reasons for deviation of relationships from those expected. Here the expected relationships were based on a physiological understanding of, and experimental findings on, the responses of chlorophyll *a* to nutrient concentrations. The deviations inevitably lead to weak explanatory ability of linear models (Figure 7-4).

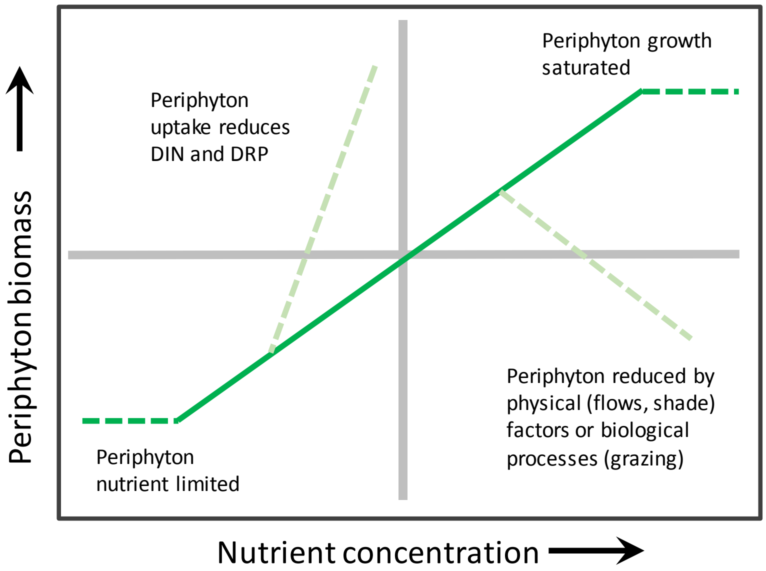


Figure 7-4: Conceptual model for explaining the interactions between periphyton chlorophyll *a*, nutrient concentrations and physical factors. The model provides a simple framework for understanding why chlorophyll *a* - nutrient relationships developed from field data rarely conform to expected patterns, as seen in nutrient enrichment experiments or predicted by growth models. The bright green line is the idealised growth trajectory. Pale green dashed lines indicate deviations from the trajectory under different conditions. The conceptual model was developed to explain a space-for-time approach, but applies even more strongly to a time-series approach. Adapted from Munn et al. (2011).

The conceptual model in Figure 7-4 was developed to explain deviations in a space-for-time approach. The analysis in Section 6 overcame the problem of accounting for nutrient uptake (top left quadrant of the model) by focusing on peak annual chlorophyll *a* (or the 92nd percentile for multiple years) and using DIN and DRP data averaged over the same time period (to represent averaged nutrient availability over times of high and low concentrations). The problem of accounting for physical disturbances (lower right quadrant in the model) was overcome by developing a metric for accrual time based on the flow magnitude threshold most likely to remove periphyton (Da_EFF). Consequenly, our models had generally higher explanatory power than any of those developed by Munn et al. (2011) (maximum R^2 = 0.54 in Munn at al. 2011, compared to maximum R^2 = 0.78 in Table 6-4).

The conceptual model applies even more strongly in the time-series approach because with time series we need to link single measurements of the dependent variable (chlorophyll *a*) with linked single measurements of the predictor variables. Physical disturbance was again accounted for by using Da_EFF as the predictor variable. As with the space-for-time approach, averaged values of DIN and DRP (over six months) appeared to account for some of the uptake effect (top right quadrant in Figure 7-4), but the effect was still clear at some sites (e.g., in the Rangitikei).

In summary, with the benefit of a long and detailed dataset at multiple sites, it has been possible to demonstrate that within-site analyses may be able to reveal relationships with potential drivers of biomass other than the effects of flow. This was achieved by either isolating peak biomass, or restricting data to that collected at times of very low flow and with no preceding high flows. In both cases, longer datasets would improve the chances of detecting "real" relationships. In the meantime, the space-for-time approach provides a useful framework for predicting chlorophyll *a* across sites in the Horizons region, recognising that there is variability (error) associated with any predictions, which may be severe at some sites.

8 Effect of using fortnightly vs monthly datasets

Key messages

- Horizons has collected periphyton data at fortnightly intervals at a subset of the monitoring sites. Data from 12 sites were used to compare the predictive ability of within-site models derived from fortnightly and monthly datasets at these sites. Data were available at each site from between 17 and 24 months.
- The models from fortnightly data performed similarly to or better than the models using monthly data in 11 of the 12 sites tested. Only at ohau_gladstone did monthly data predictd periphyton biomass more accurately than fortnightly.
- Poor performance of monthly data over the period of fortnightly surveys (17 to 24 months) may be attributable to low numbers of samples.
- The 17- to 24-month time series of fortnightly data from the 12 sites generally did not yield stronger predictive relationships than using the complete (up to 7 years) monthly dataset at the same sites (although noting that the datasets were not strictly comparable in numbers of samples or variables included).
- We concluded that fortnightly data in some cases could allow relationships to be developed over a shorter time period. Fortnightly datasets have other applications, including more accurate estimation of accrual rates, and testing of mechanistic models of periphyton growth.

8.1 Introduction

Selected sites in the Horizons dataset have been monitored fortnightly rather than monthly. Objective 2 in the contract was to determine whether using the higher resolution data would affect the outcome of analyses of within-site relationships between chlorophyll *a* and environmental variables. Specifically, the contract (see Appendix A) asked for tests to show whether fortnightly resolution improved model fit over monthly sampling and whether the strength of different predictors varied between monthly and fortnightly datasets. The possibility being considered was that certain environmental variables might vary over shorter time scales than monthly, and that the effects of these shorter-term variations might be detected in the fortnightly data set but not in the monthly dataset.

The results of the within-site analyses on time series of up to eight years (Section 7) highlighted the lack of strong positive correlations between chlorophyll *a* and nutrient concentrations when using complete time-series of monthly data, and the dominance of flow as a predictor of chlorophyll *a*. At the few sites where within-site multiple regression relationships included DIN or DRP with a positive coefficient, including the nutrient variable provided only a small increase in predictive ability, compared to using flow only (NSE improvement <0.2). Therefore, it seems unlikely that we can reasonably assess how changing sampling resolution affects the latter relationships. In addition, fortnightly datasets were available only for relatively short time-series at several stations (17 to 24 months). Consequently, we were only able to address the impact of fortnightly vs monthly monitoring frequency on the predictive ability of a predetermined regression model, not on the particular parameters retained in a stepwise regression.

8.2 Data

Fortnightly data were available from 13 sites (Table 8-1). A complete two-year time series was available for nine sites. Fortnightly monitoring commenced in August 2015 at the remaining four sites. Periphyton data and water quality were available up until April 2017, but the flow data extended to the end of 2016 only, giving 17 months of fortnightly data.

Table 8-1: List of sites with fortnightly data. Sites in order of their HRC number. FRE_eff is the mean annual frequency of effective flows. NL = nutrient limitation.

HRCn	Site abbreviation	shade (assessed in field)	Start	Finish	Flow group	FRE_eff	NL group
1	makakahi_doc		Aug-13	Sep-15	C	4.0	co
2	mangatainoka_putara	y	Aug-13	Sep-15	C	8.9	co
3	mangatainoka_lars		Aug-13	Sep-15	B	5.2	co
7	mangatainoka_huk		Aug-13	Sep-15	B	8.9	P
10	makakahi_ham	n	Aug-13	Sep-15	C	3.1	co-P
16	mangatainoka_scarb		Aug-13	Sep-15	B	4.4	P
18	mangatainoka_pahiatua		Aug-13	Sep-15	D	10.1	P
19	mangatainoka_sh2	n	Aug-13	Sep-15	B	4.3	P
23	manawatu_hop	n	Aug-13	Sep-15	A	8.2	none-N
59	waikawa_nmr	y	Aug-15	Apr-17	A	13.6	co
60	ohau_gladstone	y	Aug-15	Apr-17	D	6.8	co
61	ohau_sh1	y	Aug-15	Apr-17	A	12.4	co
62	ohau_haines		Aug-15	Apr-17	A	8.8	co

8.3 Methods

Our approach to comparing the performance of fortnightly vs. monthly data was based on within-site linear regressions as described in Section 7, using the shorter fortnightly datasets. There was no evidence of temporal autocorrelation in any of the final site models using the full monthly datasets. Therefore, this step was omitted.

Direct comparisons of significance levels of predictor variables between monthly and fortnightly datasets are not valid because F-tests and the resulting P values used to test the significance of terms in multiple regression are influenced by sample size. Therefore, to obtain datasets with the same number of datapoints for comparison between fortnightly and monthly data at each site, we first generated three datasets:

1. Month 1, the first sampling date in a calendar month;

2. Month 2, the second sampling date in a calendar month;
3. Fortnightly samples, twice as many samples as in the above datasets.

A linear regression was run on each of the three datasets for each site using log_{10} chlorophyll *a* as the response variable. Predictor variables were: days since an effective flow (Da_EFF), conductivity, log_{10}DIN and log_{10}DRP. Leave-one-out cross-validation was used to test model fit between predicted and observed values for each dataset at a site. For each datapoint, leave-one-out cross-validation runs the model without that point and predicts it based on a model generated by the other datapoints. For example, the following could be generated for a site with 24 fortnightly data points:

- Month 1 model on 12 datapoints, with observed values and leave-one-out predictions for each point;
- Month 2 model on 12 datapoints, with observed values and leave-one-out predictions for each point;
- Fortnightly model on 24 datapoints, with observed values and leave-one-out predictions for each point.

At each site, the Month 1 and Month 2 observed vs. predicted datasets were then joined back together to form a dataset with the same number of datapoints as the fortnightly dataset. The model fit between predicted and observed values was then compared between the joined monthly dataset and the fortnightly dataset.

8.4 Results

In almost all cases the models generated from the fortnightly data sets performed better than the models from the combined monthly data sets (Table 8-2, Figure 8-1). The exception was ohau_gladstone, where the combined monthly datasets predicted periphyton biomass more accurately than the fortnightly dataset.

The size of the datasets from the first and second sample in a month varied at most sites (Table 8-3) making direct comparison of model fit between the two monthly datasets difficult. However, at the three sites that had the same number of samples in the Month 1 and Month 2 datasets (mangatainoka_putara, mangatainoka_sh2 and ohau_haines) the model adjusted R^2 varied considerably between Month 1 and Month 2.

The same terms were significant in the regressions on the Month 1 and Month 2 datasets at mangatainoka_putara (no significant terms), and at ohau_haines (two significant terms) (Table 8-4). At mangatainoka_sh2 none of the predictors had significant relationships with log_{10}chlorophyll *a* in the Month 1 dataset, while in the Month 2 dataset both nutrients had significant relationships with log_{10}chlorophyll *a* (Table 8-4).

Table 8-2: Monthly versus fortnightly datasets: R^2 and NSE of cross-validation on observed vs predicted values for each site. The monthly dataset comprises combined values for the first and last samples in a month. The fortnightly dataset used all fortnightly samples. The total number of samples (N) was smaller than the expected 48 and 34 samples from the 24-month and 17-month datasets because of missing data (usually due to high flows). The last column shows the NSE for each site from the within-site multiple regressions using the 7-year monthly dataset and 6m-averaged DIN (from Table 7-3).

HRCn	Site	N	R^2		NSE		NSE, all monthly data
			monthly	fortnightly	monthly	fortnightly	
1	makakahi_doc	41	0	0.05	-0.56	-0.04	0.03
2	mangatainoka_putara	38	0.11	0.15	-0.02	0.12	0.11
3	mangatainoka_lars	39	0.52	0.48	0.44	0.46	0.47
7	mangatainoka_huk	39	0.08	0.21	-0.11	0.18	0.47
10	makakahi_ham	35	0	0	-0.52	-0.12	0.20
16	mangatainoka_scarb	39	0.21	0.25	0.09	0.22	0.18
19	mangatainoka_sh2	36	0.03	0.15	-0.36	0.12	0.50
23	manawatu_hop	35	0.28	0.40	0.02	0.39	0.45
59	waikawa_nmr	30	0.09	0.33	-0.57	0.29	0.21
60	ohau_gladstone	29	0.34	0.17	0.24	0.06	0.20
61	ohau_sh1	30	0.03	0.16	-0.67	0.10	0.38
62	ohau_haines	26	0.36	0.47	-0.04	0.43	0.28

Table 8-3: Comparison of adjusted R^2 of the regression models generated from the Month 1 and Month 2 datasets at each of the sites with fortnightly data. The model residual df (degrees of freedom) is number of samples minus number of terms in the model (the model df, 5).

HRCn	Site	Model adjusted R^2		Model residual df	
		1st in Month	2nd in Month	1st in Month	2nd in Month
1	makakahi_doc	-0.03	0.21	16	15
2	mangatainoka_putara	0.05	0.39	14	14
3	mangatainoka_lars	0.64	0.77	16	13
7	mangatainoka_huk	0.20	0.29	15	14
10	makakahi_ham	-0.14	0.04	14	11
16	mangatainoka_scarb	0.39	0.36	15	14
19	mangatainoka_sh2	-0.12	0.34	13	13
23	manawatu_hop	0.28	0.63	14	11
59	waikawa_nmr	0.34	0.40	11	9
60	ohau_sh1	0.27	0.11	11	9
61	ohau_gladstone	0.28	0.65	10	9
62	ohau_haines	0.46	0.54	8	8

Table 8-4: P values for terms from regressions on Month 1 and Month 2 datasets for the three sites with the same number of datapoints in the two datasets. Significant terms (P< 0.1) are shown in bold type.

	Mangatainoka_putara		Mangatainoka_sh2		Ohau_haines	
	Month 1	Month 2	Month 1	Month 2	Month 1	Month 2
Conductivity	0.41	0.53	0.76	0.10	0.14	0.61
log_{10}DIN	0.51	0.63	0.35	**0.06**	0.32	0.35
log_{10}DRP	0.56	0.44	0.71	**0.06**	**0.02**	**0.03**
daEFF	0.11	0.10	0.75	0.69	**0.08**	**0.09**

8.5 Discussion

The outcome of this analysis was that sampling at fortnightly intervals resulted in stronger relationships between chlorophyll *a* and the selected environmental variables than sampling at monthly intervals at all but one of the sites (ohau_gladstone). The comparison method took into account the difference in sample numbers between the sampling frequencies. All the relationships were tested with the same set of predictors.

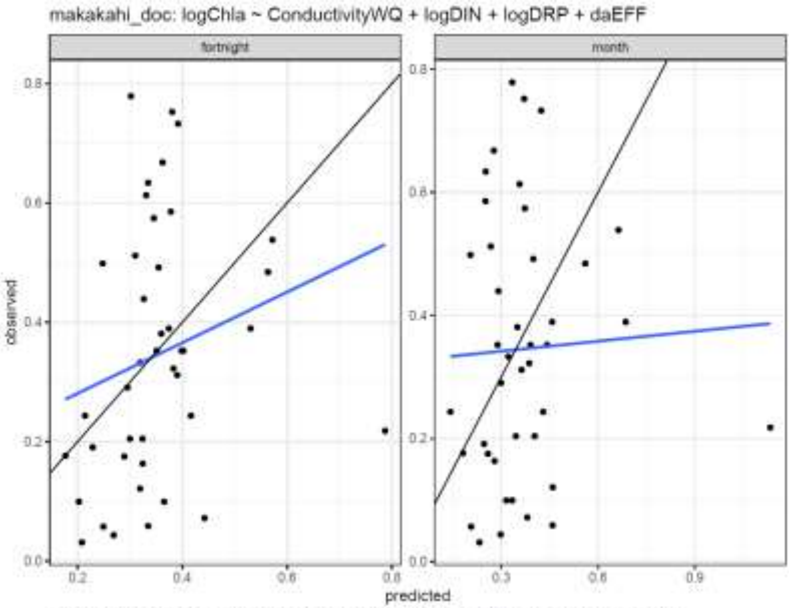
Sites other than ohau_gladstone fell into three groups in terms of their responses compared to the results of the multiple regression analyses using the entire dataset of monthly data (i.e., the complete 7-year dataset; see Section 7 and compare Table 8-2 with Table 7-4). Note that the results for the entire dataset are not strictly comparable to the present results as the DIN variables used differed. However, at most sites, correlations between chlorophyll *a* and the different DIN variables were similar (Appendix G).

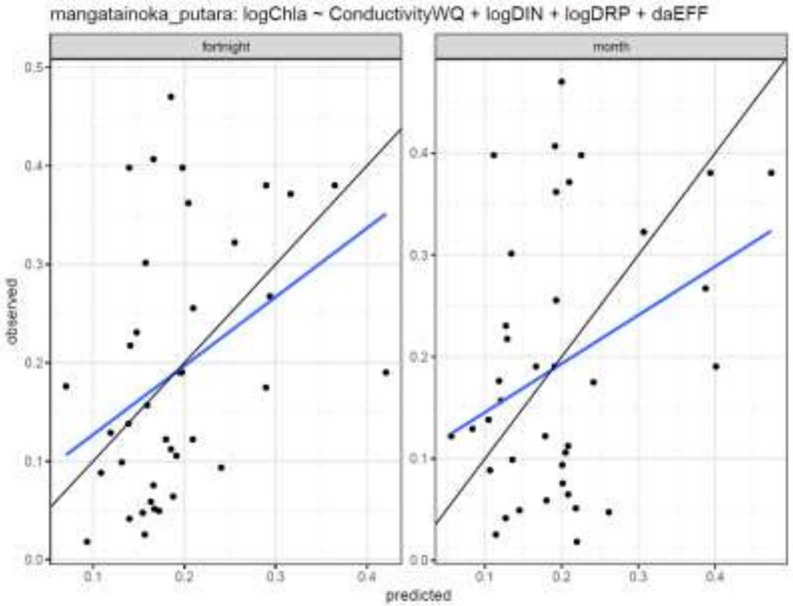
The first group comprised four sites at which relationships of comparable strength were found in all of the datasets. This applied to:

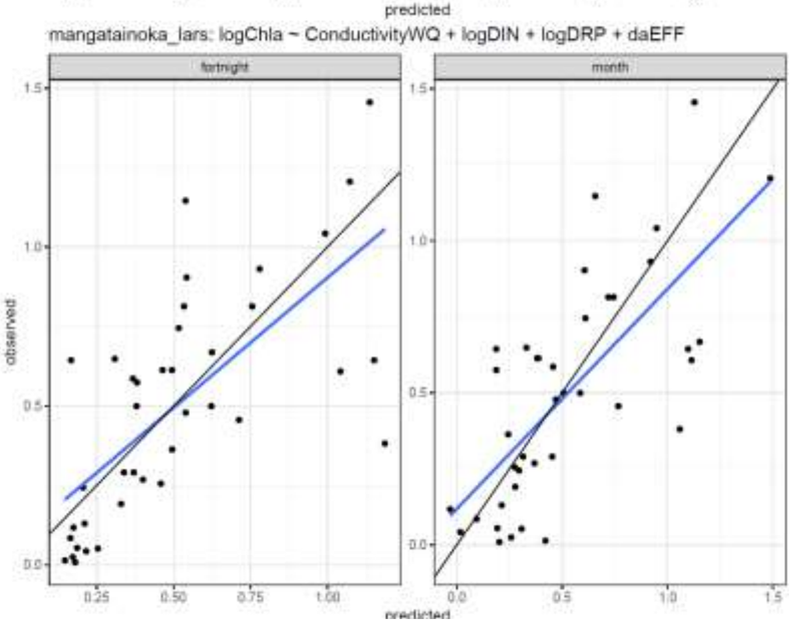
- makakahi_doc and mangatainoka_putara (effectively no predictive ability in any of the relationships) (NSE close to zero, or negative);
- mangatainoka_lars (strongest predictive ability in models from both the fortnightly and monthly datasets (NSE = 0.44, 0.46), and strong within-site multiple regression relationship, largely driven by flow (Da_EFF, NSE = 0.41);
- mangatainoka_scarb (generally weak predictive ability, including in the longer monthly dataset (maximum NSE 0.22). Note that sampling started at this site in August 2013.

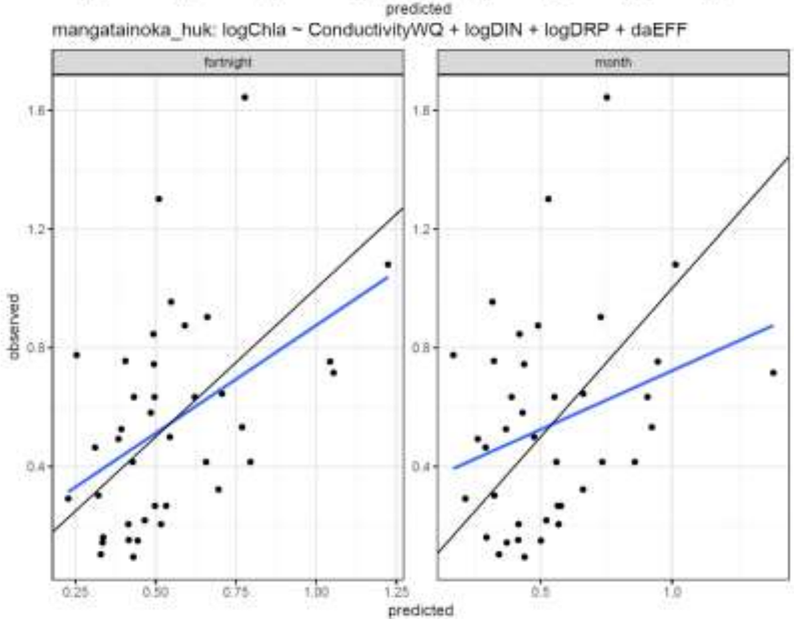
Plots on the following three pages:

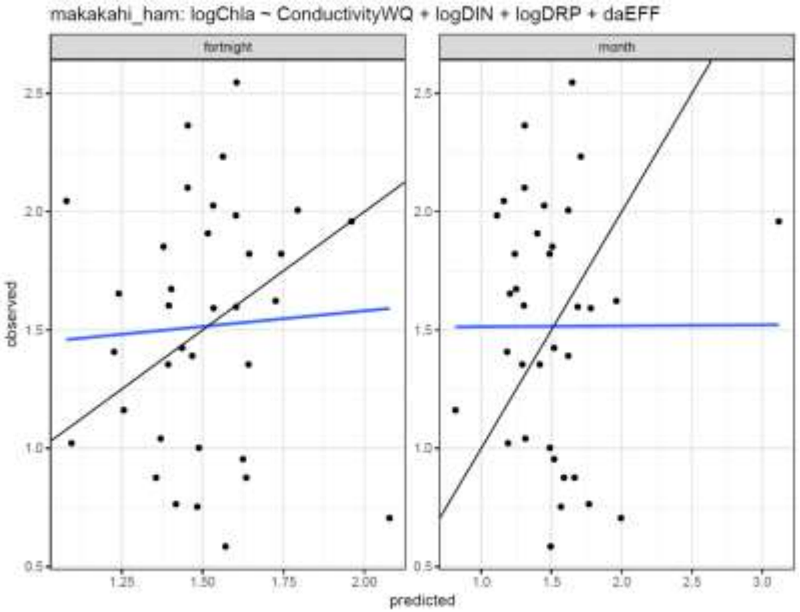
Figure 8-1: Fortnightly vs. monthly datasets: comparison of observed vs predicted relationships at 12 sites.. The black line is 1:1 and the blue line is a fitted regression between the observed and predicted points. The fortnightly data set is based on a regression using fortnightly sampling data. The monthly dataset is based on predictions from two separate datasets, the first and second sample in any given month. Predictions and observations from the two monthly datasets are combined to allow comparison of model fit with the fortnightly data set assuming the same number of sampling points..

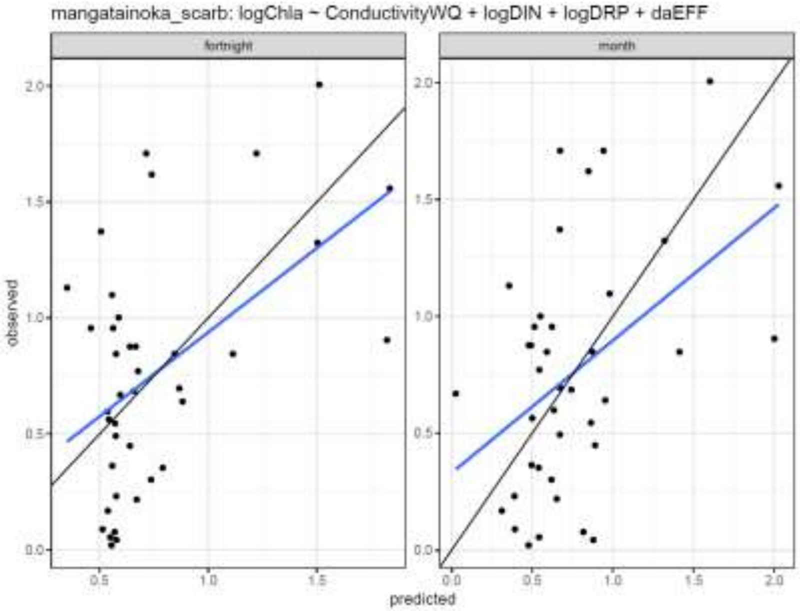

makakahi_doc: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted

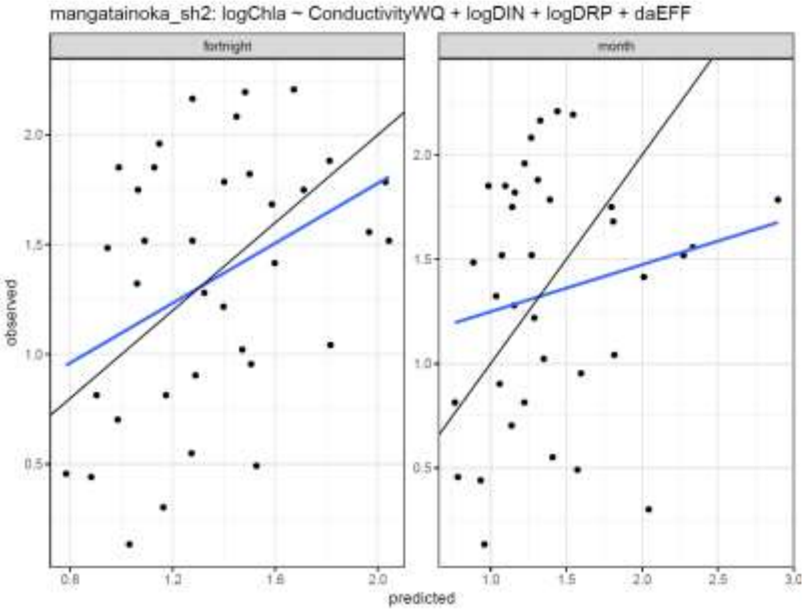

mangatainoka_putara: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted

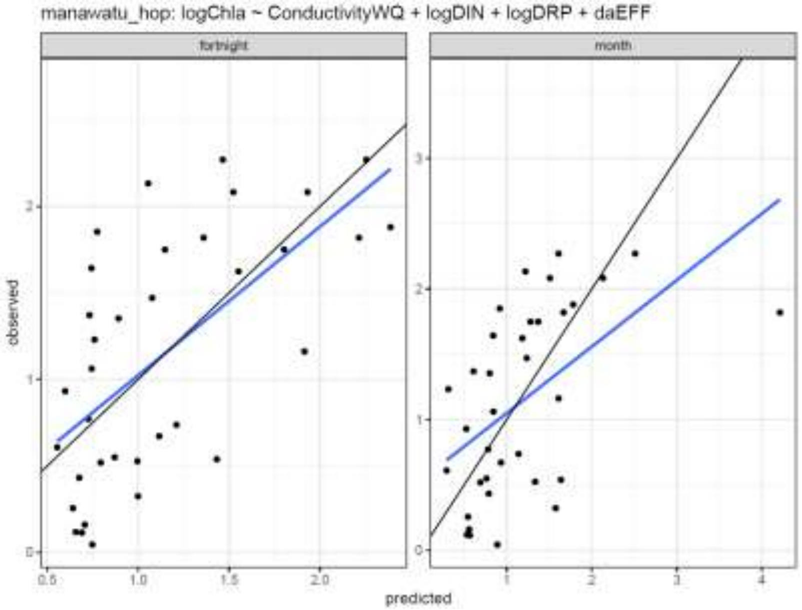

mangatainoka_lars: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed

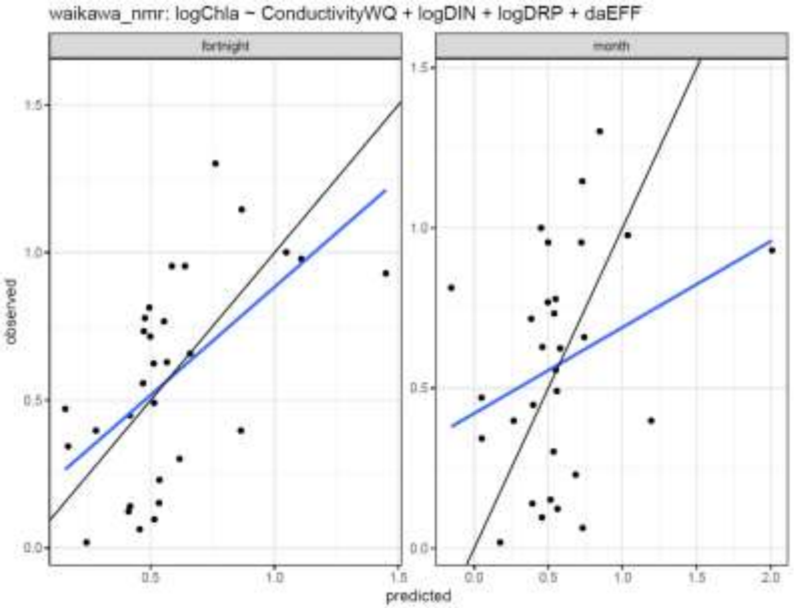

mangatainoka_huk: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted

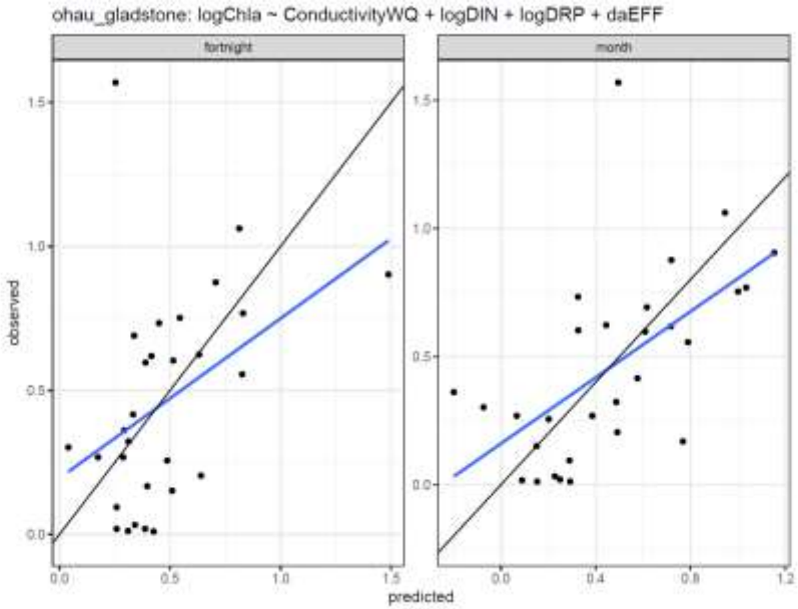

makakahi_ham: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted

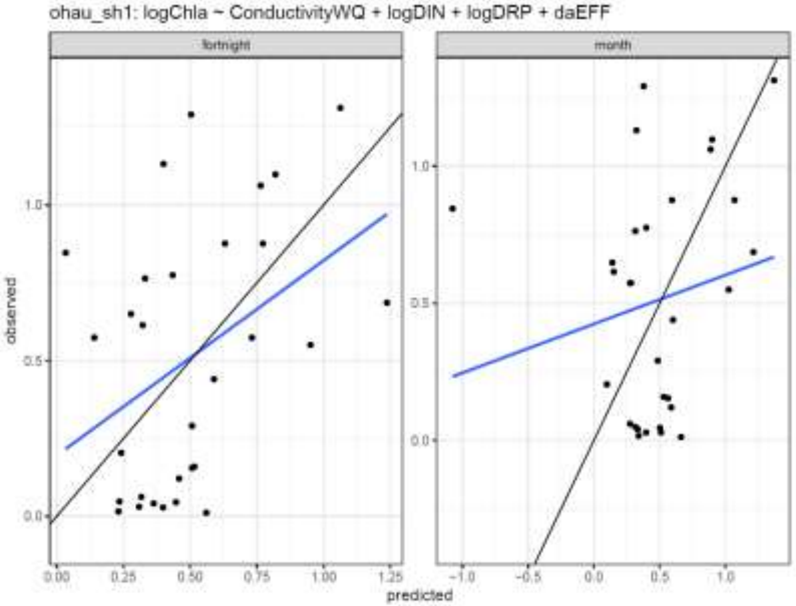

mangatainoka_scarb: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted

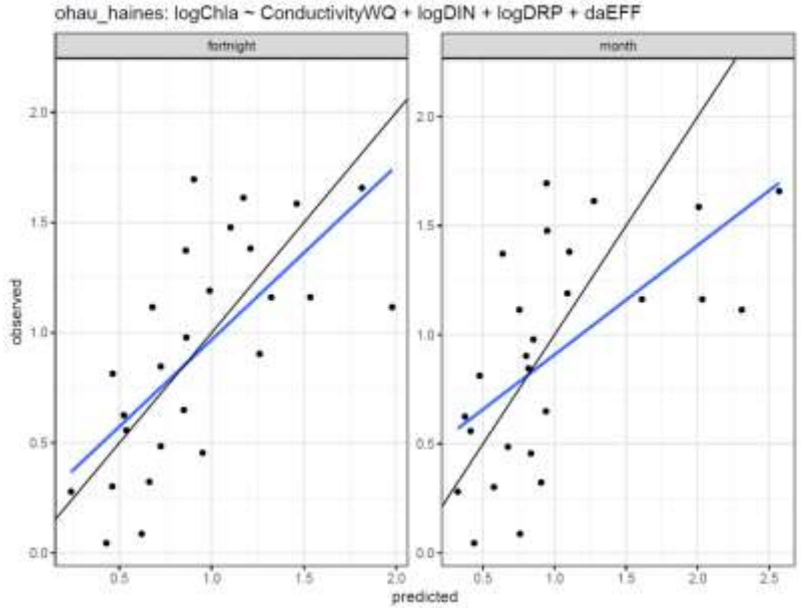

mangatainoka_sh2: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted


manawatu_hop: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted


waikawa_nmr: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted


ohau_gladstone: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted


ohau_sh1: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted


ohau_haines: logChla ~ ConductivityWQ + logDIN + logDRP + daEFF
fortnight
month
observed
predicted

The second group was made up of four sites with at which the fortnightly model outperformed the monthly model, but within-site models from the longer-term monthly dataset showed better predictive ability than either the short-term monthly or fortnightly models (assessed using NSE)). The sites were mangatainoka_huk, mangatainoka_sh2, makakahi_ham and ohau_sh1 (Table 8-2). For example, using the 7-year monthly dataset, within-site regression models at mangatainoka_sh2 showed much better predictive ability than either the short-term monthly data (no predictive ability) or fortnightly data (only weak predictive ability) (Table 8-2) and were strongly influenced by the positive effect of accrual time (Da_EFF, NSE = 0.47, see Table 7-4).

The third group included three sites at which the fortnightly model outperformed the monthly model, and had similar predictive ability to models produced using multiple regression on the whole monthly dataset (manawatu_hop, waikawa_nmr, ohau_haines).

One reason for poorer performance of the short-term monthly datasets compared to the fortnightly datasets was illustrated for mangatainoka_sh2, in Table 8-4. Table 8-4 showed that, at mangatainoka_sh2, one monthly set of chlorophyll *a* samples was significantly related to both DIN and DRP, while the alternate set was not. Since the samples were effectively collected randomly in relation to environmental conditions, there is no logical explanation for the difference except that the dataset was too small (18 monthly samples) to guarantee that the typical range of chlorophyll *a* and environmental conditions was encompassed.

While this test cannot say anything about the strength of effect of individual variables, the results suggest that, at some sites (i.e., those in the third group above), data collected at more frequent intervals than monthly could be useful because they may allow identification of relationships using a shorter time-series of data. From Table 8-1 we see that all three sites in the third group were classed as "flow-sensitive" (Groups A and B in Table 3-1) and all four had relatively high flood frequencies (at least 8 events per year exceeding the effective flow, Table 8-1). Further analysis is needed to demonstrate that this finding holds across the region, or further afield.

However, at another flow-sensitive, high flood frequency site in Table 8-1 (ohau_sh1), the fortnightly model outperformed the monthly model, but the model derived from the long-term monthly dataset was much stronger, and driven largely by flow (Da_EFF, NSE = 0.33). Therefore, identifying sites at which fortnightly sampling would be helpful is not straightforward.

Sample collection at more frequent intervals than monthly can be useful for other reasons. Differences in periphyton accrual rates between sites can be assessed more easily using data collected at closer intervals than a month (Biggs and Close 1987, Biggs and Stockseth 1995, Davie et al. 2012). Information of accrual rates at particular times of year could be used, for example, to estimate how long since a flood would be required for chlorophyll *a* to exceed a guideline or objective. It may be possible to estimate an average accrual rate using the long-term dataset (from the slope of the relationship between chlorophyll *a* and accrual period, see Section 3.2), but ability to determine accrual rates over individual growth periods could be informative for assessing the effects of season on accrual rates. Fortnightly datasets could also be used to test mechanistic models of periphyton growth.

9 Relationships between chlorophyll *a* and percentage cover

Key messages

- In addition to data on periphyton chlorophyll *a*, data on periphyton cover were available from all sites, in six categories (bare rock, film, sludge, mats, slimy green filaments, other (coarse) filaments). For the analysis sludge and mats were combined into "Mats", and slimy green and other (coarse) filaments into "Fils".
- Correlations between chlorophyll *a* and metrics of percentage cover were investigated using between-site and within-site approaches. The purpose of the analysis was to see whether it is possible to make robust conversions from visual estimates to chlorophyll *a*. If that proved to be the case, then it can be inferred that management of the environmental factors that affect chlorophyll *a* will apply to visual cover by periphyton in an equivalent way.
- We explored relationships between mean and maximum chlorophyll *a* and cover, between sites and within sites, using a range of cover metrics (in particular, weighted composite cover (WCC) and the combination of Film, Mats and Fils in a multiple regression).
- For the between-site analysis, predictive ability of the chlorophyll *a* and percentage cover relationships with mean chlorophyll *a* varied across years and was often poor (NSE < 0.3), with the exception of later years (2013–14, 2014–15). Relationships were broadly equivalent in performance between WCC, Mats or Fils, and in general were weaker for annual maxima then annual mean cover estimates.
- Within sites, the multiple regression using Film, Mats and Fils produced the strongest relationships with chlorophyll *a*. 44% of sites had RSE > 0.55. All the sites with strong predictive ability were in the wider Manawatu catchment, or in the Ohau catchment, and did not include headwater sites.
- We concluded that use of a single region-wide relationship to predict chlorophyll *a* from cover is unlikely to be robust for the Horizons dataset. Only within the mid- to lower Manawatu River and Ohau River did strong chlorophyll *a* – cover relationships suggest that the drivers of chlorophyll *a* are likely to have corresponding effects on percentage cover by periphyton.

9.1 Introduction

All of the analyses in previous sections have focussed on chlorophyll *a* as the main dependent variable. Chlorophyll *a* is a useful composite measure of standing crop of all algae[6], and is required to

[6] Note also that chlorophyll *a* content varies with both species composition (Kasprzak et al. 2008) and with environmental conditions such as nutrient concentrations and light (Baulch et al. 2009). Therefore, while chlorophyll *a* generally represents the total amount of algae in a sample, there are still issues of consistency over time and space.

assess compliance with the NPS-FM periphyton attribute. However, chlorophyll *a* does not provide direct information on the visual impact of periphyton at a site. The Horizons One Plan includes targets for percentage cover by mats and filaments; therefore, the estimates of percentage cover by different categories of periphyton carried out as part of the Horizons periphyton monitoring programme allow assessment of all sites against those targets. In addition to providing data consistent with current resource management policy in the Horizons region (and with earlier national recreational policy that focussed on periphyton cover; Biggs 2000b), another potential use of periphyton cover data is to allow estimation of chlorophyll *a*. In the narrative around the periphyton attribute in the NPS-FM (Snelder et al. 2013), the following statement allows for substitution of visual estimates in some cases:

> "Although the proposed objective is specified in terms of chlorophyll *a*, a significant proportion of monitoring could be carried out for low risk systems using the quicker and less costly visual estimate methodologies. Recently developed protocols can be used to estimate chlorophyll *a* from cover data (Kilroy et al. 2013). Should monitoring based on visual cover estimates indicate that a site is approaching the relevant periphyton abundance threshold, monitoring could then be upgraded to include measurement of chlorophyll *a*."

This advice is now formally included in MfE guidance on the application of the NPS-FM (MfE 2015). Furthermore, if cover and chlorophyll *a* are closely correlated then it can be inferred that management of the environmental factors that affect chlorophyll *a* will apply to visual cover by periphyton in an eqivalent way.

In this section correlations between chlorophyll *a* and various metrics of percentage cover are explored, first between sites, then within sites. This section addresses Objective 1 parts b1 and c1 in the contract (see Appendix A).

9.2 Between site relationships

9.2.1 Methods

Cover variables

During the surveys periphyton cover at each site on each occasion was recorded as mean estimated percent cover in six categories (bare rock, film, sludge, mats, slimy green filaments, other (coarse) filaments). Hereafter we refer to the combination of sludge and mats as Mats, and the combination of slimy green and other (coarse) filaments as Fils.

With multiple variables, there are several options for deriving variables for direct comparison with chlorophyll *a*. Examples are:

1. total percentage cover of all periphyton (e.g., sum of all categories except bare rock);
2. percentage cover of all algae with potentially high chlorophyll *a* (i.e., Mats and Fils only);
3. total percentage cover by Mats and Filaments converted to a weighted composite cover (WCC) as suggested by Matheson et al. (2012). WCC is calculated as (percentage cover by mats)/2 + percentage cover by filaments, acknowledging the fact that thresholds for nuisance levels of filamentous algae are typically set at 30%, while thresholds for mats are higher, at 60%.;

4. percentage cover by either Mats or Fils separately;
5. a combination of Mats, Fils and Films combined into a multiple regression;
6. percentage cover converted to a derived chlorophyll *a* using conversion factors (e.g., Kilroy et al. 2013).

For the analyses, chlorophyll *a* was log-transformed and Mats and Fils were fourth-root transformed. Film data were normally distributed and were used untransformed.

Trial regressions, all data

We first ran trial linear regressions between chlorophyll *a* and cover calculated using the first four options above (individual values on each survey at all sites). Total percentage cover including Film (variable 1) was not linearly related to chlorophyll *a* and was not considered further. The strongest relationship with a single variable (highest R^2 and steepest slope) was between chlorophyll *a* and WCC (variable 3), followed by between chlorophyll *a* and mats + filaments (variable 2). Both these relationships explained 56% of the variance in chlorophyll *a* (individual measurements), with the relationships:

$$Log_{10}chla = 0.354 + 0.558 \text{ (frthrt WCC) } (R^2 = 0.56, P < 0.001, S.E.M. = 0.445, n = 4806)$$

$$Log_{10}chla = 0.351 + 0.517 \text{ (frthrt Mats+Fils) } (R^2 = 0.56, P < 0.001, S.E.M. = 0.446, n = 4806)$$

Equivalent relationships with %cover mats and % cover by filaments (variable 4) were:

$$Log_{10}chla = 0.535 + 0.514 \text{ (frthrt Fils) } (R^2 = 0.43, P < 0.001, S.E.M. = 0.506, n = 4806)$$

$$Log_{10}chla = 0.578 + 0.456 \text{ (frthrt Mats) } (R^2 = 0.34, P < 0.001, S.E.M. = 0.546, n = 4806)$$

Multiple regression across all sites including Mats, Filaments and Film as separate variables (variable 5) explained 60% of the variance in chlorophyll *a*, with the relationship

$$Log_{10}chla = 0.107 + 0.005 \text{ (Film) } + 0.389 \text{ (frthrt Fils) } + 0.300 \text{ (frthrt Mats)}$$
$$(R^2 = 0.60, P < 0.001, S.E.M. = 0.427, n = 4806)$$

Relationships in different periods

On the basis of these trials we looked further at between-site relationships between chlorophyll *a* and variable 3 (WCC) and variable 5 (combined cover measures). Simple regression was run between mean or maximum WCC and mean or maximum (respectively) chlorophyll *a* in each year of the dataset, followed by stepwise multiple regression including Film, Mats and Film. Performance of the models was evaluated using leave-one-out cross validation (see Section 6.2.1).

9.2.2 Results

Regressions run on mean data in each hydrological year (all sites) yielded significant relationships between chlorophyll *a* and WCC in all years ($P < 0.001$ in all cases), with some variability across years (Table 9-1). The regressions for maximum values were weaker and had poor predictive power (data not shown).

Using multiple regression including Film, Mats and Fils (mean annual values in each year at each site) as predictor variables gave results similar to those using WCC in terms of R^2 values (Table 9-2). Film was included in the model in only one year, 2011-12. In this year, NSE < 0 indicated that the relationship across all sites had no useful predictive power (worse than just using the mean across all sites). The relationship in 2012-13 similarly had effectively no predictive ability.

For both WCC and the combination of Film, Mats and Fils in a multiple regression, the highest variance in mean annual chlorophyll *a* was explained, and the best predictive power obtained in 2013-14 (70% by mean WCC and 67% by the combination of Film, Mats and Fils). The slopes of the relationships between mean chlorophyll *a* and mean WCC differed among years, indicating shifts over time in the chlorophyll *a* equivalent of WCC, on average (Table 9-1). In the combined regression, all three variables were selected as predictors in 2011-12, Mats only in 2010-11, and Mats and Fils in the remaining five of the seven years (Table 9-2).

Table 9-1: Summary result (regression and cross-validation) of linear regressions of mean annual chlorophyll *a* against mean annual WCC. Chlorophyll *a* was log-transformed, WCC fourth-root transformed.

		Regression			Cross validation		
Period	***n***	**R^2**	**P**	**slope**	**R^2**	**NSE**	**RMSD**
2009-10	48	0.45	**<0.001**	0.531	0.4	-0.18	0.342
2010-11	53	0.55	**<0.001**	0.739	0.52	0.14	0.346
2011-12	53	0.47	**<0.001**	0.639	0.43	-0.2	0.332
2012-13	56	0.51	**<0.001**	0.665	0.48	0.02	0.311
2013-14	61	0.70	**<0.001**	0.773	0.69	0.57	0.266
2014-15	61	0.60	**<0.001**	0.619	0.58	0.3	0.309
2015-16	60	0.56	**<0.001**	0.590	0.53	0.18	0.285

Table 9-2: Summary results (regression and cross-validation) of multiple regression between mean annual chlorophyll *a* and mean Film, Mats and Fils. Chlorophyll *a* was log-transformed, Mats and Fils fourth-root transformed. Data on Film were untransformed.

		Regression		Cross validation			Terms included
Period	***n***	**R^2**	**P**	**R^2**	**NSE**	**RMSD**	
2009-10	48	0.57	**<0.001**	0.51	0.21	0.31	Mats, Fils
2010-11	53	0.64	**<0.001**	0.61	0.41	0.31	Mats
2011-12	53	0.51	**<0.001**	0.39	-0.15	0.35	Mats, Fils, Film
2012-13	56	0.49	**<0.001**	0.40	-0.14	0.34	Mats, Fils
2013-14	61	0.67	**<0.001**	0.64	0.47	0.29	Mats, Fils
2014-15	61	0.61	**<0.001**	0.57	0.31	0.31	Mats, Fils
2015-16	60	0.62	**<0.001**	0.58	0.33	0.27	Mats, Fils

9.3 Within site relationships

9.3.1 Methods

Regressions were run at each site between individual chlorophyll *a* observations and cover observations on the same day, following the methodology used in the within-sites analyses between chlorophyll *a* and multiple variables (see Section 7.3), omitting the autocorrelation step, which was not appropriate in this case. In summary, at each site a series of linear regression analyses was run between chlorophyll *a* and percentage cover variables. The cover variables were Mats+Fils, WCC, Mats, Fils, and [Film, Mats and Fils] (the latter in a multiple regression) (i.e., variables 2 to 5 in Section 9.2.1 above). The performance of the models for predicting at new times within a site was evaluated using leave-one-out cross-validation tests (see Section 6.2.1 for explanation). Output from cross validation tests was compared across the variables at each site.

9.3.2 Results

With few exceptions, the multiple regression using Film, Mats and Fils produced the strongest relationships between chlorophyll *a* and cover variables within a site (Table 9-3). The exceptions were mangatainoka_lars, mangatainoka_huk, maingatainoka_us_tir, porewa_us_hun, although differences were small. Across all site, the mean R^2 of the multiple regression models was 0.47, with a mean RMSD of 0.36. WCC and Mats + Fils had corresponding mean R^2 of 0.35 and RMSD of 0.40.

The within-site relationships showed a geographical pattern. The multiple regression relationship (chlorophyll *a* vs. Film, Mats and Fils), and also the simpler regression relationships indicated stronger correspondence between cover and chlorophyll *a* at sites in the Manawatu catchment (note concentration of red-highlighted values up to HRC site no 37 in Table 9-3), on average, than in the Rangitikei or Whangaehu catchments (Table 9-4). Relationships at the three sites in the Ohau catchment were also relatively strong. There was variability across the 36 sites within the Manawatu catchment, with weak correspondence at a few sites, notably the two headwater sites (makakahi_doc and mangatainoka_putara) and also at makuri_tuscan (Table 9-3).

Table 9-3: Summary results of leave-one-out cross-validation tests on relationships between chlorophyll *a* and percentage cover at each site. Chlorophyll *a* was log-transformed, Mats and Fils were fourth-root transformed, Film was untransformed. R^2 is the R^2 of the regression of observed vs. predicted from the leave-one-out procedure. RMSD is the root-mean squared deviation, which is the average deviation of the predictions from the observed values in the same units as the predicted variable. * indicates sites at which the combination of Film, Mats and Fils did not produce the strongest model (highest R^{String} and lowest RMSD). Under Film, Mats, Fils, leave-one-out R^2 values >0.50 are highlighted in bold red and values >0.40 ≤0.50 in bold black. NSE values not shown, but R^2 < 0.1 generally corresponds to NSE < 0.1, i.e., the predictive ability of the relationships is very poor or nil.

			Mats		Fils		Fils + Mats		WCC		Film, Mats, Fils					
HRC	Site	n	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	Film	Fils	Mats	NL
1	makakahi_doc	41	0.09	0.29	1.00	0.31	0.09	0.29	0.04	0.30	0.25	0.26	1		1	co
2	mangatainoka_putara	93	0.02	0.24	0.02	0.23	0.04	0.23	0.03	0.23	0.09	0.23	1		1	co
3	mangatainoka_lars*	42	0.18	0.38	0.51	0.29	0.51	0.29	0.44	0.31	**0.48**	0.30	1	1		co
4	tamaki_res	99	0.09	0.36	0.14	0.35	0.18	0.34	0.22	0.33	**0.55**	0.25	1	1	1	
5	mangatera_us_dan	99	0.08	0.56	0.24	0.51	0.23	0.51	0.31	0.49	0.39	0.47	1	1	1	
6	mangatera_ds_dan	98	0.30	0.61	0.31	0.61	0.37	0.58	0.42	0.56	**0.59**	0.47	1	1	1	
7	mangatainoka_huk*	41	0.39	0.37	0.26	0.41	0.45	0.36	0.41	0.37	**0.44**	0.36		1	1	P
8	kumeti_tr	99	0.31	0.38	0.13	0.43	0.32	0.38	0.37	0.36	**0.63**	0.28	1	1	1	P
9	manawatu_weber	91	0.30	0.60	0.38	0.57	0.44	0.54	0.41	0.56	**0.57**	0.47	1	1	1	co -none
10	makakahi_ham	91	0.11	0.46	0.27	0.42	0.30	0.41	0.28	0.42	0.42	0.36		1	1	co-P
11	oroua_apiti	94	0.57	0.26	0.31	0.33	0.58	0.26	0.55	0.27	**0.67**	0.23	1	1	1	co
12	tamaki_ste	93	0.36	0.43	0.04	0.52	0.43	0.40	0.44	0.40	**0.56**	0.36	1	1	1	co-P
13	oruakeretaki_sh2	96	0.46	0.43	0.36	0.47	0.54	0.40	0.56	0.39	**0.70**	0.32	1	1	1	none-P
14	makuri_tuscan	92	0.01	0.62	0.00	0.62	0.01	0.62	0.03	0.61	0.08	0.59		1	1	P
15	pohangina_pir	94	0.48	0.27	0.34	0.31	0.52	0.26	0.55	0.25	**0.59**	0.25	1	1	1	
16	mangatainoka_scarb	38	0.42	0.42	0.26	0.47	0.48	0.40	0.49	0.39	**0.54**	0.38		1	1	P
17	tiraumea_nga	72	0.13	0.57	0.21	0.54	0.21	0.54	0.21	0.54	**0.57**	0.40	1	1	1	P
18	mangatainoka_pahiatua	37	0.11	0.49	0.19	0.47	0.28	0.44	0.28	0.44	0.38	0.42	1	1	1	P
19	mangatainoka_sh2	95	0.45	0.47	0.43	0.48	0.63	0.39	0.66	0.37	**0.67**	0.36	1	1	1	P
20	mangatainoka_ds_db	97	0.47	0.43	0.37	0.47	0.67	0.34	0.59	0.38	**0.73**	0.30	1	1	1	P
21	mangatainoka_us_pah	97	0.43	0.47	0.38	0.49	0.57	0.41	0.59	0.40	**0.59**	0.38	1	1	1	P

			Mats		Fils		Fils + Mats		WCC		Film, Mats, Fils					
HRC	Site	n	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	Film	Fils	Mats	NL
22	mangatainoka_ds_pah	89	0.43	0.44	0.44	0.44	0.60	0.37	0.60	0.37	**0.69**	0.32	1	1	1	P
23	manawatu_hop	91	0.32	0.65	0.54	0.54	0.60	0.50	0.63	0.48	**0.72**	0.41	1	1	1	none-N
24	mangatainoka_us_tir*	69	0.46	0.44	0.53	0.41	0.67	0.34	0.70	0.33	**0.70**	0.33	1	1	1	P
26	mangapapa_troup	100	0.23	0.42	0.33	0.39	0.44	0.36	0.45	0.36	**0.65**	0.29	1	1	1	co-P
27	pohangina_mais	98	0.18	0.40	0.29	0.37	0.38	0.35	0.39	0.34	**0.43**	0.34	1	1	1	co
28	manawatu_ug	90	0.33	0.54	0.56	0.43	0.57	0.43	0.58	0.42	**0.74**	0.35	1	1	1	P-co
29	oroua_almadale	91	0.27	0.35	0.21	0.36	0.39	0.32	0.39	0.32	**0.47**	0.30	1	1	1	co
30	oroua_us_fei	83	0.19	0.50	0.52	0.39	0.53	0.39	0.53	0.38	**0.62**	0.34	1	1	1	
31	oroua_ds_fei	90	0.59	0.47	0.62	0.45	0.70	0.40	0.73	0.39	**0.79**	0.33	1	1	1	
32	oroua_awahuri	90	0.41	0.53	0.57	0.46	0.62	0.43	0.63	0.42	**0.69**	0.37	1	1	1	
33	manawatu_tc	86	0.08	0.55	0.53	0.39	0.53	0.39	0.52	0.40	**0.62**	0.36	1	1		co
34	manawatu_us_pncc	92	0.32	0.57	0.65	0.41	0.67	0.40	0.67	0.40	**0.74**	0.36	1	1	1	co-P
35	manawatu_ds_pncc	89	0.48	0.62	0.62	0.53	0.71	0.46	0.73	0.44	**0.75**	0.40	1	1	1	none-P
36	manawatu_opik	78	0.30	0.66	0.61	0.49	0.65	0.47	0.69	0.44	**0.80**	0.35	1	1	1	none-P
37	tokomaru_hb	100	0.33	0.41	0.33	0.41	0.46	0.37	0.43	0.38	**0.56**	0.33	1	1	1	co
38	rangitikei_puk	87	0.03	0.38	0.33	0.31	0.33	0.31	0.28	0.32	**0.50**	0.27	1	1	1	co
39	moawhango_waiouru	60	0.00	0.47	0.00	0.47	0.00	0.47	0.00	0.47	0.00	0.47				co
40	rangitikei_man	91	0.05	0.51	0.23	0.46	0.23	0.46	0.24	0.46	0.28	0.44		1	1	co
41	porewa_us_hun*	51	0.02	0.45	0.00	0.46	0.03	0.45	0.01	0.45	0.02	0.36			1	
42	porewa_ds_hun	48	0.13	0.48	0.00	0.53	0.13	0.48	0.01	0.52	0.23	0.33			1	
43	rangitikei_one	87	0.11	0.52	0.30	0.46	0.30	0.46	0.28	0.47	**0.45**	0.42	1	1	1	co
44	rangitikei_mk	83	0.25	0.58	0.33	0.55	0.39	0.52	0.41	0.51	**0.49**	0.47		1	1	co
45	mangawhero_doc	100	0.00	0.35	0.09	0.33	0.09	0.33	0.00	0.35	0.17	0.28		1		N
46	makotuku_sh49	100	0.10	0.45	0.07	0.46	0.14	0.44	0.17	0.43	0.21	0.43	1	1	1	co-N
47	mangawhero_us_oha	100	0.18	0.45	0.22	0.44	0.28	0.42	0.20	0.45	**0.46**	0.36	1	1	1	co
48	mangawhero_ds_oha	97	0.15	0.50	0.22	0.48	0.25	0.47	0.28	0.46	0.36	0.43	1	1	1	N-co

			Mats		Fils		Fils + Mats		WCC		Film, Mats, Fils					
HRC	Site	n	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	Film	Fils	Mats	NL
49	makotuku_rae	85	0.09	0.43	0.22	0.40	0.22	0.40	0.15	0.42	0.31	0.40	1	1	1	co-P
50	mangawhero_pakihi	87	0.07	0.47	0.18	0.44	0.18	0.44	0.15	0.45	0.21	0.43	1	1	1	co-N
51	mangatepopo_gi	76	0.14	0.31	0.12	0.32	0.15	0.31	0.15	0.31	0.21	0.30		1	1	
52	whanganui_ds_gen	77	0.04	0.34	0.01	0.35	0.04	0.34	0.03	0.35	0.15	0.29	1		1	N
53	whakapapa_ds_gen	72	0.27	0.35	0.21	0.37	0.27	0.35	0.27	0.35	**0.50**	0.29	1	1	1	N
54	waitangi_us_wai	100	0.01	0.50	0.01	0.50	0.01	0.50	0.01	0.50	0.18	0.46	1	1	1	
55	waitangi_ds_wai	100	0.00	0.49	0.25	0.42	0.25	0.42	0.28	0.41	0.28	0.41		1	1	
56	tokiahuru_kar	69	0.14	0.43	0.01	0.46	0.17	0.42	0.13	0.43	0.31	0.39	1		1	
57	makotuku_us_rae	76	0.11	0.42	0.04	0.43	0.13	0.41	0.10	0.42	0.25	0.37		1	1	P-co
58	makotuku_ds_rae	90	0.03	0.46	0.13	0.43	0.14	0.43	0.12	0.44	0.24	0.41		1	1	
59	waikawa_nmr	99	0.16	0.38	0.31	0.34	0.33	0.34	0.37	0.33	**0.43**	0.31	1	1	1	co
60	ohau_gladstone	99	0.11	0.34	0.14	0.33	0.16	0.33	0.15	0.33	0.31	0.30	1	1		co
61	ohau_sh1	97	0.43	0.42	0.31	0.46	0.45	0.41	0.51	0.39	**0.57**	0.36	1	1	1	co
62	ohau_haines	43	0.15	0.49	0.49	0.38	0.49	0.38	0.41	0.41	**0.64**	0.32	1	1		co

Table 9-4: Mean R^2 and RMSD from leave-one-out cross-validation tests on chlorophyll *a* vs. periphyton cover relationships, by catchment. *n* = number of sites.

		chlorophyll *a* vs. Mats + Fils		chlorophyll *a* vs. WCC		chlorophyll *a* vs. Film, Mats, Fils (best combination)	
Catchment	*n*	R^2	RMSD	R^2	RMSD	R^2	RMSD
Manawatu	36	0.45	0.40	0.46	0.40	0.57	0.35
Rangitikei	7	0.20	0.45	0.18	0.46	0.28	0.39
Whangaehu	11	0.17	0.43	0.14	0.43	0.27	0.40
Ohau	3	0.37	0.37	0.36	0.38	0.51	0.33

9.4 Discussion

Inconsistent relationships between chlorophyll *a* and the cover variables across all sites in different periods, and inconsistencies in relationships at individual sites indicated that use of cover data for river management based on standards specified as chlorophyll *a* is not straightforward. There is clearly some correspondence between chlorophyll *a* and cover variables, but *at the scale of the whole region and over time*, use of a single relationship to predict chlorophyll *a* from cover appears unlikely to permit robust reporting of algal effects across both indicators of abundance in the Horizons region. However, strong within-site relationships at many sites in the Manawatu and Ohau catchments may offer that capability, although using different relationships at each site. The fact that the relationships performed well in cross-validation tests (e.g., $R^2 > 0.5$), using a long time-series suggests that the relationships are robust over time at these individual sites.

In this analysis we did not explore in detail the approach of developing chlorophyll *a* equivalents for converting different types of periphyton cover direct to a chlorophyll *a* estimate (i.e., Method 6 in Section 9.2.1)[7]. From the wide variety of results of simpler relationships at different sites (Methods 2 – 5 in Section 9.2.1) we assume that conversion factors (if applicable at all) will vary from site to site. Such between-site variation has been observed in other datasets. For example, factors for converting cover data to chlorophyll *a* derived during a detailed study of three Canterbury rivers (Kilroy et al. 2013) were later applied to a wider range of sites in Canterbury (Kilroy et al. 2017). In that study, although overall correspondence between observed and predicted chlorophyll *a* remained relatively strong (>75% of the variance in chlorophyll *a* explained) at the new sites, slightly better correspondence was obtained by adjusting conversion factors (Kilroy et al. 2017). However, both the original (Kilroy et al. 2013) and adjusted conversion factors showed variation in their performance at different sites. In particular, we noted a different relationship between observed and predicted chlorophyll *a* at sites in alpine-fed rivers compared to hill-fed rivers. The next step for the Horizons data may be to explore possible classification of sites into groups with different relationships between cover and chlorophyll *a*.

[7] Attempts were made to apply conversion factors (e.g., in Kilroy et al. (2012), and modifications). None of the combinations trialled produced closer relationships than those found between chlorophyll *a* and the simpler cover measures used (Mats, Fils, Film, WCC).

10 Classification of sites and implications for predicting chlorophyll *a* and setting nutrient limits

Key messages

- We aimed to determine the scope for grouping sites, based on output from earlier chapters in this report, to discriminate between differing site responses of chlorophyll *a* (e.g., on the basis of effective flow, correlations between measures of standing crop, within-site relationshps with environmental factors, nutrient limitation status, conductivity, geology and catchment land use).
- Sites were assigned to groups (i.e., classified) on the basis of site characteristics, catchment characteristics, and the variables included in, or strength of, within-site relationships between chlorophyll *a* and environmental variables. We also considered the strenths of within-site relationships between chlorophyll *a* and periphyton cover.
- The strongest pattern noted was that sites with strong within-site relationships between chlorophyll *a* and cover also had the stongest within-site relationships with environmental variables (including with accrual time (days since an effective flow)). These sites included most sites in large rivers, had higher DIN and finer sediment, and were in catchments with high proportions of their area in farmland and low proportions in indigenous forest. All these variables were generally intercorrelated and it was not clear what was driving the pattern.
- Grouping sites by their within-site chlorophyll *a* – environment relationships did not generate a pattern aligned with catchment geology or (life-supporting capacity (LSC) class. Treating sites within each LSC or geological class alike in terms of management actions is therefore unlikely to deliver equivalent periphyton (chlorophyll *a*) outcomes.
- However, the strengths of the within-site relationships again showed some patterns across catchment geology classes. Sites with AL and SS geology had stronger within-site relationships than those with HS or VA geology.

10.1 Introduction

The third objective in the contract was to "Classify stations on the basis of their within-station environmental drivers, to permit generalisation of earlier driver findings, and comparison of findings to between-station inferences across all stations."

The purpose of the analysis was to determine whether there was scope for groups of sites to be managed for periphyton biomass together by focussing on an influential factor (for periphyton) common to the sites in a group(s). The analysis in this objective therefore comprised investigating whether groups of sites (derived from the results of the within-site relationships) were identifiable by other factor(s) such as geology, climate, stream order or size, position in the catchment (altitude / temperature), or to existing classifications such as Horizons LSC classes.

The task set in Objective 3 of the contract was based on an assumption that there would be clear within-site relationships with environmental variables and that these variables would vary across sites, allowing sites to be divided into groups based on inclusion of different variables in the relationships. However, the outcome of the within-site analysis was that it proved difficult to identify relalationships that suggested cause and effect. For example, the relationships were more likely to reflect within-site correlations between flows and nutrients rather than provide evidence for factors (other then flows) that might be influencing periphyton biomass. Despite this, the relationships may still tell us something about the type of site. Therefore, we proceeded to create classifications (i.e., assign sites to groups) based on the relationships, and to test whether the groups of sites defined in the classifications were distinguishable using other variables.

10.2 Methods

Classifications used in the analysis are listed in Table 10-1. The classifications fall into five categories: within-site relationships; nutrients/water quality; hydrology/physical characteristics; catchment; and "other" (pre-defined).

Once the classifications were defined (i.e., the basis of assigning sites to groups was decided), we analysed the data by addressing the following four questions.

1. Are groups of sites defined based on the **strength of within-site relationships** distinguishable using site and/or catchment characteristics?

Sites were grouped according to the strength of the within-site relationships (as indicated by NSE of cross-validated regression) (classifications 2, 3 and 4 in Table 10-1). For chlorophyll *a* vs. multiple variables we considered the relationships in Table 7-4 in which DIN and DRP were means calculated from the previous 6 months of data. For chlorophyll *a* vs. cover we used the relationships derived from a multiple regression using Film, Mats and Fils (see Table 9-3).

The values of water quality, hydrological and catchment land cover variables were compared across the groups of sites in classifications 2, 3 and 4 by first inspecting box plots. Where the variable ranges appeared to differ between groups, or were characteristic for a group (i.e., in a relatively small range), we ran non-parametric Kolmogorov-Smirnov (KS) Two-Sample to assess the significance of any differences. Note that numbers of sites in each group varied. Therefore, it was not always possible to establish differences between groups where numbers of sites were low.

Correspondence between classifications 2, 3 and 4 and multiple variables was investigated by first selecting combinations of variables that could potentially influence the relationships, then conducting non-parametric ANOSIM (Analysis of Similarities) tests to determine whether there were significant differences between groups based on the combination of variables. The explanatory variables were first normalised so that they are all expressed on the same scale (0 to 1). This adjustment prevents situations where a variable with a large range of nominal values overwhelms the effect of a variable with a small range of values. A dissimilarity matrix was then generated using the normalised data. Sites with more similar combinations of variables have lower similarity scores than those with dissimilar combinations. ANOSIM uses a randomisation test to determine whether the predetermined groups of sites are significantly different (i.e., more similar within the group than between groups). Non-metric multi-dimensional scaling (NMDS) was used to visualise the groups in a two-dimensional plot. Sites similar to each other plot closer together than sites that are dissimilar.

2. Are groups of sites distinguishable on the basis of **DIN or DRP with a positive or negative effect** contributing to the strongest within-site relationships between chlorophyll *a* and multiple variables?

Sites were assigned to one of three groups for each of DIN and DRP: DIN or DRP included with a positive effect, included with a negative effect, and not included (classifications 5 and 6 in Table 10-1). Differences between the three groups for DIN and the three groups for DRP were determined for the continuous water quality and hydrological variables using box plots followed by KS tests, as in question 1 above. Correspondence with categorical catchment variables (geology, climate and HSC class) was assessed using matrices.

3. Are site characteristics (water quality and hydrological / physical), catchment geology, climate and LSC classes associated with the strengths of within-site relationships between chlorophyll *a* and flow / nutrient variables and between chlorophyll *a* and cover?

The correspondence of categorical site and catchment groups (classifications 9 to 31 in Table 10-1) to the strengths of within-site relationships was evaluated by comparing box plots of relationship strengths (NSE as a continuous variable) against groups in each site and catchment classification, followed by KS tests, as above.

4. Are groups of sites distinguishable based on the relationship between chlorophyll a and DIN or DRP based on peak biomass only?

An interesting result in the within-site analysis was that when data at each site were reduced to annual peak biomass, in some cases, positive relationships were identifiable between chlorophyll *a* and DIN or DRP (using averaged data over the previous 6 months). This contrasted with generally negative relationships (especially with DIN) when all of the data were used, because of the pattern at many sites of highest DIN occurring at high flows and/or in winter, when periphyton tends to be low. Isolating peak biomass may allow genuine cause – effect patterns to be revealed, although relationships were weak because of the small number of datapoints (maximum of n = 9, including part years). Sites with positive, negative and no relationships between chlorophyll *a* and DIN or DRP were grouped on the basis of the correlation coefficient (classification 7 in Table 10-1). Correspondence between the groups and the continuous environmental variables was determined using box plots followed by KS tests, as above.

10.3 Results

10.3.1 Assigning sites to groups

Classifications and group memberships for each site are presented in Appendix K. Site groups based on the NSE of relationships were based on the range of NSE in each case, and used natural breaks in the data. We also judged that an NSE < 0.5 could not be considered as a ‟strong” relationship. Therefore, for the chlorophyll *a* – accrual period relationship, the highest NSE of 0.5 (at tiraumea_nga) was only assigned to the ‟mod” group.

Where the groups were derived from continuous variables, the thresholds defining different groups were based on the data range. In some cases, the thresholds were effectively arbitrary as there was no clear basis for assigning sites to groups, except around existing breaks in the data. For example,

Table 10-1: List of classifications used to partition sites into groups. Data and classifications for every site and for each classification are shown in Appendix K. The maximum number of groups of site in each classification is seven. Numbers of groups range from three to seven.

No	Classification name	Basis of classification	Classes (groups)							Group names / notes
			1	2	3	4	5	6	7	
Based on within-site relationships										
1	Definition of effective flow	Removal threshold (*n* x median) and strength of relationship	< 5 x med.	> 5 x med.	Unres-ponsive	no clear threshold				A, B, C, D (Table 3.1)
2	Chlorophyll *a* – accrual period	Relationship strength, NSE	<0.2	≥0.2<0.35	≥0.35-0.5					vweak, weak, mod (see Table 7-4)
3	Chlorophyll *a* – multiple variables	Relationship strength, NSE (6-m DIN)	<0.15	≥0.15<0.35	≥0.35<0.5	≥0.5				vweak, weak, mod, strong (Table 7-4)
4	Chlorophyll *a* – cover	Relationship strength, NSE (Film, Mats, Fils)	<0.25	≥0.25<0.5	≥0.5<0.65	≥0.65				vlow, low, mod, strong (Table 9-3)
5	Within-site correlates with chl a	Includes n (positive or negative)	n	n-	x					See Table 7-4
6	Within-site correlates with chl a	Includes p (positive or negative)	p	p-	x					See Table 7-4
7	Peak annual chl a vs DIN or DRP	Positive, negative or no relationship, correlation coefficient	>0.2	<-0.2	-0.2 – 0.2					positive, negative, none (Table 7-3)
8	Within-site relationship with cover	Variables included	FGfM	FGf	FM	GfM	Gf	M		See Table 9-3
Based on nutrients and water quality										
9	Nutrient limitation, classical, all flows	limiting nutrient, from N:P ratio	N	P	co					See Table 4-1
10	Nutrient limitation, classical, low flows	limiting nutrient, from N:P ratio	N	P	co					
11	Nutrient limitation, concs, all flows	limiting nutrient, concs	N	P	co	none				See Table 4-2
12	Nutrient limitation, concs, low flows	limiting nutrient, concs	N	P	co	none				
13	Mean DIN concentration	DIN, mg/m^3	<50	50 < 100	100 < 300	300 < 550	550 < 750	>750		vlow to xhigh
14	Mean DRP concentration	DRP, mg/m^3	<7	7<10	10<15	15<40	>40			vlow to xhigh
15	Mean conductivity	µS/cm	<90	90<130	130<200	>200				low to vhigh
16	Periphyton chlorophyll *a*	mean chl a (mg/m^2)	<5	5<15	15<30	30<50	>50			vlow to vhigh
17	Periphyton chlorophyll *a*	92^{nd} percentile (mg/m^2)	<15	15<50	50<120	120<200	>200			vlow to vhigh

No	Classification name	Basis of classification	Classes (groups) 1	2	3	4	5	6	7	Group names / notes
Based on site-specific hydrological and physical characteristics										
18	Stream size	Mean flow (m^3/s)	<4	4 < 10	10 < 30	>30				small, med, large, vlarge
19	Frequency of effective flow	Mean annual frequency (FRE_EFF)	<2	≥2<6	≥6<8	≥8<10	≥10			vlow to vhigh
20	Flood frequency	FRE3	<2	≥2<5	≥5<8.5	≥8.5<10.5	≥10.5			vlow to vhigh
21	Flood frequency	FRE10	<2	≥2<4	≥4<8.5	≥8.5				vlow to high
22	Stream order	Order from REC	1	2	3	4	5	6	7	
23	Substrate, coarse	% cover, bedrock + boulders	<2	2 < 5	5 < 10	10 < 20	>20			vlow to vhigh
24	Substrate, fine gravel / sand	% cover, sand + fine gravel	<7.5	7.5 < 10	10 < 15	>15				low, mod, high, vhigh
25	Altitude	m a.s.l.	<100	200 < 500	> 500					low, hill, upland
Based on catchment characteristics										
26	Geology	dominant top rock	AL	HS	LO	SS	VA			Horizons supplied
27	Geology	dominant base rock	AL	HS	LI	SS	VA			
28	Climate	REC class	CW	CX	CD					
29	Landcover, farms	% cover, farm	<1	1 < 10	10 < 25	25 < 50	50 < 75	>75		vlow to xhigh
30	Landcover, grassland	% cover, grassland	<0.25	0.25 < 1	1 < 10	>10				vlow to high
31	Landcover, indigenous forest	% cover, indig_for	<2.5	2.5 <10	10 < 20	20 < 50	50 < 75	>75		vlow to xhigh
Other (from Horizons)										
32	River catchment	River name / subregion								See Table J-4
33	LSC class	Horizons classification	HM	HSS	LM	UHS	Uli	UVA	UVM	
34	Waste-water treatment plant	downstream of WWTP	y	n						

the four conductivity groups defined groups of sites where there was a difference of at least 5 µS/cm between the highest value in one group and the lowest value in the next group up. The range of mean conductivity in the lowest group was 50 to 86 µS/cm. Because 50 µS/cm is not an extremely low conductivity value[8] the group was named "low" rather than "vlow". The "vhigh" group defined a group of sites at which the lowest conductivity (211 µS/cm) was 23 µS/cm higher than the highest mean value in the "high" group. The same approach was taken for other variables where there was no predefined basis for forming groups.

10.3.2 Are groups of sites defined based on the strength of within-site relationships distinguishable using site and/or catchment characteristics?

Chlorophyll *a* – accrual period relationships (classification 2, Table 10-1)

The strength of the relationship between chlorophyll *a* and days since an effective flow corresponded, in part, to the strength of the within-site relationship between chlorophyll *a* and percentage cover in that there was overlap of sites with high NSE in the two relationships. Sites with the strongest relationships between chlorophyll *a* and days since an effective flow also tended to have higher conductivity, included most sites in large rivers, and were in catchments with high proportions of their area in farmland and low proportions in indigenous forest. Peak chlorophyll *a* showed a gradient of increasing median value over the three groups but the difference between groups was not significant (Figure 10-1).

Chlorophyll *a* – multiple variable relationships (classification 3 in Table 10-1)

Only two sites (makakahi_doc and mangatainoka_putara) were included in the vweak relationships group. These two headwater sites had low DIN, DRP and conductivity, and high proportions of coarse material on the substrate and were excluded from the main comparison because of low numbers. Site and catchment characteristics of groups of sites in the mod and strong groups (n = 19 and 4, respectively) were similar. We therefore combined the mod and strong sites into one group (mod – strong) and compared the latter group (n = 23) with sites with weak relationships (n = 14).

The relationships all included the flow variable DaEFF. Therefore, the results were similar to those of the flow groups only (see above). Exceptions were that mean DIN differed significantly between the weak and mod-strong groups (KS test, P = 0.006), as did pc_coarse (KS test, P = 0.018) (Figure 10-2). Both these differences reflected the catchment differences: sites with moderate to strong relationships tended to be in catchments with high proportions of farmland in their catchments, and these sites also tended to have high DIN and generally fine substrate. They also included most of the larger rivers (Figure 10-2).

[8] As a comparison, the lowest mean conductivity across all sites in the NRWQN is about 40 µS/cm (in the Monowai River, Southland). Much lower conductivity is typical in first-order streams in undeveloped areas (e.g., < 25 µS/cm, Kilroy et al. 2006).

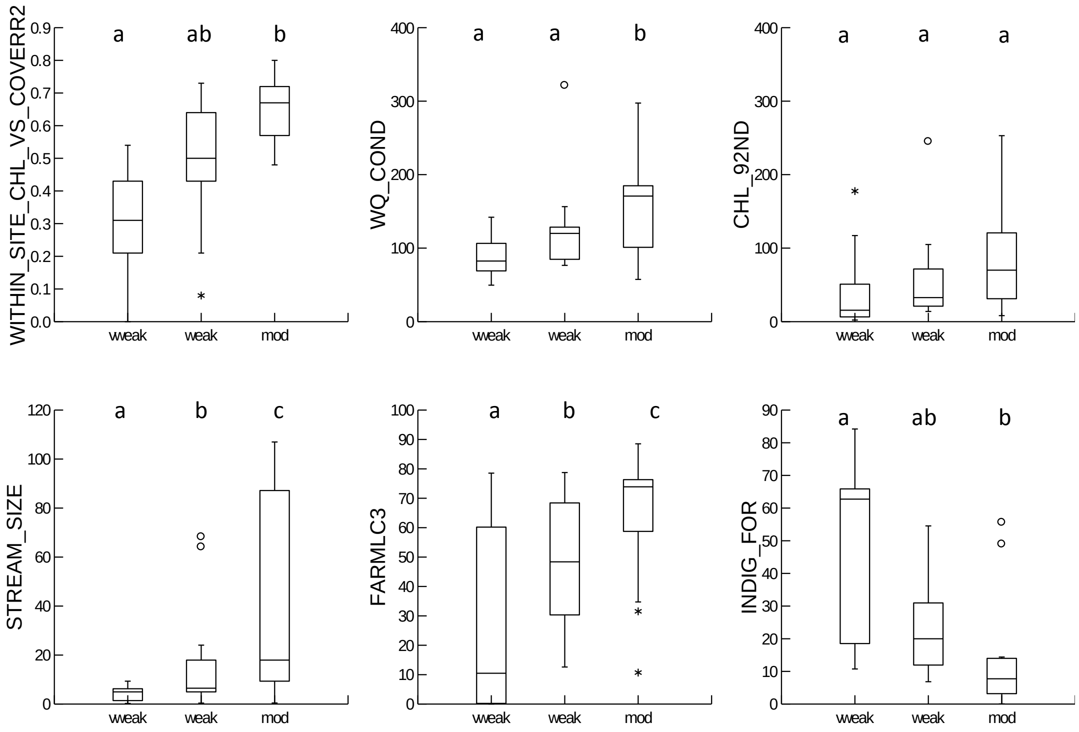


Figure 10-1: Box plots of site and catchment variables associated with site groups based on the strength of the relationship between chlorophyll *a* and days since an effective flow. Classifiation 2 in Table 10-1. Numbers of sites in the vweak, weak and mod groups were 9, 13 and 17, respectively. Different letters above each bar indicate differences between groups in a KS test (P < 0.1).

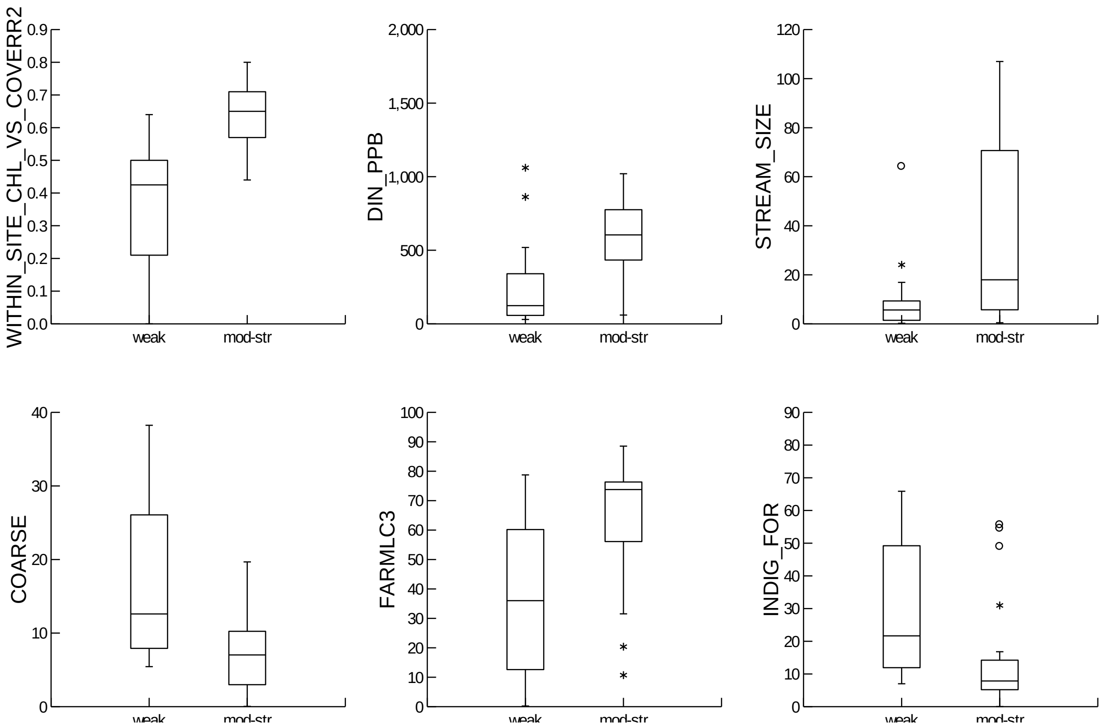


Figure 10-2: Box plots of site and catchment variables associated with site groups based on the strength of within-site relationships between chlorophyll *a* and multiple variables. Classifiation 3 in Table 10-1. Two groups were compared. Numbers of sites in the weak and mod-str groups were 14 and 23 respectively. The groups differed significantly (KS test, P < 0.05, or P < 0.1 for stream size). Refer to text for explanation of reduction of the four groups in Table 10-1 to two.

Chlorophyll *a* – cover relationships (classification 4 in Table 10-1)

The general pattern for the strength of within-site relationships between cover and chlorophyll *a* was that sites lower down in catchments had stronger relationships than sites in more upstream locations. The pattern was reflected in sites with strong relationships being mainly larger rivers at low altitudes, with higher DIN concentrations and lower proportions of coarse substrate than sites with weak relationships (Figure 10-3). Chlorophyll *a* abundance did not differ between groups except between the moderate and high-strength groups: no trend was evident across the groups (Figure 10-3). There was also no difference in flood frequencies between the groups (FRE3 shown as an example in Figure 10-3).

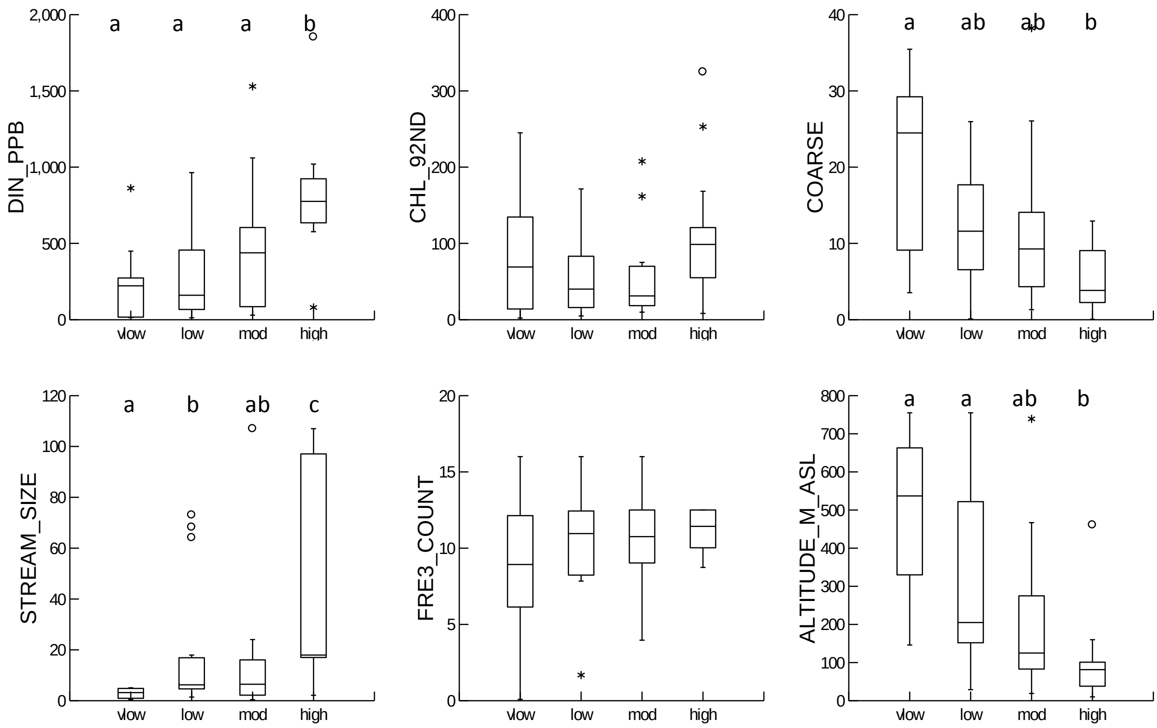


Figure 10-3: Box plots of site and catchment variables associated with site groups based on the strength of the relationship between chlorophyll *a* and percentage cover. Classification 4 in Table 10-1. Numbers of sites in the vlow, low, mod and high groups were 11, 19, 17 and 14, respectively. Different letters above each bar indicate differences between groups in a KS test ($P < 0.05$). Plots for chlorophyll *a* and FRE3 (the two centre plots) are shown to illustrate that these variables did not differ between the groups.

Distinguishing groups using combinations of variables

Strength groups for the chlorophyll *a* – accrual period (classification 2) and chlorophyll *a* – multiple variables (classification 3) relationships were not more clearly defined using a combination of variables in an ANOSIM test than using single variables. Groups defined by the strength of the relationship between percentage cover and chlorophyll *a* (classification 4) were separated in an NMDS plot with groupings reflecting the patterns seen in single variables (Figure 10-4). ANOSIM showed that all groups differed significantly ($P < 0.05$) except for the mod and high groups, which overlapped. The strongest differences were between the vlow and high and low and high groups ($P < 0.001$).

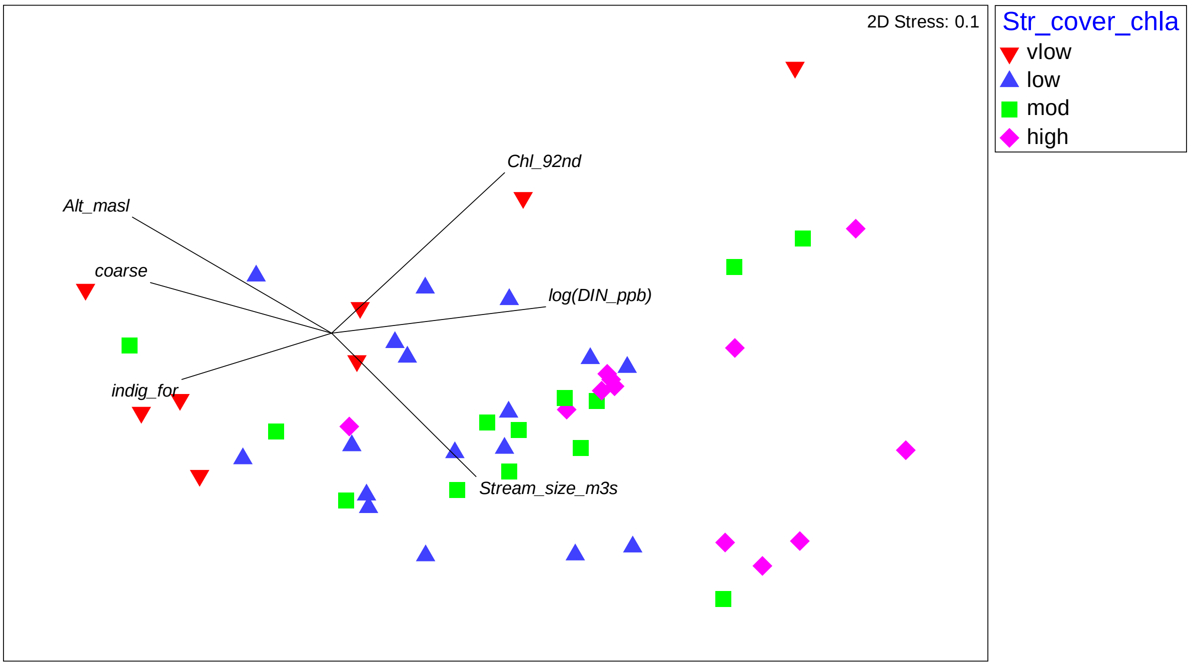


Figure 10-4: Non-metric multi-dimensional scaling ordination of sites showing separation of site groups based on the relationship between chlorophyll *a* and and percentage cover. Vectors on the plot show the direction of influence of the variables included. Sites with strong relationships are in larger rivers with higher DIN, and sites with weak relationships are at higher altitude and have higher percentages of coarse substrate. The direction of the chlorophyll *a* vector indicates no real effect of peak chlorophyll *a* in separating the groups.

10.3.3 Are groups of sites distinguishable on the basis of DIN or DRP (with a positive or negative effect) contributing to the strongest within-site relationships between chlorophyll *a* and multiple variables?

The analysis applied to classification 5 and 6 in Table 10-1. Numbers of sites varied widely across groups, limiting the comparisons. In particular, only three sites (of the 39 sites with flow records included in the analysis) had within-site relationships in which DIN had a positive coefficient.

The continuous site or catchment variables generally did not differ significantly between groups of sites based on inclusion of DIN or DRP (with a positive or negative coefficient) in within-site relationships. The only pattern noted was that all sites in which DRP was included as a predictor with a negative coefficient had moderate to high DIN (> 250 mg/m^3). However, the differences between groups were not significant because some sites with a positive coefficient or with no DRP included also had high DIN. No such pattern was evident for the groups based on DIN (Figure 10-5).

The different DIN and DRP groups were generally not concentrated at sites with particular geologies or HSC classes (Table 10-2). An exception was that 100% of the sites with AL geology (top rock and base rock) had relationships including DIN with a negative coefficient. However, with only six (top rock) and seven (base rock) sites with AL geology, we cannot be certain that the pattern is meaningful.

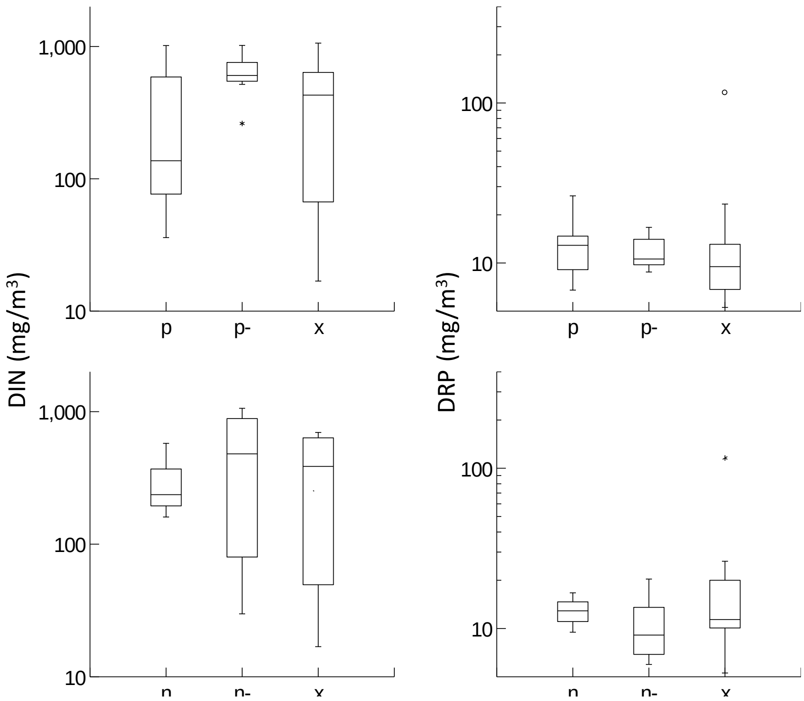


Figure 10-5: Box plots of geometric mean DIN and DRP in groups of sites with DRP or DIN included in a within-site relationship. n, p = sites with DIN or DRP included with a positive coefficient; n-, p- = DIN or DRP included with a negative coefficient; x = DIN or DRP not included. Numbers of sites: **for DRP**, p 8, p- 7, x 24; **for DIN**, n 4, n- 21, x 14.

10.3.4 Are catchment geology, climate and LSC classes associated with the strengths of within-site relationships between chlorophyll *a* and flow / nutrient variables and between chlorophyll *a* and cover?

In this analysis we looked at whether site groups based on catchment characteristics (classifications 13 to 32 in Table 10-1) had characteristic strengths of within-site relationships (with strength represented by the NSE of the relationships). The plots for classifications 13 to 25 reflected the results in Section 10.3.2 and are not presented.

For the catchment classifications (26 to 32 in Table 10-1), box plots showed overlaps in the strengths of relationships between sites with different geology and HSC class (Figure 10-6).

For the chlorophyll *a* vs. accrual period (Da_EFF) relationships, NSE did not differ significantly between groups of sites with top rock or base rock AL, HS, SS or VA. Uneven numbers of sites in the LSC classes (see Table 10-2) precluded a valid test. The outcome was similar for the chlorophyll *a* vs. multiple variables within-site relationships (Figure 10-6). The number of sites was limited to those with a flow record and sufficient data (up to 40 sites).

Within-site relationships between chlorophyll *a* and percentage cover were run at all sites (up to 62 sites) and numbers of sites in different geology classes were more even. LSC classes were still heavily biased towards HS (Figure 10-6). KS tests indicated that the within-site relationships were, on average, significantly stronger at sites with AL and SS top rock geology than at sites with VA top rock geology (KS tests, $P < 0.05$). Differences were stronger for base rock geology: all geology groups had significantly different relationship strengths (KS tests, $P < 0.05$), except for HS vs. SS ($P < 0.1$, marginally significant) and AL vs. SS ($P = 0.79$, not significant).

Table 10-2: Percentages of sites at which within-site relationships included DIN or DRP as explanatory variables, by catchment geology and LSC. Classificcations 5 and 6 in Table 10-1. DIN- and DRP- are sites at which the coefficient for DIN or DRP was negative. X = neither DIN not DRP included. Percentages are relative to the number of sites in each geology or HSC class (shown in the centre column). Note that the total number of sites was 40 (all sites with a flow record plus sufficient data for the within-site analysis).

		Classification 5: DIN included					Classification 6: DRP included		
		DIN	DIN-	X	Total n		DRP	DRP-	X
Top rock	AL	0	100	0	6	AL	17	0	83
	HS	0	62	38	14	HS	23	31	46
	SS	15	31	54	13	SS	31	15	54
	VA	14	57	29	7	VA	43	14	43
Base rock	AL	0	100	0	7	AL	14	14	71
	HS	0	54	46	13	HS	23	15	62
	LI	0	100	0	1	LI	0	100	0
	SS	13	40	47	15	SS	33	13	53
	VA	33	33	33	3	VA	67	33	0
LSC	HM	7	61	32	28	HM	29	11	57
	HSS	0	0	100	1	HSS	0	100	0
	UHS	0	60	40	5	UHS	20	20	60
	Uli	0	100	0	1	Uli	0	100	0
	UVA	33	33	33	3	UVA	67	33	0
	UVM	0	0	100	1	UVM	100	100	0

Differences across LSC classes were unclear because of large differences in numbers of sites. The most notable difference was between HM sites (n = 33) and UVA and UVM sites (n = 12 and 3 respectively) (KS tests, $P < 0.01$).

The bottom set of plots in Figure 10-6 shows the 92nd percentile of chlorophyll *a* in each geology group and HSC class. The same differences in chlorophyll *a* between geology groups were evident for base rock and top rock. In both cases, chlorophyll *a* at sites with HS geology was lower, on average, than that at sites with AL, SS and VA geology.

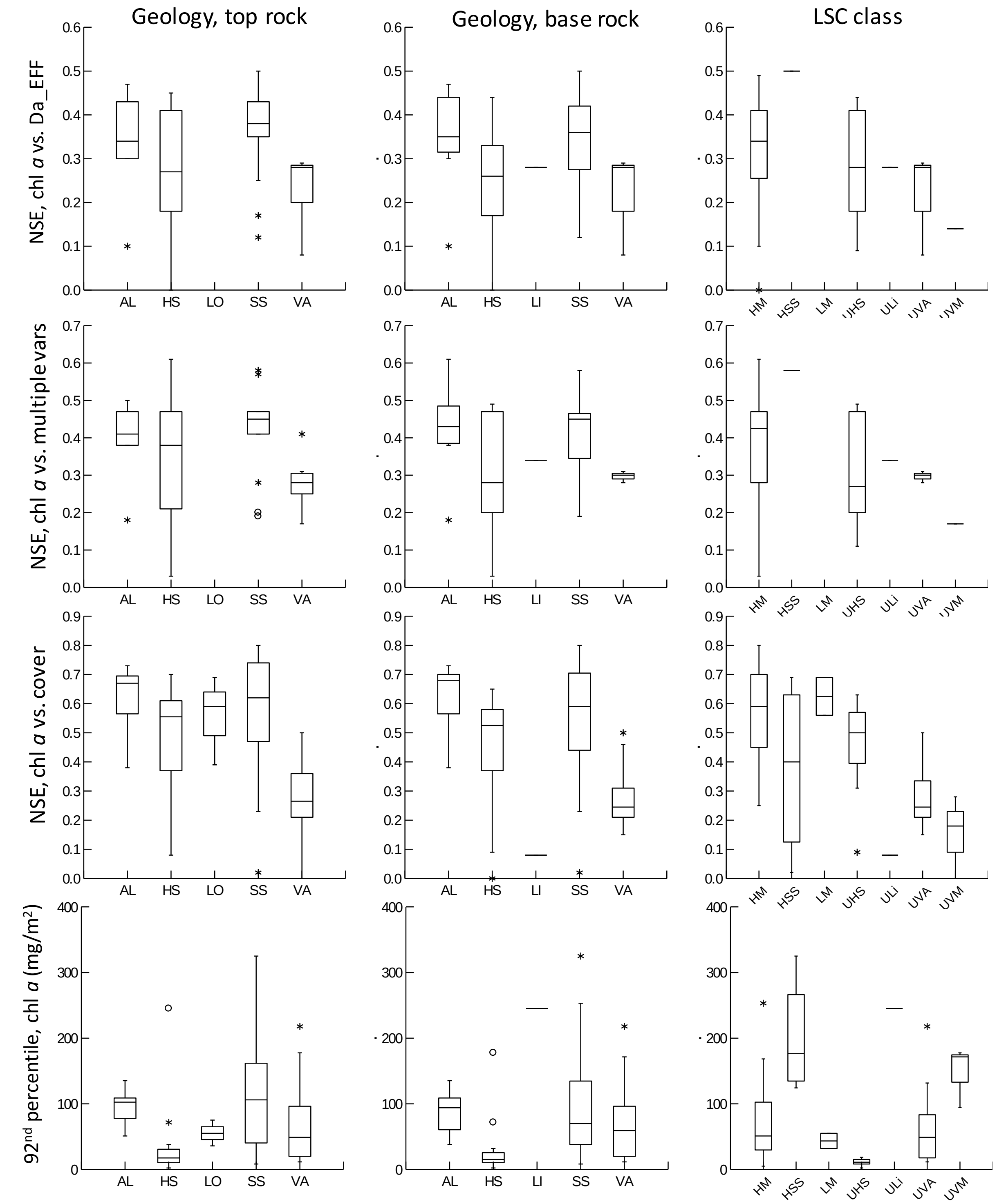


Figure 10-6: Box plots of relationship strength in groups of sites with different geology and LSC class. Refer to Appendix J for numbers of sites in the geology and HSC classes in the top two sets of pots (i.e., total of 39 sites at which an effective flow was identified). All sites were available for the lower two sets of plots. Site numbers are: top rock: AL, 7; HS, 16; LO, 3; SS, 18; VA, 18; base rock: AL, 8; HS, 16; LI, 1; SS, 23; VA, 14; LSC class: HM, 33; HSS, 4; LM, 2; UHS, 7, ULi, 1; UVA, 12; UVM, 3.

10.3.5 Are groups of sites distinguishable based on the relationship between chlorophyll *a* and DIN or DRP based on peak biomass only?

Box plots of the negative, no relationship and positive groups (from Table 7-3, classification 7 in Table 10-1) against the site and catchment characteristics showed no clear patterns: in all cases there was a wide range within each group. Plotting the r-values for each site (from Table 7-3) against the groups of sites defined by chlorophyll *a*, DIN, DRP, conductivity, and geology again showed no patterns. This analysis was not taken any further.

10.4 Discussion

Overall, groups of sites based on the results of the within-site analyses did not separate out strongly on the basis of other site characteristics such as water quality (DIN, DRP, conductivity) or catchment features (such as land cover and geology). There was always overlap between the groups, which was expected because group definitions were arbitrary and based on the data.

Nevertheless, there were indications that sites with strong relationships between chlorophyll *a* and percentage cover were concentrated in the larger rivers at low altitudes, with high DIN concentrations, low proportions of coarse substrate, and generally with periphyton sensitive to high flows. All the disinguishing variables were intercorrelated. Consequently, the reason for the pattern is unclear. One explanation is that the group of sites with strong chlorophyll *a* vs. cover relationships have generally similar characteristics and for that reason support a similar suite of periphyton taxa, so that cover by film, mats and filaments corresponds to a similar quantity of chlorophyll *a* at all sites.

The pattern for chlorophyll *a* vs. cover relationship strength was reflected by chlorophyll *a* vs. multiple variables and chlorophyll *a* vs. effective flow strength (classifications 2 and 5 in Table 10-1), with stronger relationships tending to occur in larger lowland rivers. Stronger relationships also occurred at sites with AL or SS geology than VA or HS, and this pattern is also related to position of the sites in the catchment (with AL and SS downstream). Possibly similarity of periphyton composition also drove these relationships. Taxonomic data are available for periphyton at the Horizons monitoring sites on up to three occasions. Therefore, it would be possible to check whether the idea of uniformity of community composition across these sites holds, and whether community composition or structure at other sites is really different.

Crucially, given the pupose of this exercise (see Introduction to this section) there was no indication that inclusion of DIN or DRP as a predictor variable in the within-site relationship, or the direction of the relationship (positive or negative), corresponded to groups of sites with particular characteristics (such as geology or land cover, or water quality and physical features). Therefore, there appears to be limited scope, at this stage, for identifying sites that could be managed together on the basis of a common factor influencing periphyton.

Ideally site groupings would be based on the LSC class assigned to each site by Horizons, based on geology and position in the catchment. The two largest classes (in terms of number of sites) differed in the strength of relationship between cover and chlorophyll *a*, but not in the strength of the within site relationships with flow and other variables. The assessment of whether classes varied in their response to DIN or DRP was hampered by strong bias in the dataset towards sites in the the HM class.

Other classifications could be added to the list in Table 10-1 (but are beyond the scope of the current analysis). For example, sites could be separated into groups on the basis of whether there is seasonality in DIN or DRP, or significant relationships between DIN or DRP and flow.

11 Synthesis and conclusions

The analysis described in this report focussed on time-series data (at least monthly) on periphton and associated environmental variables collected by Horizons at over 60 river sites thoughout the Manawatu-Whanganui region since late 2008. This dataset is the longest and most comprehensive of its type in New Zealand and is probably unusual worldwide in terms of length of coverage and data resolution. Accompanying hydrological data (a continuous flow record) was available for 50 of the sites.

The contract for the analysis of this dataset (see Appendix A) described in the preceding sections specified three objectives:

- Objective 1 – Establish the significance and strength of relationships between environmental factors and periphyton standing crop (max or 92nd% Chl-*a* between stations, observed Chl-*a* within station time-series);
- Objective 2 – Establish if the resolution of sampling affects the performance of environmental drivers identified in Objective 1;
- Objective 3 – Classify stations on the basis of their within-station environmental drivers, to permit generalisation of earlier driver findings, and comparison of findings to between-station inferences across all stations.

The analysis in response to all objectives was underpinned by the assumption (based on substantial previous research, summarised in Biggs 2000b) that the environmental factors of most relevance in controlling abundance of periphyton were accrual period (the time available for development of periphyton in a river, without significant removal through the effects of high flows) and dissolved nutrient concentrations (dissolved inorganic nitrogen, DIN, and dissolved reactive phosphorus, DRP). Other contributing factors include light (e.g., presence of shade or not), temperature, river bed substrate composition, water conductivity.

In all periphyton – environment analyses, periphyton abundance was represented by chlorophyll *a*. The metrics used were annual maximum chlorophyll *a* (when considering annual datasets) or the 92nd percentile of chlorophyll *a* (when considering multi-year datasets). The 92nd percentile of chlorophyll *a*, (calculated from at least three years of data) is the metric used in the periphyton attribute of the National Policy Statement for Freshwater Management (NPS-FM). Below, the two metrics are referred to together as peak chlorophyll *a*.

Below we set out the key conclusions from the three objectives, along with a brief commentary on their significance and relevance to river management.

11.1 Objective 1: Between and within-site relationships between chlorophyll *a* and environmental variables

A key part of the analysis was an initial investigation into the responses of periphyton chlorophyll *a* at each of the 50 sites with a linked flow record to preceding flows of different sizes. From this analysis we were able to identify at over 30 sites a threshold flow size (in multiples of median flow) above which periphyton was typically removed to low levels. A flood threshold of 3 x median flow has been used previously as a "rule-of-thumb" for representing flows that remove periphyton (based on an analysis by Clausen and Biggs 1996). We showed that effective flood thresholds in the Horizons region varied from 1.5 to 15 x median flow. This enabled derivation of a variable for accrual period at each site based on the effective flow (termed Da_EFF).

We first used the data from Horizons sites as independent data to test relationships between maximum annual chlorophyll *a* and mean DIN or DRP, plus mean accrual period (using 3 x median flow), developed by Biggs (2000a). The outcome of that analysis was that the Biggs (2000a) relationships were not good at predicting annual maximum chlorophyll *a* at many sites. Reasons for poor predictions included the limited range of DIN used by Biggs (2000a) and a uniform river type that may not conform to river types in the Horizons region. Consequently, development of new models specific to the Horizons region was justified.

11.1.1 Between-site analysis

The between-site analysis showed that, although mean DIN and peak chlorophyll *a* were positively correlated in all time periods, mean DIN in isolation from other variables was not a good predictor of peak chlorophyll *a* across sites within the Horizons region. Mean DRP was weakly or not correlated with peak chlorophyll *a* in all periods.

Combining DIN, DRP and Da_EFF with other variables in multiple regression analyses, the strongest models in each time period included **DIN**, **conductivity** and **Da_EFF** as predictors. The initial models explained at least 50% of the variance in peak chlorophyll *a* across sites in all time periods. Leave-one-out cross-validation (a robust method for evaluating the predictive ability of models) produced encouraging results, with high proportions of variance in observed chlorophyll *a* explained by predicted chlorophyll *a* for some periods (e.g., 75% in 2012 - 2015). Including Da_EFF as a predictor always produced stronger relationships than accrual period from 3 x median flow.

Conclusion: The best between-site models appear to be good enough for application to management. Application of the models was not discussed in the analysis section. Since one of the main objectives of the Horizons periphyton monitoring programme was to develop models for use in river management, the following suggestion is provided for applying the models to river management:

- Generate predictions of the 92nd percentile of chlorophyll *a* under a range of different scenarios.
- These scenarios could be used to predict the range of DIN or DRP under which the 92nd percentile of chlorophyll *a* would fall into different NPS-FM bands, or meet different targets in the Horizons One Plan, given different combinations of conductivity, bed substrate, water temperature and accrual period.
- A worked example is shown in Appendix L. The tables in Appendix L can be used to read off the combinations of conditions that would be expected (within the error of the relationships) to lead to exceedance of (say) 200 mg/m^2 chlorophyll *a*. Note that the predictive relationship used includes both DIN and DRP as predictors. Given the lack of association between peak chlorophyll *a* and DRP in isolation from other variables, the inclusion of DRP in the model should be treated with caution. Similar tables can easily be generated using alternative relationships.

11.1.2 Within-site analysis

Consistently with many previous studies, the between-site analysis took a "space-for-time" approach to identifying relationships between peak periphyton abunfance (as chlorophyll *a*) and environmental variables. However, the unusually long (> seven years) and detailed datasets at individual sites in the Horizons region provided an opportunity to conduct analyses within sites. The analysis proceeded in steps, and we assumed that the most important variables controlling periphyton biomass (as chlorophyll *a*) over time were DIN, DRP, accrual period (based on the effective flow) and temperature.

We first explored simple correlation relationships between chlorophyll *a* and DIN or DRP at each site, using a range of metrics including mean DIN and DRP over the previous four or six months. Use of preceding values recognises that chlorophyll *a* on a particular day is the result of preceding conditions. Because DIN and DRP can be positively correlated with flow, correlations were run on the data both including and excluding samples collected at high flows.

The outcome of the analysis was that negative correlations predominated, especially for DIN. At many sites the negative correlations persisted even when only samples taken at flows less than half median flow were included. Persistent negative correlations with DIN regardless of flow was seen at all four monitoring sites in the Rangitikei River, and in that case may represent evidence of instream uptake of DIN. Similar strong patterns were not observed for DRP.

Filtering the data further to include only the mean annual maximum value of chlorophyll *a* (i.e., only one data point per year) revealed that the DIN or DRP relationship reversed from negative (using all of chlorophyll a observations) to positive (using annual peak chlorophyll *a*) at many sites. For DIN, the change was from four (7%) to 19 (33%) sites. For DRP the change was from 22% positive relationships to 41%. We interpreted the shifts as evidence of nutrient effects on peak chlorophyll *a*.

Within-site analyses using mutlple explanatory variables were run using techniques similar to those used for the between-site analyses. Negative coefficients for DIN and DRP persisted in the prelationships, and it was clear that most of the explanatory power of the relationships at most sites was attributable to the effects of accrual time.

Conclusion: The analysis overall demonstrated that there is no clear and simple linear relationship between nutrient availability throughout the year and corresponding algal standing crop.

The results were discussed in the context of a conceptual model (from the literature) that illustrates the interactions that obscure straightforward chlorophyll *a* – nutrient relationships. Alternative approaches were suggested (e.g., parallel trend analyses, mechanistic models). It was also pointed out that a longer dataset could lead to clearer results by providing enough data to concentrate on analysing data only at times of peak chlorophyll *a*.

11.1.3 Chlorophyll *a* vs. cover

Correlations between chlorophyll *a* and metrics of percentage cover were investigated between sites and within sites. The rationale for the analysis was that if there are strong relationships, then management of the environmental factors that affect chlorophyll *a* will apply to visual cover by periphyton in an eqivalent way.

We considered a range of of cover metrics (e.g., weighted composite cover (WCC) and the combination of Film, Mats and Fils in a multiple regression) and explored relationships between mean and maximum chlorophyll *a* and cover, between sites and within sites.

For the between-site analysis, predictive ability of the relationships with mean chlorophyll *a* varied across years and was often poor (NSE < 0.3); relationships using maximum values had consistently poor predictive power. Within sites, the multiple regression using Film, Mats and Fils produced the strongest relationships with chlorophyll *a*. 44% of sites had RSE > 0.55. All the sites with strong predictive ability were in the wider Manawatu catchment, or in the Ohau catchment, and did not include headwater sites.

Conclusion: Use of a single region-wide relationship to predict chlorophyll *a* from cover is unlikely to permit robust reporting of algal effects in terms of chlorophyll *a*. However, there is scope to use conversions at individual sites in the Manawatu and Ohau catchments. The next step for the Horizons data may be to explore possible classification of sites into groups with different relationships between cover and chlorophyll *a*.

11.2 Objective 2 – Establish if the resolution of sampling affects the performance of environmental drivers identified in Objective 1.

In addition to the monthly monitoring data, Horizons has collected periphyton data at fortnightly intervals at a subset of the monitoring sites. Data from 12 sites were used to compare the predictive ability of within-site models derived from fortnightly and monthly datasets at these sites. Data were available at each site from between 17 and 24 months.

The models from fortnightly data performed similarly to or better than the models using monthly data in all sites except at one site (ohau_gladstone) where monthly data predicted periphyton biomass more accurately than fortnightly.

The fortnightly data from the 12 sites generally did not yield stronger predictive relationships than using the complete (up to seven years) monthly dataset at the same sites (although we note that the datasets were not strictly comparable in numbers of samples or variables included). Poor performance of monthly data over the period of fortnightly surveys (17 to 24 months) may be attributable to low numbers of samples.

Conclusion: Fortnightly data in some cases could allow relationships to be developed over a shorter time period. Fortnightly datasets have other applications including more accurate estimation of accrual rates, and testing of mechanistic models of periphyton growth.

11.3 Objective 3 – Classify stations on the basis of their within-station environmental drivers, to permit generalisation of earlier driver findings, and comparison of findings to between-station inferences across all stations.

The objective of this analysis was to determine whether there was scope for groups of sites (i.e, belonging to particular classes) to be managed for periphyton biomass together by focussing on an influential factor (on chlorophyll *a*) common to the groups.

Sites were assigned to groups on the basis of a range of site and catchment characteristics, and the variables included in, or strength of, within-site relationships between chlorophyll *a* and environmental variables. We also considered the strenths of within-site relationships between chlorophyll *a* and periphyton cover.

The strongest pattern noted was that sites with strong within-site relationships between chlorophyll *a* and cover also had the stongest within-site relationships with environmental variables (including with days since an effective flow). These sites included most sites in large rivers, had higher DIN and finer sediment, and were in catchments with high proportions of their area in farmland and low proportions in indigenous forest, compared to the sites with weaker relationships. In addition, the strengths of the within-site relationships showed some patterns across catchment geology classes. Sites with AL and SS geology generally had stronger within-site relationships than those with HS or VA geology.

All the variables common to sites with the strongest within-site relationships were generally intercorrelated, so that it was not possible to identify a single factor that caused the pattern. The sites with strong and weak relationships did not differ in mean or peak chlorophyll *a*. It is suggested that the concentration of highest-strength within-site relationships at sites in larger, lowland rivers (with AL and SS geology) may reflect a more uniform community composition at those sites than in the smaller upland sites. This could be checked using existing data on species composition.

Conclusion: Overall, groups of sites based on different predictor variables included in the within-site chlorophyll *a* – environment relationships were not clearly distinguishable from their geology or Horizons-assigned LSC (life-supporting capacity) class. Therefore, we identified no basis for management of sites together, as suggested by the initial objective for the analysis.

12 Acknowledgements

This report was funded by DairyNZ and Horizons Regional Council, in a collaborative arrangement with NIWA, which included contributing to the report. We thank Brent Wood (Horizons) for supplying the hydrological data and Kathy Walter for initial data processing ready for the analysis. We very much appreciated advice from Doug Booker on data analysis techniques, and Clive Howard-Williams' constructive and thoughtful review comments.

13 References

Aertsen, W., Kint, V., van Orshoven, J., Ozkan, K., Muys, B. (2010) Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling*, 221(8): 1119-1130.

Augustinus, P., Reid, M., Andersson, S., Deng, Y., Horrocks, M. (2006) Biological and geochemical record of anthropogenic impacts in recent sediments from Lake Pupuke, Auckland City, New Zealand. *Journal of Paleolimnology,* 35(4): 789-805.

Ballantine, D.J., Davies-Colley, R.J. (2014) Water quality trends in New Zealand rivers: 1989-2009. *Environmental Monitoring and Assessment,* 186(3): 1939-1950.

Banet, A.I., Trexler, J.C. (2013) Space-for-Time Substitution Works in Everglades Ecological Forecasting Models. PLOS one, 8(11): 10.

Baulch, H.M., Turner, M.A., Findlay, D.L., Vinebrooke, R.D., Donahue, W.F. (2009) Benthic algal biomass - measurement and errors. *Canadian Journal of Fisheries and Aquatic Sciences,* 66(11): 1989-2001.

Bernot, M.J., Tank, J.L., Royer, T.V., David, M.B. (2006) Nutrient uptake in streams draining agricultural catchments of the midwestern United States. *Freshwater Biology*, 51(3): 499-509.

Biggs, B. (2000a) Eutrophication of streams and rivers: dissolved nutrient-chlorophyll relationships for benthic algae. *Journal of the North American Benthological Society*, 19: 17-31.

Biggs (2000b) New Zealand Periphyton guideline: detecting, monitoring and managing enrichment of streams. Ministry for the Environment, Wellington. 122 p.

Biggs, B., Stockseth, S. (1996) Hydraulic habitat suitability for periphyton in rivers. *Regulated Rivers: Research & Management,* 12: 251-261.

Biggs, B.J.F. (1995) The contribution of flood disturbance, catchment geology and land use to the habitat template of periphyton in stream ecosystems. *Freshwater Biology*, 33: 419-438.

Biggs, B.J.F., Close, M.E. (1989) Periphyton biomass dynamics in gravel bed rivers: the relative effects of flows and nutrients. *Freshwater Biology*, 22: 209 - 231.

Biggs, B.J.F., Gerbeaux, P.J. (1993) Periphyton development in relation to macro-scale (geology) and microscale (velocity) limiters in two gravel-bed rivers, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 27: 39-53.

Biggs, B., Goring, D., Nikora, V. (1998) Subsidy and stress responses of stream periphyton to gradients in water velocity as a function of community growth form. *Journal of Phycology,* 34: 598-607.

Biggs, B.J.F., Nikora, V.I., Snelder, T.H. (2005) Linking scales of flow variability to lotic ecosystem structure and function. *River Research and Applications,* 21(2-3): 283-298.

Biggs, B., Price, G. (1987) A survey of filamentous algal proliferations in New Zealand rivers. *New Zealand Journal of Marine and Freshwater Research*, 21(2): 175-191.

Blois, J.L., Williams, J.W., Fitzpatrick, M.C., Jackson, S.T., Ferrier, S. (2013) Space can substitute for time in predicting climate-change effects on biodiversity. Proceedings of the National Academy of Sciences of the United States of America, 110(23): 9374-9379.

Boeff, K.A., Strock, K.E., Saros, J.E. (2016) Evaluating planktonic diatom response to climate change across three lakes with differing morphometry. *Journal of Paleolimnology*, 56(1): 33-47.

Booker, D. (2013) Spatial and temporal patterns in the frequency of events exceeding three times the median flow (FRE3) across New Zealand. *Journal of Hydrology (NZ),* 52(1): 1-25.

Bothwell, M.L. (1988) Growth rate responses of lotic periphytic diatoms to experimental phosphorus enrichment: The influence of temperature and light. *Canadian Journal of Fisheries and Aquatic Sciences,* 45: 261 - 270.

Bothwell, M.L. (1989) Phosphorus-limited growth dynamics of lotic periphytic diaotm communities: areal biomass and cellular growth rate responses. *Canadian Journal of Fisheries and Aquatic Sciences,* 46: 1293-1301.

Boeff, K.A., Strock, K.E., Saros, J.E. (2016) Evaluating planktonic diatom response to climate change across three lakes with differing morphometry. Journal of Paleolimnology, 56(1): 33-47.

Brett, M. (2004) When is a correlation between non-independent variables "spurious"? Oikos, 105: 647-656.

Carlson, R. (1977) A trophic state index for lakes. *Limnology and Oceanography*, 22: 361-369.

Clausen, B., Biggs, B.J.F. (1997) Relationships between benthic biota and hydrological indices in New Zealand streams. *Freshwater Biology,* 38: 327 - 342.

Corman, J.R., Moody, E.K., Elser, J.J. (2016) Calcium carbonate deposition drives nutrient cycling in a calcareous headwater stream. *Ecological Monographs,* 86(4): 448-461.

Davie, A.W., Mitrovic, S.M., Lim, R.P. (2012) Succession and accrual of benthic algae on cobbles of an upland river following scouring. Inland Waters, 2(2): 89-100.

Dodds, W.K., Smith, V.H., Lohman, K. (2002) Nitrogen and phosphorus relationships to benthic algal biomass in temperate streams. *Canadian Journal of Fisheries and Aquatic Sciences,* 59(5): 865-874.

Dodds, W.K., Smith, V.H., Lohman, K. (2006) Nitrogen and phosphorus relationships to benthic algal biomass in temperate streams (vol 59, pg 865, 2002). *Canadian Journal of Fisheries and Aquatic Sciences*, 63(5): 1190-1191.

Dodds, W. (2003a) Misuse of inorganic N and soluble reactive P concentrations to indicate nutrient status of surface waters. *Journal of the North American Benthological Society,* 22: 171-181.

Dodds, W.K. (2003b) The role of periphyton in phosphorus retention in shallow freshwater aquatic systems. *Journal of Phycology,* 39(5): 840-849.

Dodds, W.K. (2007) Trophic state, eutrophication and nutrient criteria in streams. *Trends in Ecology & Evolution,* 22(12): 669-676.

Dodds, W.K., Welch, E.B. (2000) Establishing nutrient criteria in streams. *Journal of the North American Benthological Society*, 19(1): 186-196.

Earl, S.R., Valett, H.M., Webster, J.R. (2006) Nitrogen saturation in stream ecosystems. Ecology, 87(12): 3140-3151.

Ellison, M., Brett, M. (2006) Particulate phosphorus availability as a function of stream flow and land cover. *Water Research*, 40: 1258-1268.

Francoeur, S.N., Biggs, B.J.F., Smith, R.A., Lowe, R.L. (1999) Nutrient limitation of algal biomass accrual in streams: seasonal patterns and a comparison of methods. *Journal of the North American Benthological Society*, 18(2): 242-260.

Francoeur, S.N., Biggs, B.J.F. (2006) Short-term effects of elevated velocity and sediment abrasion on benthic algal communities. *Hydrobiologia,* 561: 59-69.

Geyer, C.J. (2003) Model selection in R. http://www.stat.umn.edu/geyer/5931/mle/sel.pdf

Gao, P., Zhang, L.J. (2016) Determining spurious correlation between two variables with common elements: event area-weighted suspended sediment yield and event mean runoff depth. *Professional Geographer*, 68(2): 261-270.

Goldsmith, S.T., Carey, A.E., Lyons, W.B., Hicks, D.M. (2008) Geochemical fluxes and weathering of volcanic terrains on high standing islands: Taranaki and Manawatu-Wanganui regions of New Zealand. *Geochimica Et Cosmochimica Acta,* 72(9): 2248-2267.

Graham, M.H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, 84(11), 2809-2815.

Haiddeker, S., Wade, O. (2016) Tutaekuri and Ngaruroro Catchments: Nutrient Limitation of Algal Growth. HBRC Report No. RM16-06 – 4783

Harding, J.S., Young, R.G., Hayes, J.W., Shearer, K.A., Stark, J.D. (1999) Changes in agricultural intensity and river health along a river continuum. *Freshwater Biology,* 42(2): 345-357.

Hart, D.D., Biggs, B.J.F., Nikora, V.I., Flinders, C.A. (2013) Flow effects on periphyton patches and their ecological consequences in a New Zealand river. *Freshwater Biology,* 58(8): 1588-1602.

Hoyle, J.T., Kilroy, C., Hicks, D.M., Brown, L. (2017) The influence of sediment mobility and channel geomorphology on periphyton abundance. *Freshwater Biology,* 62(2): 258-273.

Huang, J.C., Gao, J.F., Mooij, W.M., Hormann, G., Fohrer, N. (2014) A Comparison of Three Approaches to Predict Phytoplankton Biomass in Gonghu Bay of Lake Taihu. *Journal of Environmental Informatics,* 24(1): 39-51.

Jarvie, H.P., Withers, P.J.A., Neal, C. (2002) Review of robust measurement of phosphorus in river water: sampling, storage, fractionation and sensitivity. *Hydrology and Earth System Sciences,* 6(1): 113-131.

Jellyman, P.G., Harding, J.S. (2016) Disentangling the stream community impacts of *Didymosphenia geminata*: How are higher trophic levels affected? *Biological Invasions,* 18(12): 3419-3435.

Kasprzak, P., Padisak, J., Koschel, R., Krienitz, L., Gervais, F. (2008) Chlorophyll *a* concentration across a trophic gradient of lakes: An estimator of phytoplankton biomass? *Limnologica,* 38(3-4): 327-338.

Katz, S.B., Segura, C., Warren, D.R. (2018) The influence of channel bed disturbance on benthic Chlorophyll *a*: A high resolution perspective. *Geomorphology,* 305: 141-153.

Keck, F., Lepori, F. (2012) Can we predict nutrient limitation in streams and rivers? *Freshwater Biology,* 57(7): 1410-1421.

Kilroy, C., Biggs, B., Vyverman, W., Broady, P. (2006) Benthic diatom communities in subalpine pools in New Zealand: relationships to environmental variables. *Hydrobiologia*, 561: 95-110.

Kilroy, C. (2017) Periphyton - nutrient relationships in rivers: A literature review and New Zealand perspective. For DairyNZ. *NIWA Client Report* No: 2016113CH. 79 p.

Kilroy, C., Booker, D.J., Drummond, L., Wech, J.A., Snelder, T.H. (2013) Estimating periphyton standing crop in streams: a comparison of chlorophyll *a* sampling and visual assessments. *New Zealand Journal of Marine and Freshwater Research,* 47(2): 208-224.

Kilroy, C., Wech J.A., Kelly, D., Clarke, G. (2017) Analysis of a three-year dataset of periphyton biomass and cover in Canterbury Rivers. For Environment Canterbury. *NIWA Client Report* No: 2017085CH. 112 p.

Kilroy, C., Wech, J. (2015) Spatial and temporal patterns in nutrient concentrations and periphyton in the Hurunui River. *NIWA Client Report* CHC2015_086. For Ngai Tahu Forest Estates Ltd. 96 p.

Kilroy, C., Wech, J., Chakraborty, M.; Brown, L.; Watson, B., Patterson, M. Patterson, M., Shell, T. (2016) Periphyton in the Manawatu - Whanganui region. State, trends and seasonality, 2009 - 2015. *NIWA Client Report* CHC2016_013. For Horizons Regional Council. 160 p.

Kilroy, C.; Wech, J.; Booker, D.; Brown, L.; Chakraborty, M.; Nicholson, C.; Patterson, M.; Markham, M.; Roygard, J. (2012). Periphyton in the Manawatu-Whanganui region: review of three years of monitoring. *NIWA Client Report* CHC2012-105. Prepared for Horizons Regional Council.

Kim, J., Furumai, H. (2013) Improved calibration of a rainfall-pollutant-runoff model using turbidity and electrical conductivity as surrogate parameters for total nitrogen. *Water and Environment Journal,* 27(1): 79-85.

Klose, K., Cooper, S.D., Leydecker, A.D., Kreitler, J. (2012) Relationships among catchment land use and concentrations of nutrients, algae, and dissolved oxygen in a southern California river. *Freshwater Science,* 31(3): 908-927.

Lake, L., Li, Y.L., Casal, J.J., Sadras, V.O. (2016) Negative association between chickpea response to competition and crop yield: Phenotypic and genetic analysis. *Field Crops Research,* 196: 409-417.

Larned, S.T. (2010) A prospectus for periphyton: recent and future ecological research. *Journal of the North American Benthological Society,* 29(1): 182-206.

Larned, S.T., Snelder, T., Unwin, M.J., McBride, G.B. (2016) Water quality in New Zealand rivers: current state and trends. *New Zealand Journal of Marine and Freshwater Research*, 50(3): 389-417.

Leclere, J., Oberdorff, T., Belliard, J., Leprieur, F. (2011) A comparison of modeling techniques to predict juvenile 0+fish species occurrences in a large river system. *Ecological Informatics,* 6(5): 276-285.

Lessard, J., Hicks, D.M., Snelder, T.H., Arscott, D.B., Larned, S.T., Booker, D., Suren, A.M. (2013) Dam Design can Impede Adaptive Management of Environmental Flows: A Case Study from the Opuha Dam, New Zealand. *Environmental Management,* 51(2): 459-473.

Lewis, W.M., McCutchan, J.H. (2010) Ecological responses to nutrients in streams and rivers of the Colorado mountains and foothills. *Freshwater Biology,* 55(9): 1973-1983.

Liess, A., Le Gros, A., Wagenhoff, A., Townsend, C.R., Matthaei, C.D. (2012) Landuse intensity in stream catchments affects the benthic food web: consequences for nutrient supply, periphyton C:nutrient ratios, and invertebrate richness and abundance. *Freshwater Science,* 31(3): 813-824.

Li, G.Y., Han, H.Y., Du, Y., Hui, D.F., Xia, J.Y., Niu, S.L., Li, X.N., Wan, S.Q. (2017) Effects of warming and increased precipitation on net ecosystem productivity: A long-term manipulative experiment in a semiarid grassland. *Agricultural and Forest Meteorology,* 232: 359-366.

Likens, G.E., Bormann, F.H., Johnson, N.M., Fisher, D.W., Pierce, R.S. (1970) Effects of forest cutting and herbicide treatment on nutrient budgets in Hubbard Brook watershed ecosystem. *Ecological Monographs,* 40(1): 23-

Lutz, B.D., Mulholland, P.J., Bernhardt, E.S. (2012) Long-term data reveal patterns and controls on stream water chemistry in a forested stream: Walker Branch, Tennessee. *Ecological Monographs,* 82(3): 367-387.

Ma, J., Xiao, X.M., Bu, R.C., Doughty, R., Hu, Y.M., Chen, B.Q., Li, X.P., Zhao, B. (2017) Application of the space-for-time substitution method in validating long-term biomass predictions of a forest landscape model. *Environmental Modelling & Software,* 94: 127-139.

Matheson, F., Quinn, J., Hickey, C. (2012) Review of the New Zealand instream plant and nutrient guidelines and development of an extended decision-making framework: Phases 1 and 2 final report. Prepared for the Ministry of Science & Innovation Envirolink Fund. *NIWA Client Report* HAM2012-081. http://www.envirolink.govt.nz/PageFiles/31/Review of the NZ instream plant and nutrient guidelines and development of an extended decision making framework.pdf

McArthur, K.J., Roygard, J., Clark, M. (2010) Understanding variations in the limiting nitrogen and phosphorus status of rivers in the Manawatu-Wanganui Region, New Zealand. *Journal of Hydrology (NZ)* 49 (1): 15-33.

McDowell, R.W., Larned, S.T., Houlbrooke, D.J. (2009) Nitrogen and phosphorus in New Zealand streams and rivers: control and impact of eutrophication and the influence of land management. *New Zealand Journal of Marine and Freshwater Research,* 43(4): 985-995.

MfE (2017) Draft Guide to Attributes in Appendix 2 of the National Policy Statement for Freshwater Management 2014 (as amended 2017). Wellington 43pp.

Mulholland, P.J., Helton, A.M., Poole, G.C., Hall, R.O., Hamilton, S.K., Peterson, B.J., Tank, J.L., Ashkenas, L.R., Cooper, L.W., Dahm, C.N., Dodds, W.K., Findlay, S.E.G., Gregory, S.V., Grimm, N.B., Johnson, S.L., McDowell, W.H., Meyer, J.L., Valett, H.M., Webster, J.R., Arango, C.P., Beaulieu, J.J., Bernot, M.J., Burgin, A.J., Crenshaw, C.L., Johnson, L.T., Niederlehner, B.R., O'Brien, J.M., Potter, J.D., Sheibley, R.W., Sobota, D.J., Thomas, S.M. (2008) Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature,* 452(7184): 202-U246.

Mundry, R., Nunn, C.L. (2009) Stepwise model fitting and statistical inference; turning noise into signal pollution. *The American Naturalist* 173(1): 119-123.

Munn, M., Black, R., Gruber, S. (2002) Response of benthic algae to environmental gradients in an agriculturally dominated landscape. *Journal of the North American Benthological Society*, 21: 221-237.

Munn, M., Frey, J., Tesoriero, A. (2010) The Influence of Nutrients and Physical Habitat in Regulating Algal Biomass in Agricultural Streams. *Environmental Management,* 45(3): 603-615.

Niyogi, D.K., Koren, M., Arbuckle, C.J., Townsend, C.R. (2007) Stream communities along a catchment land-use gradient: Subsidy-stress responses to pastoral development. *Environmental Management,* 39(2): 213-225.

New Zealand Government (2017) National Policy Statement for Freshwater Management 2014 (Amended 2017). http://www.mfe.govt.nz/publications/fresh-water/national-policy-statement-freshwater-management-2014-amended-2017.

Oppel, S., Meirinho, A., Ramirez, I., Gardner, B., O'Connell, A.F., Miller, P.I., Louzao, M. (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation,* 156: 94-104.

Peterson, B.J., Wollheim, W.M., Mulholland, P.J., Webster, J.R., Meyer, J.L., Tank, J.L., Marti, E., Bowden, W.B., Valett, H.M., Hershey, A.E., McDowell, W.H., Dodds, W.K., Hamilton, S.K., Gregory, S., Morrall, D.D. (2001) Control of nitrogen export from watersheds by headwater streams. *Science,* 292(5514): 86-90.

Pineiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M. (2008) How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, 216(3-4): 316-322.

Potapova, M., Charles, D. (2003) Distribution of benthic diatoms in US rivers in relation to conductivity and ionic composition. *Freshwater Biology,* 48: 1311-1328.

Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* 9:181–199.

Redfield A.C. (1958) The biological control of chemical factors in the environment. *American Scientist,* 46, 205–221.

Rier, S.T., Stevenson, R.J. (2006) Response of periphytic algae to gradients in nitrogen and phosphorus in streamside mesocosms. *Hydrobiologia,* 561: 131-147.

Rode, M., Halbedel Angelstein, S., Anis, M.R., Borchardt, D., Weitere, M. (2016) Continuous in-stream assimilatory nitrate uptake from high frequency sensor measurements. Environmental Science & Technology, 50(11): 5685-5694.

Rott, E., Schneider, S.C. (2014) A comparison of ecological optima of soft-bodied benthic algae in Norwegian and Austrian rivers and consequences for river monitoring in Europe. *Science of the Total Environment,* 475: 180-186.

Rutherford, J.C. (2013) Effects of land use on streams – Phase 2 modelling studies in the Tukituki River, Hawke's Bay. Prepared for Hawke's Bay Regional Investment Company Limited. *NIWA Client Report No:* HAM2013-026.

Rutherford, J.C., Scarsbrook, M.R., Broekhuizen, N. (2000) Grazer control of stream algae: Modeling temperature and flood effects. *Journal of Environmental Engineering-Asce,* 126(4): 331-339.

Schneider, S.C., Petrin, Z. (2017) Effects of flow regime on benthic algae and macroinvertebrates - A comparison between regulated and unregulated rivers. *Science of the Total Environment,* 579: 1059-1072.

Sharma, S., Walker, S.C., Jackson, D.A. (2008) Empirical modelling of lake water-temperature relationships: a comparison of approaches. *Freshwater Biology,* 53(5): 897-911.

Snelder, T., Biggs, B.J.F., Kilroy, C., Booker, D. (2013) National Objective Framework for periphyton. *NIWA Client Report* CHC2013_122. For Ministry for the Environment. 39 p.

Snelder, T., Booker, D., Unwin, M., Wood, S., Wilcock, B. (2014). State and Trends of River Water Quality in the Manawatū River Catchment. Horizons Report 2014/EXT/1364. 69 p.

Stevenson, R.J., Hill, B.H., Herlihy, A.T., Yuan, L.L., Norton, S.B. (2008) Algae-P relationships, thresholds, and frequency distributions guide nutrient criterion development. *Journal of the North American Benthological Society,* 27(3): 783-799.

Suplee, M.W., Watson, V., Dodds, W.K., Shirley, C. (2012) Response of algal biomass to large-scale nutrient controls in the Clark Fork River, Montana, United States. *Journal of the American Water Resources Association*, 48(5): 1008-1021.

Suren, A., Biggs, B., Kilroy, C., Bergey, E. (2003) Benthic community dynamics during summer low-flows in two rivers of contrasting enrichment 1. Periphyton. *New Zealand Journal of Marine and Freshwater Research* (37): 53-70.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement,* 55, 525-534.

Verberg, P. (2016) Nutrient limitation of algal biomass in the Waikato River. NIWA Client Report HAM2016-053. Prepared for the Technical Leaders Group of the Waikato – Waipa Rivers Wai Ora Project. 227 p.

Vilches, C., Giorgi, A., Casco, M.A. (2013) Periphyton responses to non-point pollution in naturally eutrophic conditions in Pampean streams. *Fundamental and Applied Limnology*, 183(1): 63-74.

Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P. (2006) Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, 75(5): 1182-1189.

Withers, P.J.A., Jarvie, H.P. (2008) Delivery and cycling of phosphorus in rivers: *A review. Science of the Total Environment,* 400(1-3): 379-395.

Wood, S.A., Depree, C., Brown, L., McAllister, T., Hawes, I. (2015) Entrapped Sediments as a Source of Phosphorus in Epilithic Cyanobacterial Proliferations in Low Nutrient Rivers. *PLOS one,* 10(10): 17.

Appendix A Requirements for the analysis (from the contract)

The following is the proposed approach to address objectives (1) to (3). Each objective was required to be separately addressed in the report.

Objective 1 – Establish the significance and strength of relationships between environmental factors and periphyton standing crop (max or 92nd% Chl-a between stations, observed Chl-a within station time-series).

a. Determine hydrological sensitivity of periphyton standing crop at each station by exploring within-station relationships between Chl-a and flow. At each station compute a series of flow metrics (days since *n**median flow where *n* = 1.5, 2, 3, 5, 7, 9, 10, 12 and 15), using the 2000-2015 interval for estimating station median flow. Regress observed Chl-a against each metric and identify the threshold of *n* with greatest R^2 and a significant relationship (p<0.05). N x median in the relationship with highest R^2 is interpreted as the most effective (or optimum) flow threshold for removing periphyton at that site, provided that the relationship is statistically significant (P < 0.05). Assign each station to one of the following categories: "Stable" (high flow threshold, e.g., 7 x median or greater, with significant relationship), "Unstable" (low flow threshold, e.g., 5 x median or lower, with significant relationship), "Mixed" (where high and low n-thresholds perform equivalently in terms of R2) and "Unresponsive" (where neither high nor low n-thresholds are significant or possess R2>0.4). (Note that these definitions may be adjusted in light of the results of the analysis.)

b. Between-station variance to be explored using Chl-a as response :

1. Perform correlation analyses between periphyton response variables (e.g., Chl-a, % cover) to determine whether environmental variables for max or 92nd% Chl-a will be shared across other periphyton biomass metrics (i.e., whether subsequent driver information on Chl-a will be of use in managing % cover).
2. Use simple linear regression to test for relationships between observed Chl-a (annual or multi-year max or 92nd%) against preceding average DIN and DRP (e.g., 4-month means), or annual or multi-year averages of DIN and DRP.
3. Test the Biggs (2000) DIN, DRP and Days of accrual (D_a) equations for each water-year, ensuring that max Chl-a for a station is paired against that water-year's average conditions. Calculate D_a as in Biggs (2000) [365 days divided by average annual number of freshes exceeding 3 x median]. Then recalculate D_a as 365 days divided by the average annual number of freshes exceeding *n* x median, where *n* is the optimum value identified for each station in (1a). Inspect output of both methods. Use classical DIN/DRP theory to classify stations into N-limited (e.g., DIN : DRP <8), P-limited (>15) and co-limited (>8 <15) groups. Determine whether the Biggs (2000) DIN or DRP equations performed significantly differently across the three N:P groups, using 1-way, 3-factor ANOVA.
4. Multiple stepwise linear regression – undertake variance inflation filtering beforehand to determine which environmental drivers to drop due to collinearity, then examine suite of multiple instream and catchment predictors of Chl-a (max or 92nd%), for:
    1. 3-year interval (2012-2015)
    2. 6-year interval (2009-2015)
    3. Stable, Unstable and Mixed stations separately
    (Suite of environmental predictors to include: DIN, DRP, FREn, TDP, TN, TP, TSS, conductivity, %substrate, %native bush, % exotic high producing grassland, %exotic grassland)
5. Quantile multiple stepwise regression – utilise the 75th%, 85th% and 95th% for a response (amongst the integrated station estimates of the max or 92nd% observed Chl-a) determining if predictive performance can be improved on 1b2, 1b3, and 1b4 above.

c. Within-station variance along the full gradient of Chl-a recorded by each station (largely as above, only utilising variation in Chl-a and environment at each station, over time – acknowledging likelihood of pseudo-replication):

1. Paired correlations between periphyton response variables (e.g., Chl-a, % cover) – to determine whether driver analysis will be consistent across multiple periphyton metrics at each station (i.e., whether subsequent driver information on Chl-a will be of use in managing % cover). Determine if stations can be grouped on account of the apparent or lack of consistent pattern between periphyton biomass indicators, before then commenting on whether drivers of max or 92nd% Chl-a at groups of stations are shared across periphyton indicators.

2. Simple linear regression – testing observed Chl-a against preceding average environmental conditions for 4 months immediately prior to develop simple linear regressions to DIN and DRP.
3. Multiple stepwise linear regression – undertake variance inflation filtering beforehand to determine which environmental drivers to drop due to collinearity (at each station), then examine suite of multiple instream and catchment predictors of observed Chl-a, for:
    1. 6-year interval (2009-2015)
       (Suite of environmental predictors to include: DIN, DRP, FREn, TDP, TN, TP, TSS, conductivity,– use accrual days as predictor, determined from best-performing n threshold in 1a)
       (Consider whether to include a decimal time factor to account for autocorrelation – e.g., convert date to decimal time, include as continuous variable. Won't interfere with objective as should not be correlated to environmental variance except where there is a strong monotonic trend component to any driver).
4. Quantile multiple stepwise regression – utilise the 75^{th}%, 85^{th}% and 95^{th}% for a response (within each station time-series over the 6-year interval) determining if predictive performance can be improved on 1bb and 1bc, above.

Objective 2 – Establish if the resolution of sampling affects the performance of environmental drivers identified in Objective 1.

1. For subset of stations with fortnightly sampling of periphyton biomass, repeat Objective 1c, for within-station analyses **only** but using the fortnightly equivalent information on environmental condition and periphyton biomass, contrasting the findings on the significance and strength of predictors determined with output from Objective 1.
2. If differences arise, reflect on whether the environmental indicators identified as differing in their performance between fortnightly or monthly, are likely to vary at fortnightly or shorter intervals – if so, than likelihood being that monthly-based regressions are unlikely to record the actual effects of those environmental predictors.

Objective 3 – Classify stations on the basis of their within-station environmental drivers, to permit generalisation of earlier driver findings, and comparison of findings to between-station inferences across all stations.

In each section below, output is to discuss whether groupings suggest likelihood of being able to manage periphyton biomass through focus on one or more drivers at a group of stations, and if so, whether those groupings correspond to some other likely factor (e.g., geology, climatic type, stream order, upland/lowland, point-source affected, FMU's, by river catchment).

1. Visual inspection of summary statistics on power and significance of drivers at each station, over 3- and 6-year intervals, and by earlier findings of Objective 1 (e.g., on basis of hydrological sensitivity);

Classify sites quantitatively. Ordinate sites on matrix of drivers coded into binary responses (1=significant; 0=insignificant) and visually inspect as above to see if sites with similar mix of significant drivers correspond with a-priori classes (e.g., on basis of hydrological sensitivity, climate, geology, stream order, upland/lowland, point-source affected). Alternatively, cluster on binary matrix of driver significance, and compare cluster groupings (leaves) to a-priori coded classes (e.g., hydrological sensitivity, climate…) in a confusion table.

Appendix B Notes on statistical approaches

The scope and contract for this work specified the statistical approaches to be used for analysing the data (see Appendix A). As the work proceeded, we have reviewed the suggested methods and in some cases revised the approach, considering (a) the features of the dataset, and (b) literature related to the methods. The following is a brief description of the statistical approaches specified. We then describe issues with the techniques (including those encountered during the analysis) and how we dealt with them.

Linear regression

The basic approach to identifying relationships was to focus on linear regression techniques. Linear regression (with its variants, including multiple linear regression) is one of the oldest techniques used by ecologists to build predictive relationships between biota and environmental variables (Guisan et al. 2002). Although many other techniques are now available[9], linear regression still has utility because of its ease of use and interpretation compared with more complex techniques (Aertsen et al. 2010, Huang et al. 2014), and superior predictive ability in some cases (Sharma et al. 2008). Nevertheless, the requirements and assumptions of linear regression limit its application. These include: (a) normally distributed residuals (difference between observed value of the predicted variable and the predicted value); (b) homoscedasticity in the variance of the independent variable (i.e., the same amount of scatter around the mean value); (c) a linear relationship between the dependent and explanatory variables, which in many cases is not supported by data and observations; (d) no collinearity among the independent variables.

Requirements (a) and (b) can often be overcome by appropriate transformations of the data. Assumption (c) can also be addressed by data transformation, or the addition of polynomial or interaction terms to the model so that the dependent variable can display a non-linear pattern as the parameter increases linearly. Collinearity (d) can make it difficult to identify optimal sets of explanatory variables from a range of candidate environmental parameters and can mask the effects of strong predictors (Graham 2003).

The contract specified stepwise multiple regression to identify relationships between periphyton and environmental variables. Removing predictors in a stepwise manner has some particular difficulties (see below).

Stepwise multiple linear regression

When datasets comprise biological observations along with a broad suite of potential explanatory variables, stepwise multiple regression can be used to try to identify the combinations of variables that best explain the observations. In stepwise regression, significant explanatory variables are in turn added (in forward selection) or non-significant terms are taken out (in backward removal) until a single, supposedly optimal, model is arrived at. Statistics assessing model fit, such as Akaike Information Criteria (AIC), are used at each step to assess whether terms improve model fit and thus should be retained or removed. The procedure is available in most statistical packages and is commonly applied in ecological studies. However, the drawbacks of stepwise multiple regression

[9] Relatively recently developed statistical techniques for the analysis of large and complex datasets such as the Horizons periphyton dataset include classification and regression trees (De'ath & Fabricius 2000) and their variants (boosted regression trees (BRT), random forests (RF)) and Artificial Neural Networks (ANN). These methods have the advantage that no distributions are assumed for the data and therefore transformations need not be applied. Machine-learning techniques can outperform traditional regression methods (Cunningham et al. 2009, Leclere et al. 2011). Different techniques have been applied and compared in a range of studies (e.g., Segurado & Araujo 2004, Aertsen et al. 2010). It is worth noting that their use does not guarantee that a strong or usable model will be identified (Oppel et al. 2012).

have been known for decades (e.g., Hocking 1976) and numerous authors have recommended that it should not be used (James and McCulloch 1990, Whittingham et al. 2006, Mundry and Nunn 2009).

The problems include:

1. inflation of Type 1 errors (inferring an effect when there is none) through inflating statistical significance by ignoring the fact that multiple tests are run (at each step) using the same data (this problem gets worse as the number of explanatory variables increases, especially if variables are correlated, Whittingham et al. 2006);
2. inability to select the true "best subset" and to show that alternative subsets of variables may provide solutions with similar explanatory power (refer to Whittingham et al. 2006, and Mundry and Nunn 2009, for more details).

Both problems are reduced when the sample size is very large, the effects of the predictors on the dependent variable are strong, and the number of predictor variables is small (Thompson 1995). The second problem (selection of an appropriate model) can be solved by using an information theoretic (IT) approach in which complete multiple linear regression is run on all subsets of the candidate variables, and the outputs are ranked according to various criteria that indicate an optimum or "best" model. Criteria include R^2 and adjusted R^2, the Akaike Information Criterion (AIC, Akaike 1974), Mallows' C_p (see Geyer 2003 for an explanation of each).

In the present analyses, the two main problems with stepwise multiple regression were minimised by meeting two of the criteria suggested by Thomson (1995) (see above): strong effects and small number of predictor variables. In addition, we used an IT approach to reconfirm that the stepwise approach did indeed select optimal models.

It is important to realise that identifying the best model using an IT approach in no way increases the chances that the relationship represents cause and effect. In fact, no technique can distinguish cause-effect relationships from correlations (Graham 2003), particularly when multiple predictor variables are correlated. In the present analysis we suspect that there will be an element of cause and effect because we are certain that flows and nutrients influence periphyton chlorophyll *a* (see Section 4.1). However, the main purpose of the regression analyses is for prediction rather than explanation. For that reason, we allow inclusion of other variables (e.g., conductivity; see Section 7.6.2) that improve predictability, even though their precise effect on periphyton abundance may not be understood.

Collinearity and variance inflation factors

The problems associated with collinearity in multiple regression cannot easily be overcome, although techniques exist which can at least quantify the independent contribution of correlated variables in explaining the dependent variable (Graham 2003).

The simplest procedure for reducing collinearity is to first check the correlations between all pairs of candidate explanatory variables, and include only one variable from each set of strongly correlated variables. In addition, variables can be restricted to those known (from the literature) to directly or indirectly affect the dependent variable. However, it is possible for pairwise correlations to be small but for there to be a linear dependence among three or more variables. The existence of multi-variate interdependence is why use of variance inflation factors (VIF) is often preferred over pairwise correlations for detecting collinearity. A VIF is calculated for each explanatory variable using the r-squared value of the regression of that variable against all other explanatory variables. The definition of ‘high’ VIF, indicating collinearity, is somewhat arbitrary but values in the range of 5-10 are commonly used to exclude parameters.

The contract asked for variance inflation factors (VIF) to be calculated before the stepwise regressions to determine which environmental drivers to drop due to collinearity. However, VIF is a diagnostic tool done post-estimation, i.e., calculated within models, and will vary as terms are added and removed throughout a stepwise procedure. VIF is calculated once the final model is determined, and does not help with *a priori* deciding the parameters to include in a stepwise analysis.

While there are methods that use VIF values as criterion for automatically excluding parameters in a stepwise fashion, doing this removes the variables automatically and may remove parameters that should be retained because they are biologically sensible. Because of the large number of models in our analyses, and the use of the stepwise procedure, it was more efficient to determine parameters to include prior to the analyses. Thus, we used pairwise correlations prior to analyses to *a priori* determine variables to include. Correlation matrices were constructed for each subset of data (i.e., all year-flow category combinations). Where strong pairwise correlations were detected we omitted one of the pair of correlated variables. As a final check we conducted VIF analysis on the final best models in the between sites analysis (see Section7). None of the variables retained in the final best model had a VIF >2.5, indicating little collinearity in the models.

Quantile regression

Linear regression compares the mean of values of a dependent variable along a gradient of independent (or predictor) variable values across sites. Quantile regression is similar but uses quantiles, and can address whether, for example, the 25th or 75th quantile behaves differently from the mean. Like linear regression, quantile regression produces regression coefficients that estimate an independent variable’s effect on a specified quantile of the dependent variable. In the present analysis, quantile regression could address the question: are different parameters correlated with (and therefore potentially influencing) chlorophyll *a* at sites with low or high periphyton biomass? Quantile regression is a useful method for trying to identify the main drivers of periphyton because multiple environmental factors affect periphyton standing crop, creating ‟noisy” datasets, environmental drivers may differ at sites with high and low periphyton biomass and because the method is less sensitive to non-normal errors and outliers than linear regression

Quantile regression uses the full data set for estimating the relation between predictor variables and the response when fitting quantiles across the range of the predictor variables. Regular linear regression provides an estimate of how much predictor variable X affects the mean outcome of response Y. Quantile regression on the median (50th percentile) provides an example of how much covariate X affects the median of response Y. The full data distribution is required to determine the position of the median (i.e., the point at which %50 of Y is above and 50% is below). The same reasoning holds for other percentiles, e.g., 25th or 75th percentiles. Estimating a 75th quantile regression fits a line such that 25% of the points are above the line and 75% are below.

Trials of stepwise quantile regression run on subsets of the between sites datasets showed that the results for the 50th quantile (median) were very similar to linear regression (mean) – i.e., the stepwise procedure retained the same parameters and they had similar coefficients. However, when the stepped quantile procedure was run on higher quantiles (e.g.,75th, 85th, 95th) all of the entered predictor variables were retained.

The contract requested investigating whether quantile regression increased predictive performance of the models over stepwise linear regression. Taking into account the results of the stepped quantile regression trials, we settled on the following approach for the between-sites analysis:

1. run stepwise linear models on all hydrological year-flow category combinations of sites, with maximum chlorophyll *a* (annual datasets) and 92nd percentile chlorophyll *a* (multi-year datasets) as the dependent variable;

2. explore some subsets of year-flow category combinations to investigate whether stepwise quantile regression removes any variables. Generally, the stepped procedure did not remove variables at the quantiles we were tasked with investigating (75th 85th, 95th). So we proceeded with:

    1. non-stepped quantile regressions with graphs of model coefficients in comparison with coefficients from linear regression models across all quantiles of:

        a. the best simplified models identified by stepped linear regression and best subsets analysis for each temporal dataset (annual, 3-year and 7-year datasets) (from Table 6-4);

        b. models with all predictor variables originally included in the stepped regressions.

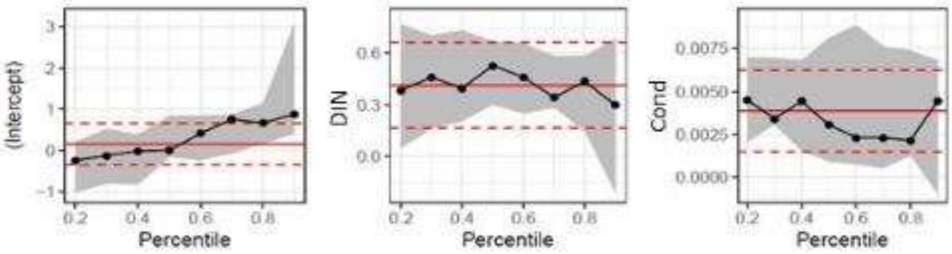
Results of quantile regressions on the best simplified model for each temporal dataset are shown below (Figure B-1). Results from quantile regressions with all predictor variables included were very similar to those from the simplified models and are not presented here.

Coefficients from quantile regressions using the same variables identified from stepped multiple linear regressions were rarely significantly different from coefficients generated from an ordinary least squares (OLS) linear regression for all the temporal datasets (Figure B-1). When coefficient estimates for quantile regressions did occur outside the 95% confidence intervals for the linear regression coefficients it was generally only at the highest quantile (0.9 or 90th percentile). For lower percentiles, the quantile coefficient was almost always within the 95% CI of the linear regression coefficient estimate. This indicates the same linear equation can explain the relationship between chlorophyll *a* biomass and the predictor variables for sites with both high and low levels of chlorophyll *a* biomass.

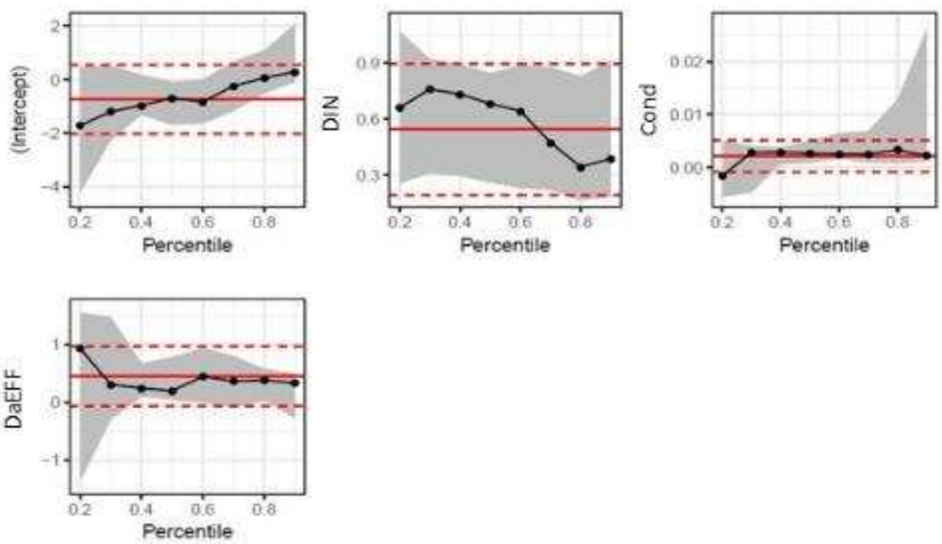
Plots on the following five pages:

Figure B-1: Coefficients from quantile regressions using variables from the best stepped linear regression for all temporal datasets. Each plot is for one term in the model (i.e., the intercept and a slope term associated with each predictor). The y axis is the coefficient value for each term and the x axis indicates individual regressions at different quantiles (0.2 to 0.9 percentiles). Quantile regression coefficients are estimated for each quantile (percentile from 0.2 to 0.9; black dots, with associated upper and lower bounds calculated using a rank inversion method; grey shading). Red solid and dashed lines indicate the coefficient and 95% CI, respectively, from a OLS linear regression. Quantile regression coefficients can be assumed to be different from linear regression coefficients at a particular percentile when the black dots occur outside of the red dashed lines.

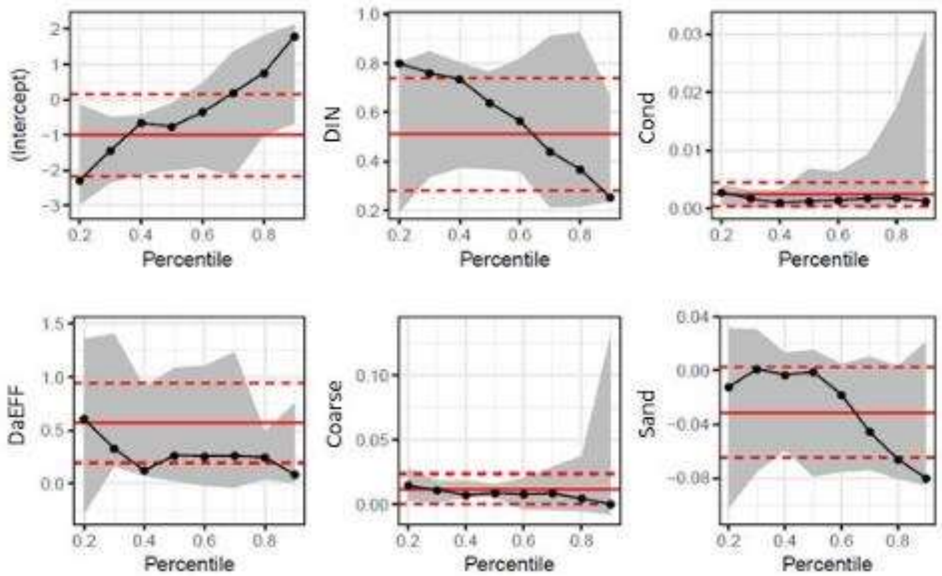
2009-10, n = 43



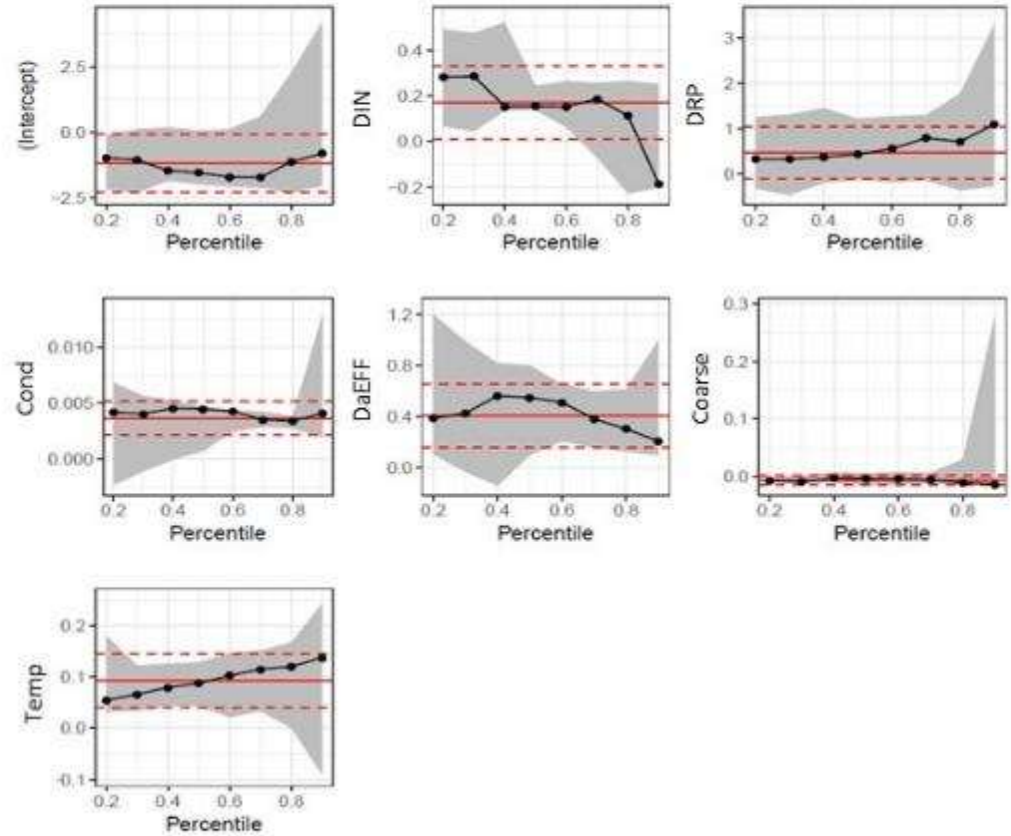
2010-11, n = 33



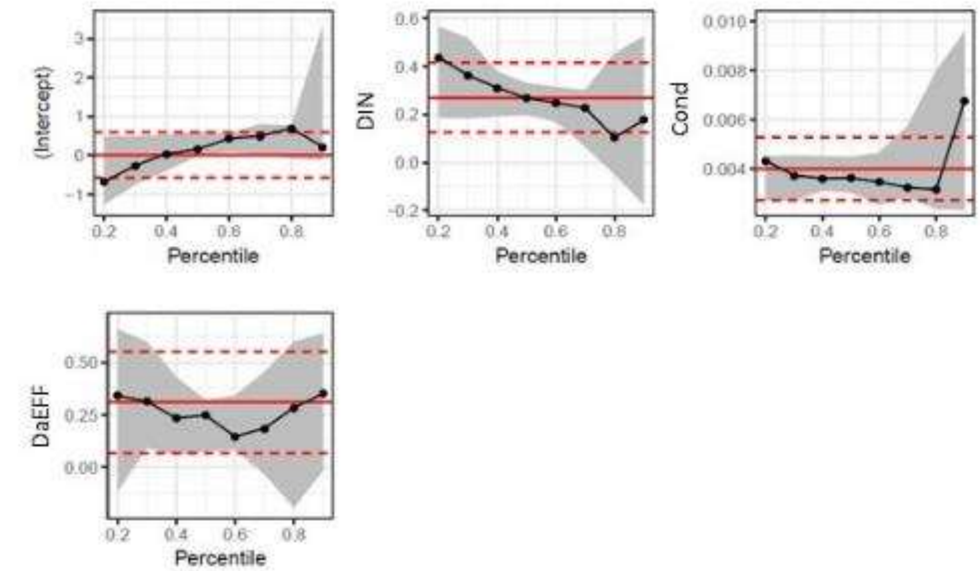
2011-12, n = 35

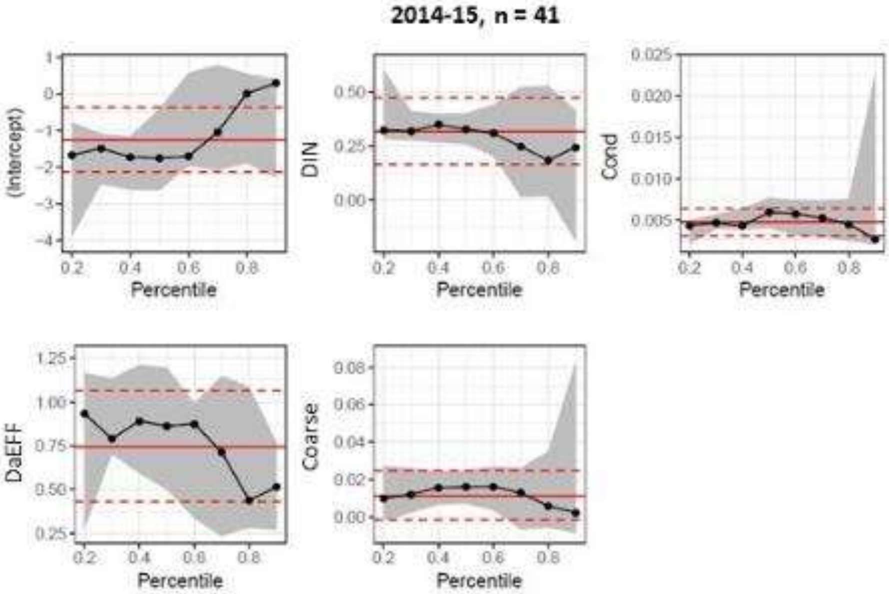


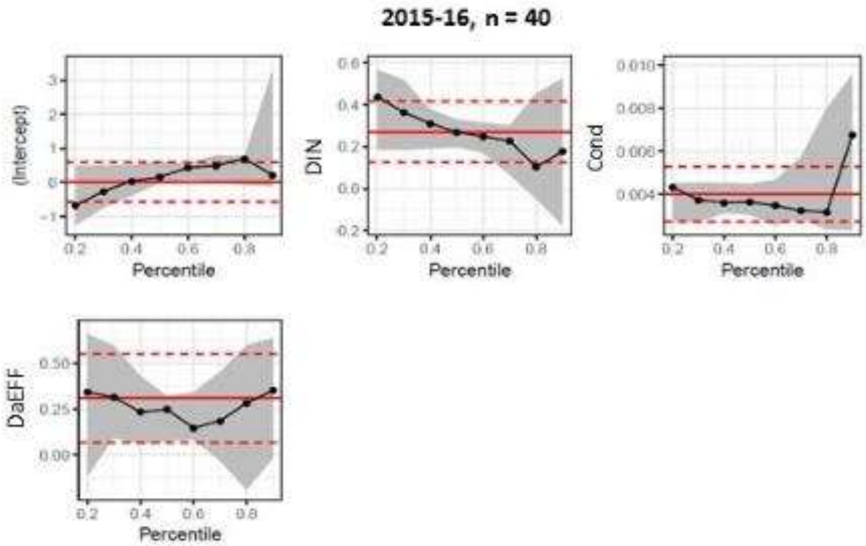
2012-13, n = 37

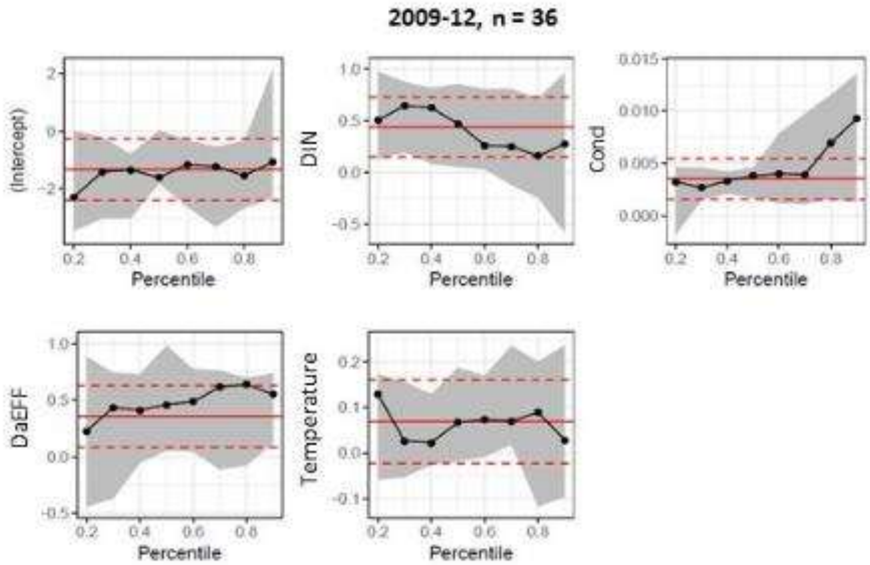


2013-14, n = 40

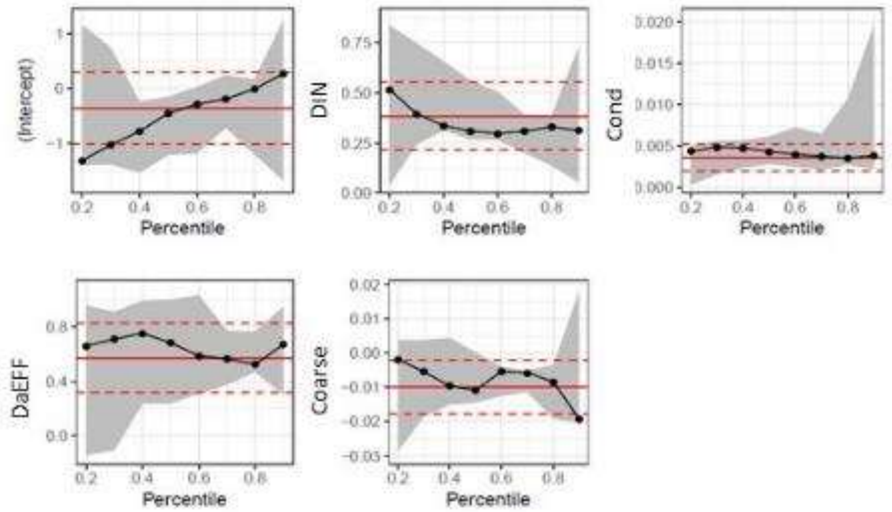



2014-15, n = 41
(Intercept)
DIN
Cond
DaEFF
Coarse
Percentile

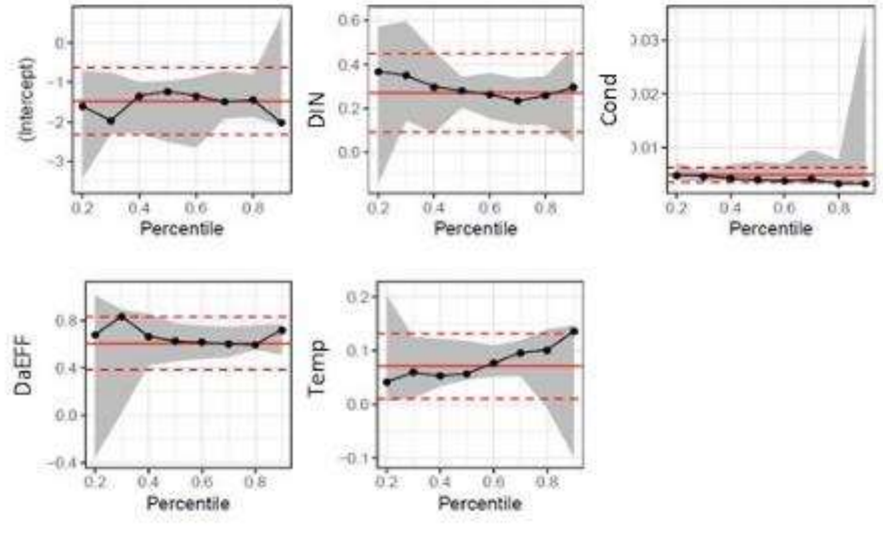

2015-16, n = 40
(Intercept)
DIN
Cond
DaEFF
Percentile


2009-12, n = 36
(Intercept)
DIN
Cond
DaEFF
Temperature
Percentile

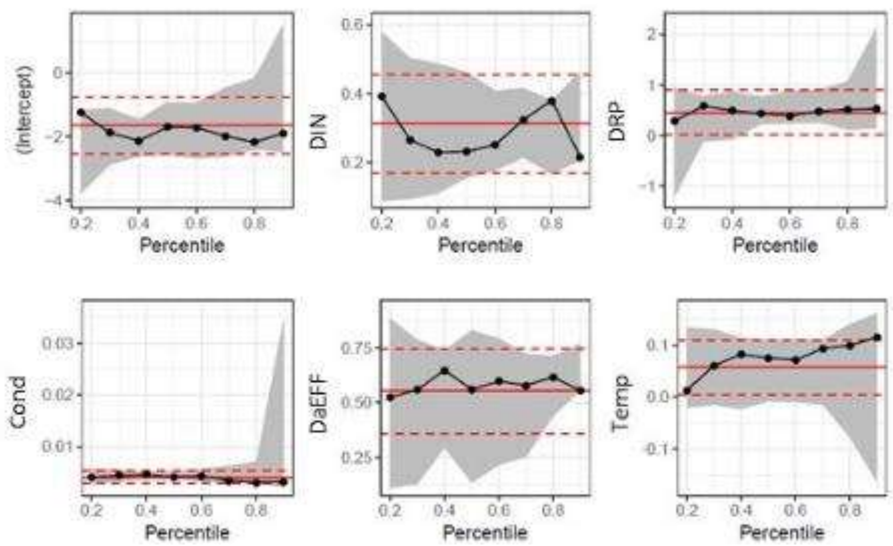
2010-13, n = 37

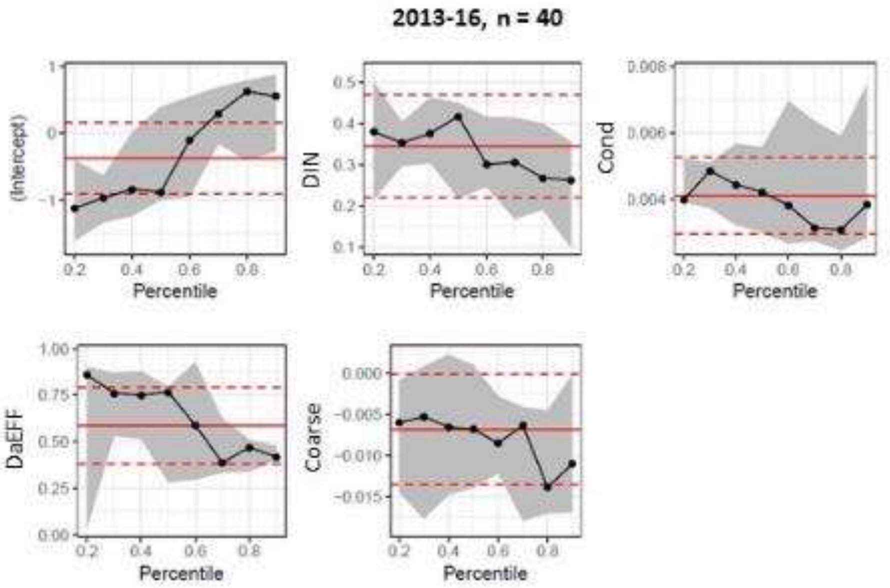


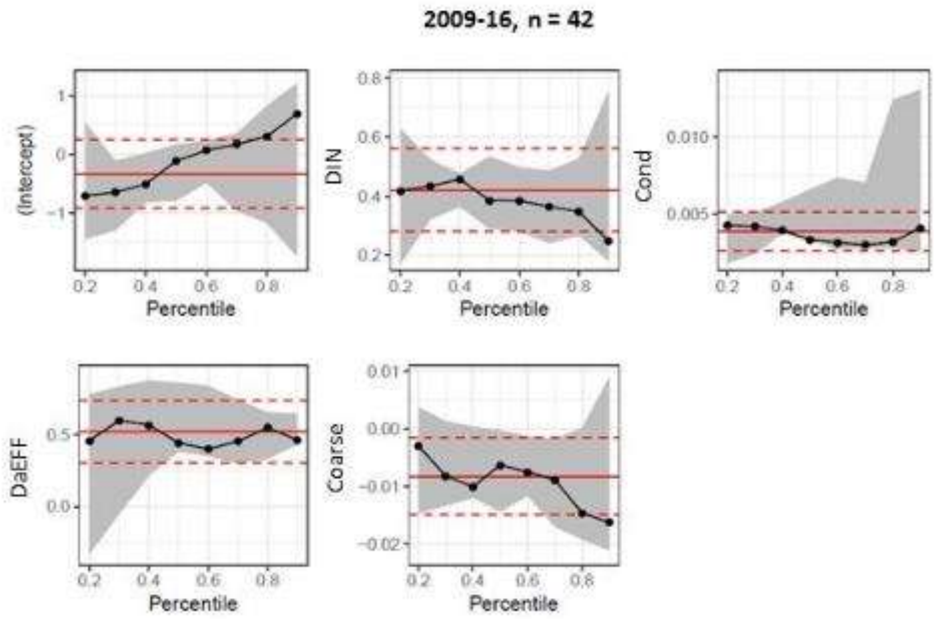
2011-14, n = 40



2012-15, n = 40







We did not investigate the quantile regressions at extreme values (>80th or <20th) any further as the temporal datasets are relatively small (maximum number of sites = 42). Although, and in fact because, the full distribution of points is used in determining the quantiles, the precision of the analysis depends on the both the sample size and the quantile being modelled (Cade and Richards 2006). Data are sparser at the extremes of the distribution. Therefore, modelling extreme quantiles like the 5th or 95th will have lower associated precision than modelling the median, particularly in small datasets (Cade and Richards 2006). Likewise, the relative insensitivity of quantile regression to outliers is greatly reduced in small datasets.

As an example, quantile regressions using the dataset from sites in year 2011-12 (n = 35) of maximum chlorophyll *a* against log DIN across quantiles from the 20th to the 90th show that the slope of this relationship flattens greatly at high percentiles (90th and 95th; Figure B-2). There are four and two datapoints, respectively, above the regression line at the 90th and 95th percentiles. One site (moawhango_waiouru) has a high chlorophyll *a* biomass at relatively low DIN (log DIN ~1.3). This one point has a disproportionate effect on the slope of the regression line at the 90th and 95th quantiles, and is the sole cause of the change in slope of the regression at the 90th and 95th percentiles (Figure B-2, left panel). However, if this site is removed from the analysis the slope of the 90th and 95th quantiles remain similar to other quantiles (Figure B-2, right panel). With larger datasets the influence of such single data points on coefficient estimates at extreme quantiles is likely to be less.

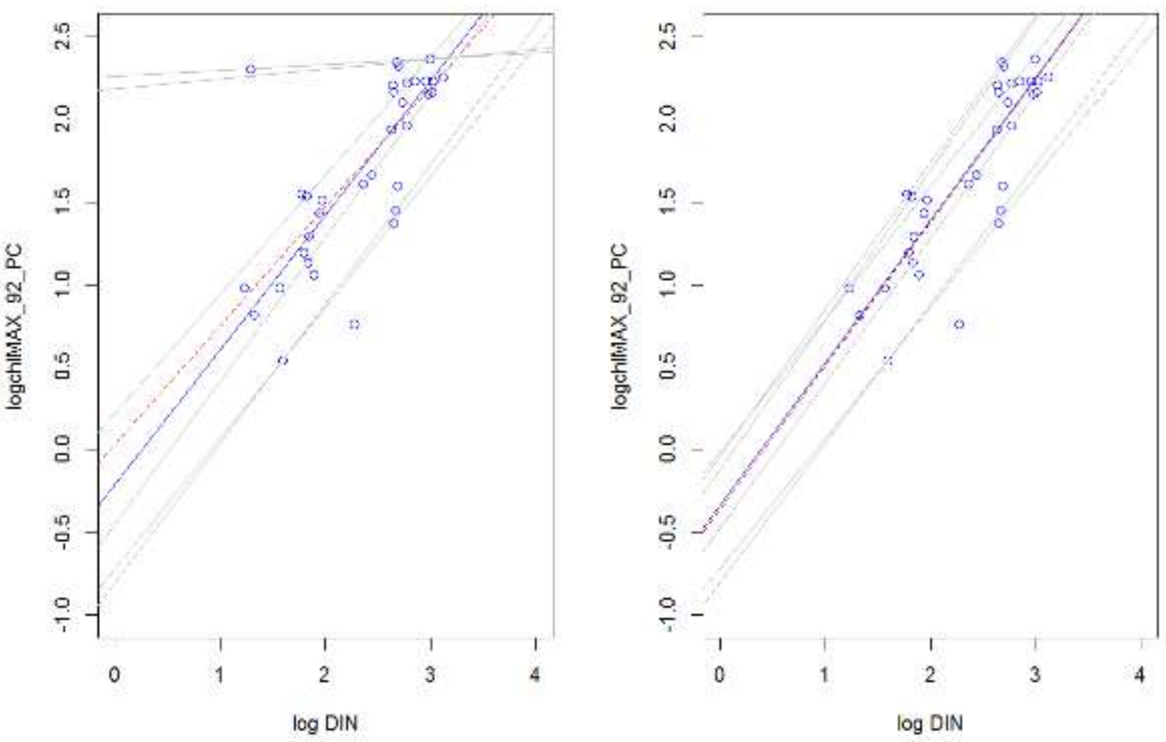


Figure B-2: Example of quantile regression using data from all sites in year 2011-2012 using a single predictor (Log DIN). The solid blue line is the median regression (50th quantile), the dashed red line is the OLS regression (mean), and the grey lines are regression lines from quantile regressions on the 5th, 10th, 25th, 75th, 90th and 95th percentiles. The plot on the left includes all 35 datapoints. In the plot on the right, the site moawhango_waiouru has been removed (n = 34).

References

Aertsen, W., Kint, V., van Orshoven, J., Ozkan, K., Muys, B. (2010) Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. *Ecological Modelling,* 221(8): 1119-1130.

Cade, B.S., Richards, J.D. (2006) A permutation test for quantile regression. *Journal of agricultural, biological and environmental sciences,* 11: 106-126.

Cunningham, S.C., Mac Nally, R., Read, J., Baker, P.J., White, M., Thomson, J.R., Griffioen, P. (2009) A robust technique for mapping vegetation condition across a major river system. *Ecosystems,* 12(2): 207-219.

Geyer, C.J. (2003) Model selection in R. http://www.stat.umn.edu/geyer/5931/mle/sel.pdf

Graham, M.H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology,* 84(11), 2809-2815.

Graham, M.H. (2003) Confronting multicollinearity in ecological multiple regression. *Ecology,* 84(11), 2809-2815.

Guisan, A., Edwards, T.C., Hastie, T. (2002) Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling,* 157(2-3): 89-100.

Huang, J.C., Gao, J.F., Mooij, W.M., Hormann, G., Fohrer, N. (2014) A Comparison of Three Approaches to Predict Phytoplankton Biomass in Gonghu Bay of Lake Taihu. *Journal of Environmental Informatics,* 24(1): 39-51.

Leclere, J., Oberdorff, T., Belliard, J., Leprieur, F. (2011) A comparison of modeling techniques to predict juvenile 0+fish species occurrences in a large river system. *Ecological Informatics,* 6(5): 276-285.

Mundry, R., Nunn, C.L. (2009) Stepwise model fitting and statistical inference; turning noise into signal pollution. *The American Naturalist* 173(1): 119-123.

Munn, M., Black, R., Gruber, S. (2002) Response of benthic algae to environmental gradients in an agriculturally dominated landscape. Journal of the North American *Benthological Society,* 21: 221-237.

Oppel, S., Meirinho, A., Ramirez, I., Gardner, B., O'Connell, A.F., Miller, P.I., Louzao, M. (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation,* 156: 94-104.

Segurado, P., Araujo, M.B. (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography*, 31(10): 1555-1568.

Sharma, S., Walker, S.C., Jackson, D.A. (2008) Empirical modelling of lake water-temperature relationships: a comparison of approaches. *Freshwater Biology*, 53(5): 897-911.

Appendix C Summary results of linear regressions between chlorophyll *a* and days since high flows

Table C-1: Summary results of linear regressions between chlorophyll *a* and days since high flows. All flows were daily mean flows. Blank cells indicate that no events exceeding the flow threshold occurred at that site. Median flow was the long-term median for each site (from 2000, or the start of the flow record, if later than 2000). Sites in order of Horizons site number (HRCn).

		R^2 of regression relationship between log_{10}chlorophyll *a* and log_{10}days since a high flow of 1.5 to 15 x median flow														
HRCn	**Site**	**1.5**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	makakahi_doc	0.01	0.01	0.01	0.03	0.08	0.11	0.10	0.10	0.12	0.12	0.14	0.08	0.04	0.12	0.12
2	mangatainoka_putara	0.01	0.01	0.04	0.06	0.09	0.08	0.08	0.13	0.13	0.14	0.13	0.10	0.11	0.12	0.13
3	mangatainoka_lars	0.12	0.13	0.22	0.21	0.31	0.35	0.37	0.43	0.44	0.44	0.44	0.47	0.49	0.53	0.43
7	mangatainoka_huk	0.12	0.14	0.15	0.14	0.20	0.22	0.26	0.28	0.29	0.31	0.31	0.25	0.25	0.27	0.26
8	kumeti_tr	0.46	0.44	0.41	0.45	0.34	0.28	0.20	0.15	0.17	0.14	0.06	0.04	0.01	0.01	0.00
9	manawatu_weber	0.32	0.35	0.34	0.30	0.39	0.35	0.38	0.35	0.34	0.31	0.30	0.30	0.30	0.28	0.28
10	makakahi_ham	0.01	0.01	0.02	0.03	0.03	0.04	0.04	0.07	0.06	0.08	0.08	0.11	0.13	0.10	0.10
11	oroua_apiti	0.35	0.39	0.40	0.33	0.29	0.17	0.15	0.09	0.09	0.08	0.03	0.01	0.01	0.00	0.00
12	tamaki_ste	0.43	0.41	0.46	0.44	0.38	0.31	0.20	0.18	0.11	0.07	0.06	0.06	0.06	0.06	0.06
13	oruakeretaki_sh2	0.42	0.49	0.44	0.44	0.44	0.39	0.36	0.25	0.14	0.15	0.10	0.09	0.08	0.10	0.07
14	makuri_tuscan	0.13	0.12	0.31	0.24	0.25	0.26	0.24	0.14	0.15	0.08	0.13	0.07	0.03	0.03	0.05
16	mangatainoka_scarb	0.16	0.19	0.32	0.31	0.24	0.29	0.22	0.22	0.21	0.20	0.21	0.20	0.16	0.19	0.19
17	tiraumea_nga	0.27	0.41	0.46	0.52	0.50	0.46	0.39	0.40	0.37	0.36	0.28	0.19	0.15	0.10	0.13
18	mangatainoka_pahiatua	0.14	0.15	0.25	0.28	0.23	0.28	0.18	0.13	0.16	0.21	0.21	0.18	0.12	0.13	0.13
19	mangatainoka_sh2	0.13	0.15	0.25	0.36	0.36	0.35	0.30	0.35	0.38	0.43	0.33	0.32	0.24	0.22	0.21
20	mangatainoka_ds_db	0.03	0.03	0.06	0.11	0.09	0.13	0.10	0.22	0.26	0.32	0.32	0.31	0.24	0.14	0.14
21	mangatainoka_us_pah	0.13	0.13	0.22	0.31	0.29	0.26	0.27	0.35	0.38	0.45	0.36	0.36	0.30	0.30	0.27
22	mangatainoka_ds_pah	0.02	0.03	0.07	0.12	0.11	0.17	0.17	0.27	0.31	0.35	0.37	0.35	0.34	0.29	0.27
23	manawatu_hop	0.47	0.46	0.42	0.46	0.38	0.39	0.39	0.37	0.36	0.36	0.34	0.27	0.22	0.21	0.12
24	mangatainoka_us_tir	0.01	0.01	0.07	0.15	0.11	0.19	0.18	0.22	0.26	0.31	0.35	0.35	0.29	0.27	0.28
26	mangapapa_troup	0.21	0.24	0.28	0.27	0.25	0.26	0.23	0.28	0.24	0.29	0.27	0.23	0.23	0.23	0.19
27	pohangina_mais	0.15	0.15	0.27	0.29	0.15	0.10	0.08	0.07	0.07	0.08	0.08	0.03	0.03	0.07	0.02
28	manawatu_ug	0.38	0.40	0.42	0.34	0.26	0.22	0.14	0.06	0.01	0.00	0.01	0.01	0.01	0.01	0.01

		R^2 of regression relationship between log_{10}chlorophyll *a* and log_{10}days since a high flow of 1.5 to 15 x median flow														
HRCn	**Site**	**1.5**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
29	oroua_almadale	0.16	0.18	0.21	0.15	0.11	0.07	0.03	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.01
33	manawatu_tc	0.41	0.45	0.41	0.36	0.27	0.18	0.10	0.03	0.00	0.01	0.02	0.01	0.01	0.01	0.00
34	manawatu_us_pncc	0.42	0.48	0.51	0.43	0.33	0.22	0.13	0.07	0.01	0.02	0.02	0.02	0.02	0.02	0.01
35	manawatu_ds_pncc	0.35	0.37	0.40	0.43	0.39	0.31	0.18	0.09	0.02	0.05	0.05	0.04	0.04	0.04	0.02
36	manawatu_opik	0.41	0.51	0.44	0.42	0.37	0.25	0.19	0.08	0.01	0.04	0.05	0.04	0.04	0.04	0.02
37	tokomaru_hb	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.02	0.02
38	rangitikei_puk	0.31	0.28	0.25	0.35	0.21	0.16	0.10	0.09	0.07	0.03	0.01	0.04	0.09	0.09	0.09
39	moawhango_waiouru	0.00	0.07	0.13	0.09	0.21	0.21	0.11								
40	rangitikei_man	0.20	0.25	0.28	0.30	0.17	0.11	0.08	0.06	0.06	0.09	0.09	0.06	0.03	0.03	0.03
43	rangitikei_one	0.33	0.31	0.31	0.34	0.24	0.13	0.02	0.04	0.04	0.03	0.11	0.08	0.14	0.14	0.14
44	rangitikei_mk	0.28	0.34	0.32	0.39	0.36	0.31	0.22	0.14	0.09	0.09	0.09	0.04	0.01	0.03	0.04
45	mangawhero_doc	0.01	0.02	0.05	0.01	0.04	0.05	0.03	0.06	0.00	0.01	0.03	0.00	0.00	0.03	0.03
46	makotuku_sh49	0.02	0.03	0.04	0.06	0.05	0.08	0.12	0.13	0.09	0.05	0.03	0.02	0.03	0.02	0.01
47	mangawhero_us_oha	0.00	0.02	0.05	0.01	0.02	0.02	0.01	0.06	0.01	0.05	0.08	0.00	0.03	0.00	0.00
48	mangawhero_ds_oha	0.00	0.03	0.05	0.02	0.04	0.04	0.03	0.13	0.00	0.01	0.05	0.00	0.00	0.02	0.02
49	makotuku_rae	0.01	0.01	0.00	0.01	0.02	0.03	0.02	0.02	0.05	0.05	0.05	0.08	0.09	0.02	0.03
50	mangawhero_pakihi	0.05	0.08	0.14	0.09	0.18	0.20	0.15	0.31	0.07	0.02	0.01	0.07	0.09	0.00	0.00
51	mangatepopo_gi	0.00	0.00	0.02	0.04	0.04	0.22	0.22	0.23	0.24	0.24	0.27	0.29	0.32	0.35	0.35
52	whanganui_ds_gen	0.00	0.02													
53	whakapapa_ds_gen	0.02	0.09	0.32	0.16	0.00	0.03	0.00								
57	makotuku_us_rae	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.04	0.03	0.00	0.00
58	makotuku_ds_rae	0.09	0.11	0.08	0.08	0.11	0.13	0.11	0.13	0.14	0.11	0.12	0.15	0.15	0.05	0.05
59	waikawa_nmr	0.18	0.23	0.15	0.16	0.14	0.11	0.11	0.13	0.12	0.09	0.04	0.06	0.05	0.05	0.09
60	ohau_gladstone	0.07	0.15	0.16	0.18	0.19	0.22	0.10	0.14	0.15	0.17	0.15	0.11	0.08	0.07	0.09
61	ohau_sh1	0.20	0.28	0.35	0.33	0.29	0.24	0.15	0.14	0.12	0.13	0.10	0.05	0.06	0.07	0.04
62	ohau_haines	0.10	0.17	0.23	0.25	0.32	0.29	0.16	0.12	0.12	0.15	0.15	0.10	0.09	0.09	0.06

Appendix D Cross-validation plots: DIN versus chlorophyll *a* across sites

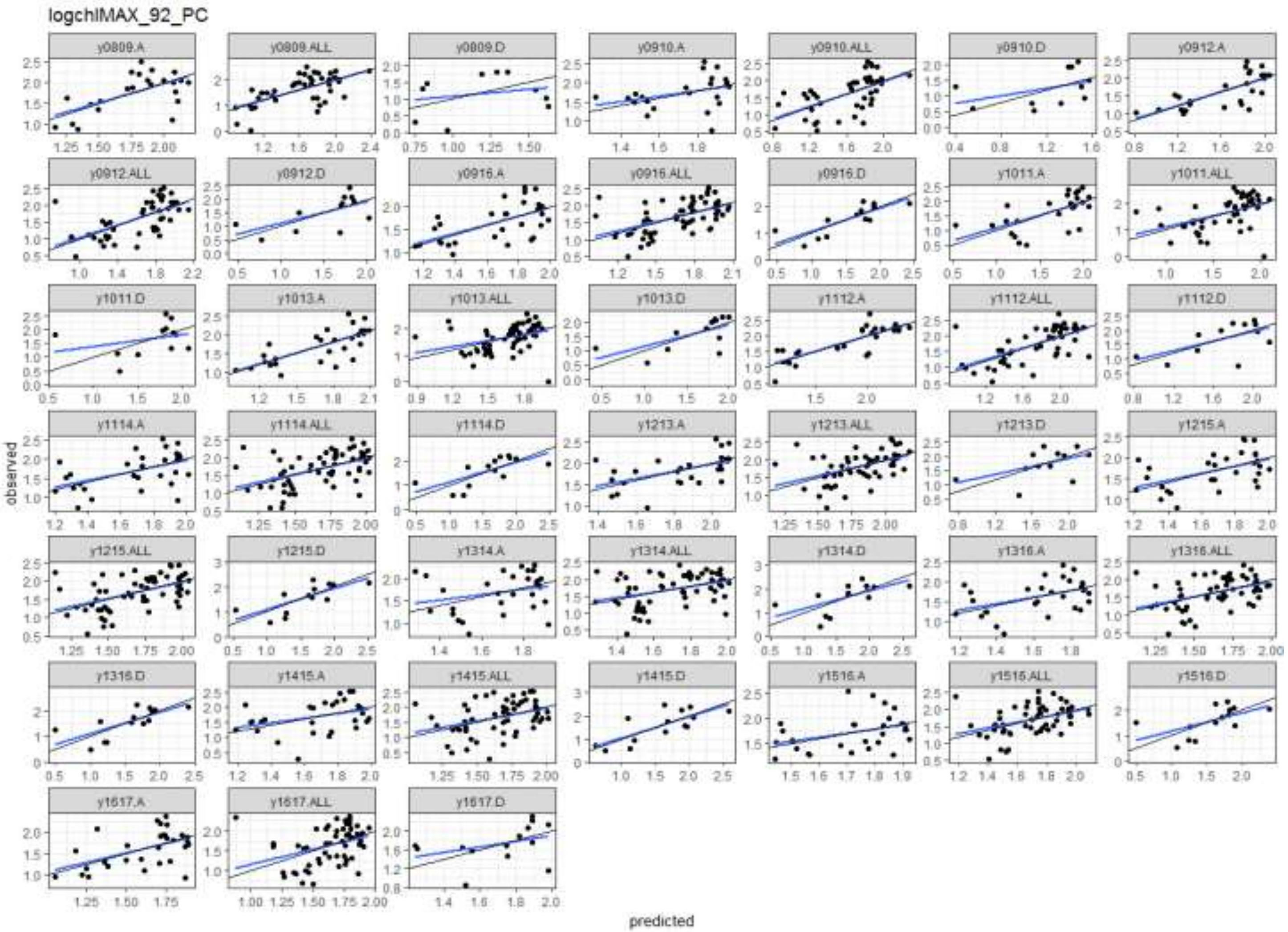


Figure D-1: Cross-validation plots: DIN versus chlorophyll *a* across sites. A = flow-sensitive sites, D = flow-insensitive sites. The dependent variable is maximum chlorophyll *a* for single years (y0910, etc.) and the 92nd percentile for multiple years (y0912, etc.). Some part years are included (y0809 and y1617).

Appendix E Between-site correlation matrices for water quality and catchment variables

Table E-1: Between-site correlation matrices for water quality and catchment variables. Matrices for the multi-year datasets are shown. For abbreviations and units refer to Table 2 2. In all regression models, a correlation matrix of the included variables was generated so that cross-correlations could be re-checked post-analysis.

Period	Variable	DIN	DRP	TN	TP	TDP	Temp	Cond	% sand	% coarse	%Farm	%Indig. For.	All Indig	%Grass	mChla	92Chl	WCC	matfil	TSS
y0912	DIN	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y0912	DRP	0.46	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y0912	TN	0.98	0.46	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y0912	TP	0.53	0.83	0.60	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y0912	TDP	0.43	0.92	0.45	0.78	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.
y0912	Temp	0.59	0.15	0.63	0.34	0.01	1.00	.	.	.	.	.	.	.	.	.	.	.	.
y0912	Cond	0.61	0.43	0.65	0.66	0.34	0.40	1.00	.	.	.	.	.	.	.	.	.	.	.
y0912	%sand	-0.02	0.48	-0.07	0.22	0.58	-0.35	-0.07	1.00	.	.	.	.	.	.	.	.	.	.
y0912	%coarse	-0.62	-0.55	-0.65	-0.69	-0.47	-0.72	-0.44	0.02	1.00	.	.	.	.	.	.	.	.	.
y0912	%Farm	0.82	0.53	0.87	0.71	0.48	0.58	0.76	-0.11	-0.61	1.00	.	.	.	.	.	.	.	.
y0912	%Indig. For.	-0.66	-0.43	-0.71	-0.66	-0.38	-0.50	-0.60	0.23	0.58	-0.72	1.00	.	.	.	.	.	.	.
y0912	%All. Indig	-0.81	-0.56	-0.86	-0.73	-0.50	-0.60	-0.76	0.07	0.62	-0.99	0.72	1.00	.	.	.	.	.	.
y0912	%Grass_lo	-0.14	-0.20	-0.12	-0.20	-0.14	0.06	-0.24	0.13	0.08	-0.30	0.13	0.21	1.00	.	.	.	.	.
y0912	mChla	0.60	0.04	0.58	0.10	0.04	0.19	0.54	-0.13	0.02	0.58	-0.31	-0.57	-0.13	1.00	.	.	.	.
y0912	92Chla	0.76	0.31	0.77	0.47	0.25	0.50	0.71	-0.13	-0.38	0.79	-0.57	-0.80	-0.08	0.83	1.00	.	.	.
y0912	WCC	0.44	0.13	0.44	0.27	0.12	0.21	0.52	-0.09	-0.06	0.52	-0.37	-0.53	-0.14	0.82	0.83	1.00	.	.
y0912	matfil	0.41	-0.07	0.38	-0.01	-0.02	0.09	0.30	-0.10	0.11	0.40	-0.19	-0.39	-0.13	0.88	0.70	0.90	1.00	.
y0912	TSS	0.24	0.29	0.33	0.57	0.17	0.46	0.36	-0.24	-0.60	0.37	-0.64	-0.37	-0.10	-0.23	0.09	-0.08	-0.30	1
y1013	DIN	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1013	DRP	0.24	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1013	TN	0.86	0.37	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1013	TP	0.31	0.89	0.54	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Period	Variable	DIN	DRP	TN	TP	TDP	Temp	Cond	% sand	% coarse	%Farm	%Indig. For.	All Indig	%Grass	mChla	92Chl	WCC	matfil	TSS
y1013	TDP	0.20	0.94	0.37	0.89	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.
y1013	Temp	0.44	0.07	0.64	0.25	0.02	1.00	.	.	.	.	.	.	.	.	.	.	.	.
y1013	Cond	0.31	0.44	0.62	0.68	0.45	0.46	1.00	.	.	.	.	.	.	.	.	.	.	.
y1013	%sand	0.05	0.52	-0.10	0.31	0.50	-0.35	-0.08	1.00	.	.	.	.	.	.	.	.	.	.
y1013	%coarse	-0.46	-0.41	-0.58	-0.55	-0.41	-0.67	-0.45	-0.06	1.00	.	.	.	.	.	.	.	.	.
y1013	%Farm	0.62	0.29	0.83	0.54	0.30	0.64	0.69	-0.24	-0.57	1.00	.	.	.	.	.	.	.	.
y1013	%Indig. For.	-0.41	-0.38	-0.59	-0.58	-0.39	-0.47	-0.60	0.10	0.55	-0.64	1.00	.	.	.	.	.	.	.
y1013	%All. Indig	-0.57	-0.33	-0.83	-0.57	-0.34	-0.68	-0.73	0.24	0.58	-0.99	0.63	1.00	.	.	.	.	.	.
y1013	%Grass_lo	-0.10	-0.24	-0.15	-0.25	-0.19	0.03	-0.25	0.04	0.08	-0.28	0.19	0.21	1.00	.	.	.	.	.
y1013	mChla	0.48	0.25	0.62	0.33	0.23	0.21	0.50	-0.01	0.00	0.41	-0.23	-0.43	-0.15	1.00	.	.	.	.
y1013	92Chla	0.70	0.32	0.81	0.48	0.29	0.46	0.64	-0.07	-0.32	0.67	-0.46	-0.68	-0.10	0.86	1.00	.	.	.
y1013	WCC	0.24	0.27	0.55	0.49	0.30	0.36	0.66	-0.16	-0.13	0.50	-0.40	-0.56	-0.18	0.78	0.76	1.00	.	.
y1013	matfil	0.14	0.21	0.48	0.37	0.26	0.31	0.53	-0.20	0.00	0.37	-0.27	-0.44	-0.17	0.78	0.64	0.93	1.00	.
y1013	TSS	0.22	0.16	0.30	0.36	0.18	0.39	0.40	-0.13	-0.63	0.45	-0.60	-0.43	-0.07	-0.25	0.07	-0.01	-0.21	1
y1114	DIN	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1114	DRP	0.17	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1114	TN	0.91	0.25	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1114	TP	0.27	0.86	0.45	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1114	TDP	0.22	0.97	0.34	0.90	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.
y1114	Temp	0.55	-0.03	0.69	0.19	0.03	1.00	.	.	.	.	.	.	.	.	.	.	.	.
y1114	Cond	0.34	0.40	0.57	0.69	0.46	0.41	1.00	.	.	.	.	.	.	.	.	.	.	.
y1114	%sand	0.00	0.51	-0.06	0.40	0.46	-0.28	0.06	1.00	.	.	.	.	.	.	.	.	.	.
y1114	%coarse	-0.41	-0.38	-0.53	-0.55	-0.41	-0.61	-0.48	-0.30	1.00	.	.	.	.	.	.	.	.	.
y1114	%Farm	0.66	0.20	0.83	0.47	0.28	0.69	0.64	-0.16	-0.51	1.00	.	.	.	.	.	.	.	.
y1114	%Indig. For.	-0.45	-0.37	-0.56	-0.56	-0.39	-0.52	-0.63	0.02	0.52	-0.64	1.00	.	.	.	.	.	.	.
y1114	%All. Indig	-0.61	-0.23	-0.83	-0.51	-0.33	-0.72	-0.69	0.15	0.53	-0.99	0.64	1.00	.	.	.	.	.	.

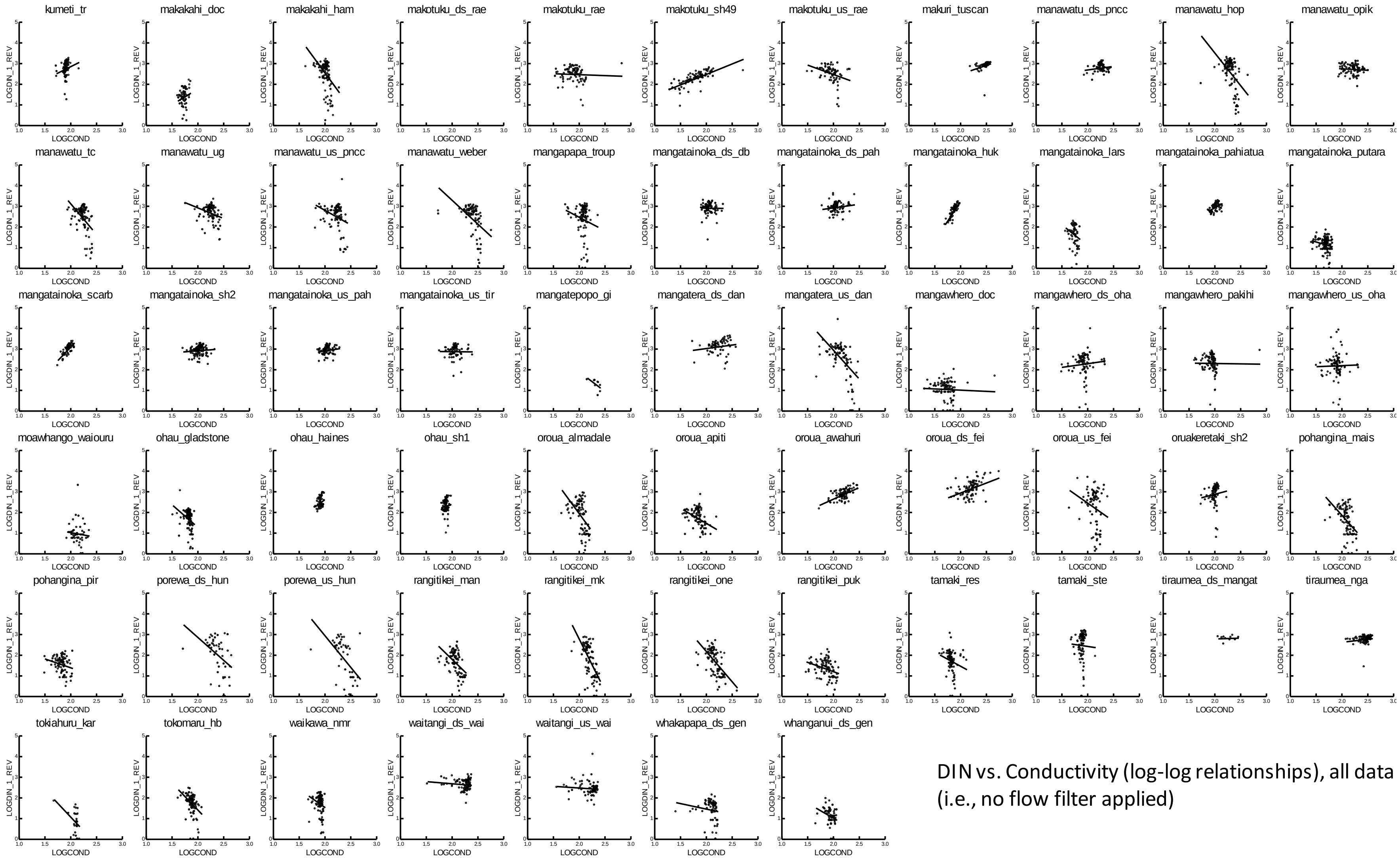
Period	Variable	DIN	DRP	TN	TP	TDP	Temp	Cond	% sand	% coarse	%Farm	%Indig. For.	All Indig	%Grass	mChla	92Chl	WCC	matfil	TSS
y1114	%Grass_lo	-0.17	-0.21	-0.19	-0.23	-0.20	0.02	-0.23	0.03	0.05	-0.28	0.19	0.22	1.00	.	.	.	.	.
y1114	mChla	0.52	0.19	0.62	0.34	0.27	0.23	0.54	-0.10	-0.01	0.43	-0.29	-0.45	-0.12	1.00	.	.	.	.
y1114	92Chla	0.67	0.28	0.78	0.50	0.35	0.44	0.67	-0.05	-0.28	0.64	-0.48	-0.65	-0.08	0.87	1.00	.	.	.
y1114	WCC	0.33	0.12	0.54	0.40	0.22	0.35	0.67	-0.18	-0.07	0.50	-0.41	-0.55	-0.17	0.81	0.77	1.00	.	.
y1114	matfil	0.23	0.05	0.46	0.24	0.16	0.30	0.47	-0.27	0.05	0.37	-0.28	-0.43	-0.15	0.75	0.62	0.92	1.00	.
y1114	TSS	0.08	0.20	0.13	0.47	0.19	0.11	0.40	0.17	-0.49	0.30	-0.30	-0.29	-0.08	-0.09	0.17	0.02	-0.22	1
y1215	DIN	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1215	DRP	0.03	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1215	TN	0.92	0.11	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1215	TP	0.24	0.83	0.42	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1215	TDP	0.13	0.96	0.26	0.92	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.
y1215	Temp	0.55	-0.20	0.69	0.16	-0.06	1.00	.	.	.	.	.	.	.	.	.	.	.	.
y1215	Cond	0.31	0.35	0.50	0.67	0.48	0.39	1.00	.	.	.	.	.	.	.	.	.	.	.
y1215	%sand	-0.05	0.39	-0.08	0.32	0.36	-0.20	0.10	1.00	.	.	.	.	.	.	.	.	.	.
y1215	%coarse	-0.33	-0.30	-0.46	-0.53	-0.39	-0.54	-0.48	-0.43	1.00	.	.	.	.	.	.	.	.	.
y1215	%Farm	0.66	0.10	0.83	0.48	0.26	0.73	0.62	-0.13	-0.44	1.00	.	.	.	.	.	.	.	.
y1215	%Indig. For.	-0.42	-0.29	-0.53	-0.54	-0.38	-0.50	-0.65	0.01	0.48	-0.63	1.00	.	.	.	.	.	.	.
y1215	%All. Indig	-0.62	-0.14	-0.82	-0.52	-0.31	-0.74	-0.66	0.12	0.47	-0.99	0.64	1.00	.	.	.	.	.	.
y1215	%Grass_lo	-0.18	-0.17	-0.20	-0.22	-0.18	0.02	-0.21	0.00	0.03	-0.28	0.19	0.22	1.00	.	.	.	.	.
y1215	mChla	0.52	0.12	0.61	0.35	0.24	0.24	0.60	-0.12	-0.04	0.47	-0.36	-0.49	-0.15	1.00	.	.	.	.
y1215	92Chla	0.63	0.20	0.73	0.49	0.33	0.42	0.68	-0.04	-0.27	0.62	-0.50	-0.64	-0.09	0.90	1.00	.	.	.
y1215	WCC	0.25	-0.01	0.44	0.31	0.14	0.37	0.65	-0.16	-0.06	0.45	-0.36	-0.50	-0.13	0.82	0.75	1.00	.	.
y1215	matfil	0.17	0.01	0.39	0.26	0.15	0.29	0.51	-0.20	-0.01	0.35	-0.27	-0.42	-0.13	0.76	0.62	0.93	1.00	.
y1215	TSS	0.15	0.14	0.22	0.48	0.21	0.24	0.41	0.08	-0.47	0.40	-0.35	-0.37	-0.11	-0.01	0.19	0.05	-0.11	1
y1316	DIN	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1316	DRP	0.09	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

Period	Variable	DIN	DRP	TN	TP	TDP	Temp	Cond	% sand	% coarse	%Farm	%Indig. For.	All Indig	%Grass	mChla	92Chl	WCC	matfil	TSS
y1316	TN	0.94	0.14	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1316	TP	0.27	0.79	0.45	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y1316	TDP	0.16	0.96	0.28	0.91	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.
y1316	Temp	0.54	-0.21	0.65	0.22	-0.07	1.00	.	.	.	.	.	.	.	.	.	.	.	.
y1316	Cond	0.33	0.38	0.49	0.69	0.52	0.33	1.00	.	.	.	.	.	.	.	.	.	.	.
y1316	%sand	0.04	0.32	0.00	0.19	0.29	-0.17	0.04	1.00	.	.	.	.	.	.	.	.	.	.
y1316	%coarse	-0.37	-0.31	-0.47	-0.56	-0.42	-0.49	-0.44	-0.37	1.00	.	.	.	.	.	.	.	.	.
y1316	%Farm	0.67	0.17	0.81	0.53	0.31	0.69	0.63	-0.13	-0.40	1.00	.	.	.	.	.	.	.	.
y1316	%Indig. For.	-0.47	-0.28	-0.56	-0.52	-0.35	-0.50	-0.66	0.09	0.48	-0.68	1.00	.	.	.	.	.	.	.
y1316	%All. Indig	-0.63	-0.20	-0.81	-0.58	-0.36	-0.71	-0.67	0.13	0.43	-0.99	0.68	1.00	.	.	.	.	.	.
y1316	%Grass	-0.21	-0.15	-0.21	-0.20	-0.16	0.02	-0.23	-0.01	0.05	-0.32	0.20	0.25	1.00	.	.	.	.	.
y1316	mChla	0.49	0.13	0.57	0.29	0.23	0.14	0.59	-0.01	-0.02	0.47	-0.36	-0.49	-0.18	1.00	.	.	.	.
y1316	92Chla	0.58	0.23	0.67	0.46	0.34	0.33	0.69	0.04	-0.23	0.61	-0.51	-0.63	-0.15	0.91	1.00	.	.	.
y1316	WCC	0.27	-0.04	0.43	0.28	0.12	0.30	0.64	-0.15	0.02	0.48	-0.36	-0.53	-0.17	0.83	0.77	1.00	.	.
y1316	matfil	0.25	-0.07	0.41	0.18	0.07	0.20	0.52	-0.14	0.10	0.40	-0.27	-0.45	-0.18	0.82	0.69	0.95	1.00	.
y1316	TSS	0.04	0.11	0.16	0.51	0.21	0.32	0.39	-0.01	-0.46	0.30	-0.28	-0.31	-0.09	-0.10	0.06	0.04	-0.10	1
y0916	DIN	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y0916	DRP	0.15	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y0916	TN	0.93	0.20	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y0916	TP	0.27	0.85	0.44	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.	.
y0916	TDP	0.14	0.97	0.26	0.90	1.00	.	.	.	.	.	.	.	.	.	.	.	.	.
y0916	Temp	0.59	-0.15	0.70	0.17	-0.09	1.00	.	.	.	.	.	.	.	.	.	.	.	.
y0916	Cond	0.36	0.40	0.55	0.71	0.48	0.38	1.00	.	.	.	.	.	.	.	.	.	.	.
y0916	%sand	0.02	0.59	-0.05	0.42	0.54	-0.33	0.07	1.00	.	.	.	.	.	.	.	.	.	.
y0916	%coarse	-0.45	-0.39	-0.55	-0.60	-0.43	-0.58	-0.49	-0.27	1.00	.	.	.	.	.	.	.	.	.
y0916	%Farm	0.69	0.18	0.84	0.49	0.26	0.71	0.65	-0.16	-0.51	1.00	.	.	.	.	.	.	.	.

Period	Variable	DIN	DRP	TN	TP	TDP	Temp	Cond	% sand	% coarse	%Farm	%Indig. For.	All Indig	%Grass	mChla	92Chl	WCC	matfil	TSS
y0916	%Indig. For.	-0.45	-0.35	-0.57	-0.58	-0.39	-0.46	-0.65	0.03	0.53	-0.65	1.00	.	.	.	.	.	.	.
y0916	%All. Indig	-0.65	-0.22	-0.83	-0.54	-0.30	-0.72	-0.69	0.16	0.53	-0.99	0.65	1.00	.	.	.	.	.	.
y0916	%Grass	-0.17	-0.19	-0.18	-0.22	-0.17	0.01	-0.24	0.04	0.06	-0.30	0.21	0.23	1.00	.	.	.	.	.
y0916	mChla	0.50	0.15	0.59	0.29	0.21	0.18	0.55	-0.11	-0.02	0.46	-0.34	-0.48	-0.19	1.00	.	.	.	.
y0916	92Chla	0.67	0.25	0.76	0.49	0.29	0.42	0.71	0.00	-0.34	0.68	-0.52	-0.69	-0.17	0.86	1.00	.	.	.
y0916	WCC	0.33	0.12	0.52	0.42	0.22	0.34	0.68	-0.19	-0.09	0.54	-0.43	-0.59	-0.21	0.83	0.80	1.00	.	.
y0916	matfil	0.30	0.05	0.48	0.28	0.16	0.24	0.53	-0.23	0.05	0.44	-0.32	-0.48	-0.21	0.86	0.72	0.94	1.00	.
y0916	TSS	0.13	0.22	0.25	0.52	0.24	0.34	0.48	0.02	-0.63	0.40	-0.49	-0.40	-0.08	-0.12	0.17	0.07	-0.10	1.00

Appendix F Plots of DIN versus conductivity within sites (log – log relationships)

Sites are arranged in alphabetical order. The plots show that DIN and conductivity are positively correlated at only a few sites (meaningthat conductivity could potentially be used as a surrogate for DIN at those sites. Negative correlations at many sites reflect negative conductivity vs. flow and positive DIN vs. flow.



DIN vs. Conductivity (log-log relationships), all data (i.e., no flow filter applied)

Appendix G Within-site correlations between chlorophyll *a* and nutrient variables

Table G-1: Within-site correlations between chlorophyll a and nutrient variables. Pearson correlation coefficients between chlorophyll a and DIN or DRP variables, within each site. Data were reduced to monthly frequency at 12 sites at which part of the data collection was fortnightly. All data were log10transformed. Grey-shaded cells are negative correlations. Note weak correlations in most cases (R < 0.3). Sites in order of Horizons site number (HRCn).

				DIN variable					DRP variable						
HRCn	site	n	dataset	DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m	DIN : DRP	Limit.
1	makakahi_doc	32	all data	0.21	0.10	0.09	0.22	0.16	0.05	-0.04	-0.08	-0.14	-0.21	0.15	co
		24	accrual	0.19	0.07	-0.06	0.10	0.11	-0.10	0.27	0.08	0.07	0.02	0.25	
		18	< median	0.34	0.09	-0.01	0.26	0.21	0.03	0.01	0.27	0.16	0.01	0.37	
		12	<0.5med.	0.12	0.02	-0.06	0.10	0.10	0.01	0.22	0.02	0.27	0.19	0.13	
2	mangatainoka_putara	83	all data	0.08	-0.06	-0.07	-0.05	-0.08	0.08	0.09	-0.02	0.09	0.09	0.01	co
		51	accrual	0.05	-0.07	-0.27	-0.21	-0.18	0.10	0.14	0.11	0.10	0.13	-0.03	
		34	< median	0.07	-0.01	-0.22	-0.15	-0.11	-0.05	0.17	0.08	0.05	-0.03	0.09	
		18	<0.5med.	0.15	0.11	-0.18	-0.07	-0.14	-0.10	0.07	0.24	-0.01	-0.01	0.19	
3	mangatainoka_lars	33	all data	-0.29	-0.23	-0.18	-0.32	-0.35	-0.04	0.07	0.11	0.01	-0.15	-0.24	co
		22	accrual	-0.22	-0.33	-0.40	-0.50	-0.44	-0.09	0.11	0.04	-0.07	-0.15	-0.16	
		13	< median	-0.24	-0.40	-0.44	-0.55	-0.52	-0.57	0.00	0.12	-0.41	-0.40	0.02	
		6	<0.5med.	-0.36	-0.26	-0.24	-0.53	-0.66	-0.71	-0.17	0.32	-0.24	-0.31	0.05	
7	mangatainoka_huk	33	all data	-0.11	-0.22	-0.62	-0.65	-0.75	-0.15	-0.14	0.07	-0.06	0.01	0.05	P
		18	accrual	-0.17	-0.28	-0.52	-0.66	-0.74	-0.10	-0.17	0.24	0.14	0.18	-0.04	
		13	< median	-0.43	-0.33	-0.39	-0.70	-0.70	-0.18	-0.31	-0.09	-0.18	-0.16	-0.14	
		6	<0.5med.	-0.75	-0.60	0.18	-0.82	-0.89	0.25	-0.64	0.01	0.11	0.36	-0.74	

HRCn	site	n	dataset	DIN variable					DRP variable					DIN : DRP	Limit.
				DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m		
8	kumeti_tr	85	all data	-0.25	-0.04	-0.04	-0.15	-0.17	-0.46	-0.25	-0.19	-0.44	-0.41	0.05	P
		37	accrual	-0.12	0.03	-0.10	-0.14	-0.26	-0.35	-0.23	-0.40	-0.52	-0.54	0.12	
		36	< median	-0.15	0.00	-0.12	-0.16	-0.27	-0.38	-0.27	-0.42	-0.55	-0.56	0.11	
		21	<0.5med.	-0.43	-0.16	0.00	-0.22	-0.22	-0.20	-0.40	-0.26	-0.43	-0.54	-0.28	
9	manawatu_weber	75	all data	-0.42	-0.48	-0.52	-0.60	-0.59	-0.19	-0.17	-0.15	-0.24	-0.31	-0.38	co -none
		44	accrual	-0.19	-0.34	-0.44	-0.48	-0.57	-0.03	-0.01	-0.11	-0.03	-0.11	-0.19	
		33	< median	-0.25	-0.43	-0.54	-0.62	-0.73	-0.01	0.13	-0.25	-0.11	-0.26	-0.24	
		20	<0.5med.	-0.20	-0.53	-0.59	-0.68	-0.75	0.17	0.16	-0.26	-0.05	-0.06	-0.23	
10	makakahi_ham	80	all data	-0.15	-0.12	-0.25	-0.27	-0.36	-0.16	0.16	0.06	0.00	0.07	-0.09	co-P
		69	accrual	-0.05	-0.06	-0.20	-0.20	-0.30	-0.07	0.23	0.05	0.11	0.17	-0.03	
		41	< median	-0.12	-0.07	-0.12	-0.16	-0.20	-0.17	0.20	0.02	0.06	0.16	-0.07	
		25	<0.5med.	-0.08	0.04	-0.03	-0.07	-0.07	0.01	0.14	0.15	0.21	0.32	-0.09	
11	oroua_apiti	81	all data	-0.42	-0.32	-0.19	-0.41	-0.45	-0.13	0.14	0.05	0.01	0.00	-0.30	co
		35	accrual	-0.11	-0.15	0.01	-0.14	-0.37	0.03	0.30	0.21	0.25	0.28	-0.10	
		26	< median	0.14	-0.11	0.03	-0.05	-0.40	-0.02	0.13	0.12	0.10	0.19	0.11	
		19	<0.5med.	-0.05	-0.16	0.09	-0.08	-0.44	0.02	0.31	0.50	0.34	0.40	-0.04	
12	tamaki_ste	76	all data	-0.63	-0.48	-0.31	-0.51	-0.35	-0.44	-0.12	0.06	-0.27	-0.17	-0.44	co-P
		39	accrual	-0.41	-0.39	-0.12	-0.36	-0.41	-0.37	-0.10	0.02	-0.26	-0.26	-0.21	
		29	< median	-0.34	-0.28	-0.05	-0.26	-0.32	-0.33	-0.04	0.15	-0.13	-0.15	-0.17	

HRCn	site	n	dataset	DIN variable: DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP variable: DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m	DIN : DRP	Limit.
		20	<0.5med.	-0.33	-0.29	-0.22	-0.39	-0.46	-0.25	0.21	0.25	0.04	0.11	-0.16	
13	oruakeretaki_sh2	82	all data	-0.38	-0.31	-0.34	-0.47	-0.48	-0.23	-0.09	0.03	-0.13	-0.07	-0.26	none-P
		34	accrual	-0.38	-0.49	-0.05	-0.34	-0.35	-0.43	-0.26	0.03	-0.30	-0.18	-0.19	
		27	< median	-0.43	-0.44	0.11	-0.26	-0.42	-0.34	-0.17	0.04	-0.25	-0.05	-0.29	
		20	<0.5med.	-0.38	-0.34	-0.13	-0.33	-0.50	-0.30	-0.01	0.08	-0.25	0.01	-0.24	
14	makuri_tuscan	81	all data	0.07	0.01	0.14	0.22	0.14	-0.15	-0.10	-0.34	-0.33	-0.43	0.16	P
		44	accrual	0.11	-0.13	-0.13	0.32	0.20	0.03	0.02	-0.44	-0.27	-0.19	0.06	
		35	< median	0.29	-0.10	-0.12	0.40	0.24	0.07	0.00	-0.54	-0.29	-0.25	-0.02	
		1	<0.5med.												
16	mangatainoka_scarb	32	all data	-0.23	-0.14	-0.34	-0.45	-0.48	-0.46	-0.04	-0.21	-0.25	-0.44	0.18	P
		24	accrual	-0.33	-0.20	-0.43	-0.57	-0.49	-0.39	-0.28	-0.09	-0.27	-0.43	0.04	
		15	< median	-0.51	-0.33	-0.44	-0.75	-0.71	-0.44	-0.26	-0.08	-0.27	-0.46	-0.10	
		10	<0.5med.	-0.42	-0.11	-0.17	-0.57	-0.68	-0.36	-0.51	-0.45	-0.38	-0.42	-0.06	
17	tiraumea_nga	62	all data	0.25	0.20	0.06	0.16	0.00	-0.30	-0.26	-0.35	-0.48	-0.53	0.34	P
		38	accrual	0.17	0.31	0.16	0.35	0.17	0.04	-0.09	-0.43	-0.39	-0.52	0.03	
		33	< median	0.17	0.35	0.16	0.38	0.19	0.08	-0.07	-0.42	-0.38	-0.54	-0.01	
		17	<0.5med.	-0.20	0.00	0.10	0.16	0.09	0.31	0.08	-0.34	-0.06	-0.33	-0.34	
19	mangatainoka_sh2	83	all data	-0.29	-0.21	-0.32	-0.36	-0.38	-0.41	-0.02	-0.11	-0.31	-0.29	0.12	P
		66	accrual	-0.24	-0.23	-0.39	-0.36	-0.41	-0.35	0.10	-0.15	-0.23	-0.22	0.11	

HRCn	site	n	dataset	DIN variable DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP variable DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m	DIN : DRP	Limit.
		38	< median	-0.18	-0.24	-0.33	-0.30	-0.44	-0.38	0.03	-0.18	-0.44	-0.32	0.15	
		21	<0.5med.	-0.11	-0.17	-0.08	-0.12	-0.23	-0.30	0.23	-0.12	-0.26	-0.13	0.15	
20	mangatainoka_ds_db	74	all data	-0.29	-0.23	-0.18	-0.38	-0.40	-0.16	-0.08	-0.16	-0.15	-0.16	-0.05	P
		61	accrual	-0.17	-0.23	-0.21	-0.37	-0.43	-0.12	-0.14	-0.08	-0.09	-0.06	-0.01	
		38	< median	-0.10	-0.24	-0.20	-0.30	-0.36	-0.21	-0.05	-0.28	-0.28	-0.21	0.07	
		21	<0.5med.	-0.26	-0.22	-0.20	-0.28	-0.23	-0.04	0.15	-0.26	0.01	0.12	-0.17	
21	mangatainoka_us_pah	77	all data	-0.31	-0.32	-0.31	-0.42	-0.34	0.14	-0.15	0.13	0.06	-0.05	-0.26	P
		67	accrual	-0.19	-0.31	-0.35	-0.40	-0.34	0.14	-0.02	0.11	0.17	0.05	-0.21	
		38	< median	-0.04	-0.19	-0.31	-0.27	-0.28	0.13	0.03	0.29	0.26	0.20	-0.14	
		22	<0.5med.	-0.08	-0.01	-0.23	-0.20	-0.13	0.02	0.17	0.21	0.22	0.12	-0.06	
22	mangatainoka_ds_pah	65	all data	-0.31	-0.40	-0.42	-0.52	-0.49	0.23	0.13	-0.02	0.12	0.08	-0.40	P
		58	accrual	-0.25	-0.36	-0.38	-0.49	-0.45	0.25	0.24	0.12	0.26	0.25	-0.39	
		33	< median	-0.30	-0.31	-0.54	-0.50	-0.48	0.15	0.31	0.00	0.20	0.23	-0.37	
		19	<0.5med.	-0.55	-0.48	-0.74	-0.71	-0.68	0.15	0.30	-0.49	-0.05	-0.01	-0.45	
23	manawatu_hop	70	all data	-0.53	-0.38	-0.35	-0.49	-0.46	-0.23	-0.04	0.00	-0.18	-0.25	-0.44	none-N
		26	accrual	-0.24	-0.14	-0.11	-0.24	-0.24	-0.14	0.26	0.16	0.11	0.16	-0.15	
		25	< median	-0.24	-0.14	-0.11	-0.24	-0.24	-0.14	0.26	0.17	0.11	0.16	-0.15	
		19	<0.5med.	-0.13	-0.11	-0.20	-0.24	-0.21	-0.17	0.23	0.16	0.03	0.18	-0.04	
24	mangatainoka_us_tir	60	all data	-0.29	-0.37	-0.33	-0.43	-0.44	-0.17	-0.21	-0.24	-0.26	-0.30	-0.07	P

HRCn	site	n	dataset	DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m	DIN : DRP	Limit.
				DIN variable					DRP variable						
		52	accrual	-0.17	-0.33	-0.35	-0.39	-0.44	-0.09	-0.20	-0.33	-0.31	-0.27	-0.04	
		32	< median	-0.06	-0.23	-0.36	-0.31	-0.38	-0.06	-0.35	-0.19	-0.39	-0.33	0.01	
		16	<0.5med.	-0.13	-0.03	-0.25	-0.18	-0.17	-0.38	-0.05	-0.36	-0.39	-0.46	0.12	
26	mangapapa_troup	87	all data	-0.45	-0.46	-0.45	-0.56	-0.50	-0.10	-0.06	-0.04	-0.17	-0.27	-0.40	co-P
		60	accrual	-0.36	-0.39	-0.39	-0.47	-0.45	-0.01	0.01	-0.11	-0.10	-0.20	-0.33	
		38	< median	-0.39	-0.39	-0.43	-0.46	-0.46	-0.03	0.09	-0.08	-0.08	-0.17	-0.35	
		25	<0.5med.	-0.31	-0.21	-0.29	-0.26	-0.29	-0.03	0.05	-0.04	-0.06	-0.16	-0.28	
27	pohangina_mais	77	all data	-0.23	-0.11	-0.18	-0.20	-0.21	0.09	0.07	0.14	0.18	0.17	-0.21	co
		47	accrual	0.21	0.16	-0.10	0.07	-0.08	-0.06	-0.09	0.10	0.08	0.17	0.19	
		36	< median	0.12	0.08	-0.08	0.03	-0.07	-0.02	-0.17	0.07	0.06	0.16	0.11	
		15	<0.5med.	0.33	0.05	-0.11	0.04	-0.18	0.15	0.27	0.29	0.36	0.39	0.24	
28	manawatu_ug	77	all data	-0.55	-0.49	-0.31	-0.43	-0.24	-0.39	-0.30	-0.14	-0.35	-0.25	-0.23	P-co
		36	accrual	-0.32	-0.40	-0.27	-0.32	-0.22	-0.16	-0.26	-0.10	-0.22	-0.01	-0.19	
		31	< median	-0.27	-0.37	-0.26	-0.30	-0.20	-0.13	-0.23	-0.11	-0.22	-0.17	-0.16	
		20	<0.5med.	-0.17	-0.28	-0.07	-0.17	-0.15	-0.08	-0.27	-0.20	-0.26	-0.35	-0.11	
29	oroua_almadale	77	all data	-0.41	-0.10	0.04	-0.10	-0.04	-0.33	0.01	0.06	-0.13	-0.15	-0.31	co
		38	accrual	0.09	0.12	0.24	0.23	0.11	-0.39	-0.07	0.32	-0.05	0.10	0.16	
		28	< median	-0.05	-0.06	0.01	0.02	-0.16	-0.50	-0.23	0.37	-0.06	0.09	0.07	
		21	<0.5med.	-0.22	-0.06	0.03	0.02	-0.16	-0.57	-0.19	0.42	0.01	0.19	0.01	

HRCn	site	n	dataset	DIN variable: DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP variable: DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m	DIN : DRP	Limit.
33	manawatu_tc	75	all data	-0.58	-0.48	-0.27	-0.47	-0.28	-0.34	-0.16	-0.08	-0.31	-0.16	-0.44	co
		28	accrual	-0.15	-0.20	-0.17	-0.24	-0.24	-0.09	-0.06	0.16	-0.11	0.02	-0.11	
		27	< median	-0.10	-0.15	-0.12	-0.19	-0.18	-0.14	-0.02	0.26	-0.08	0.04	-0.04	
		20	<0.5med.	0.18	-0.11	-0.08	-0.07	-0.12	-0.09	-0.10	0.27	-0.08	0.10	0.21	
34	manawatu_us_pncc	82	all data	-0.59	-0.48	-0.38	-0.55	-0.30	-0.36	-0.24	-0.19	-0.39	-0.31	-0.46	co-P
		37	accrual	-0.34	-0.35	-0.23	-0.36	-0.19	-0.19	-0.15	-0.03	-0.20	0.05	-0.27	
		33	< median	-0.32	-0.24	-0.23	-0.32	-0.13	-0.17	-0.02	-0.02	-0.12	0.03	-0.26	
		21	<0.5med.	-0.04	-0.01	0.09	-0.05	-0.07	-0.15	-0.08	0.49	0.18	0.20	0.03	
35	manawatu_ds_pncc	49	all data	-0.50	-0.34	-0.33	-0.50	-0.34	-0.41	-0.26	-0.06	-0.45	-0.32	0.19	none-P
		35	accrual	-0.41	-0.40	-0.13	-0.44	-0.22	-0.30	-0.29	-0.16	-0.50	-0.27	0.15	
		28	< median	-0.27	-0.33	-0.23	-0.37	-0.20	-0.26	-0.24	-0.08	-0.36	-0.16	0.20	
		15	<0.5med.	0.23	-0.09	-0.01	0.05	0.11	-0.29	-0.29	-0.05	-0.50	-0.07	0.37	
36	manawatu_opik	67	all data	-0.46	-0.25	-0.30	-0.38	-0.24	-0.55	-0.37	-0.38	-0.55	-0.40	0.26	none-P
		28	accrual	0.11	-0.10	-0.16	-0.13	-0.08	-0.57	-0.27	-0.30	-0.59	-0.42	0.60	
		27	< median	0.05	0.00	-0.13	-0.11	-0.01	-0.59	-0.08	-0.20	-0.47	-0.30	0.59	
		23	<0.5med.	0.07	0.16	0.10	0.10	0.14	-0.56	-0.02	-0.05	-0.37	-0.19	0.60	
38	rangitikei_puk	67	all data	-0.43	-0.23	-0.22	-0.39	-0.39	0.01	0.10	0.08	0.12	0.19	-0.38	co
		48	accrual	-0.27	-0.25	-0.20	-0.38	-0.42	0.03	-0.01	0.07	0.08	0.17	-0.23	
		34	< median	-0.24	-0.25	-0.30	-0.39	-0.45	0.10	0.06	0.02	0.17	0.29	-0.23	
		20	<0.5med.	-0.09	-0.16	-0.24	-0.28	-0.34	0.19	0.37	0.01	0.18	0.23	-0.14	

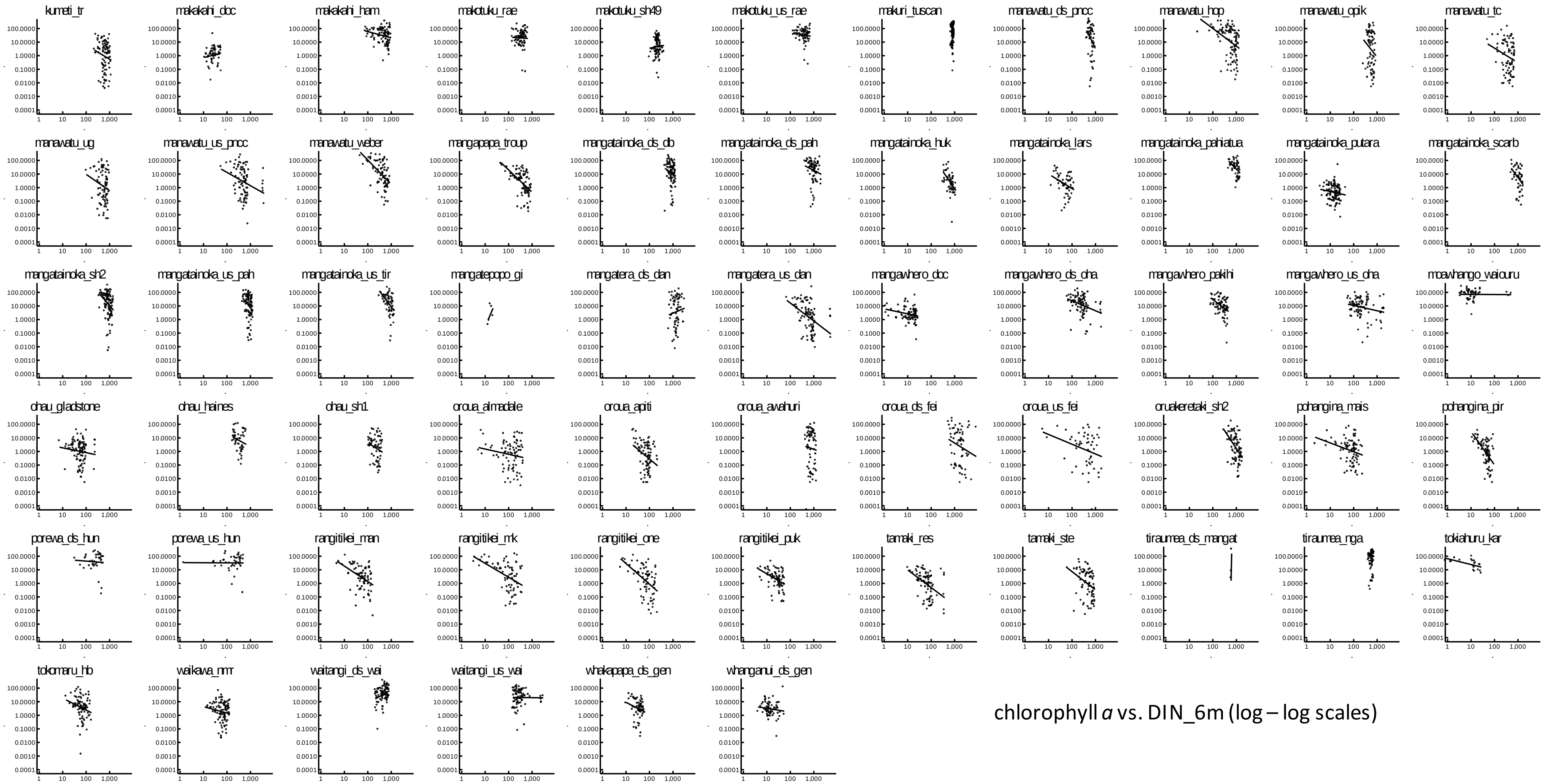
HRCn	site	n	dataset	DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m	DIN : DRP	Limit.
				DIN variable					DRP variable						
39	moawhango_waiouru	28	all data	-0.10	-0.17	-0.16	-0.40	-0.32	0.11	0.27	0.34	0.29	0.32	-0.14	co
		28	accrual	-0.10	-0.17	-0.16	-0.40	-0.32	0.11	0.27	0.34	0.29	0.32	-0.14	
		24	< median	-0.09	-0.06	-0.22	-0.37	-0.33	0.10	0.35	0.34	0.31	0.35	-0.12	
		0	<0.5med.												
40	rangitikei_man	72	all data	-0.30	-0.29	-0.34	-0.38	-0.42	-0.05	0.00	-0.12	-0.10	-0.09	-0.30	co
		55	accrual	-0.20	-0.24	-0.28	-0.32	-0.39	0.02	-0.04	-0.12	-0.08	-0.12	-0.21	
		36	< median	-0.17	-0.34	-0.48	-0.48	-0.61	0.30	0.17	-0.18	0.13	0.00	-0.26	
		20	<0.5med.	-0.06	-0.39	-0.53	-0.52	-0.64	0.37	0.07	-0.08	0.11	-0.01	-0.21	
43	rangitikei_one	75	all data	-0.52	-0.35	-0.51	-0.57	-0.56	-0.06	-0.01	-0.13	-0.12	-0.10	-0.51	co
		53	accrual	-0.39	-0.30	-0.47	-0.51	-0.55	0.11	0.10	-0.15	0.03	0.01	-0.41	
		38	< median	-0.27	-0.31	-0.73	-0.67	-0.65	0.23	0.20	-0.24	0.12	0.11	-0.35	
		19	<0.5med.	-0.05	-0.33	-0.75	-0.65	-0.71	0.23	0.24	-0.07	0.12	0.07	-0.12	
44	rangitikei_mk	71	all data	-0.42	-0.35	-0.46	-0.52	-0.52	0.08	0.19	-0.06	0.13	0.16	-0.45	co
		52	accrual	-0.33	-0.34	-0.46	-0.48	-0.51	0.20	0.16	-0.05	0.21	0.20	-0.38	
		35	< median	-0.27	-0.52	-0.75	-0.77	-0.76	0.39	0.23	-0.20	0.29	0.09	-0.44	
		16	<0.5med.	-0.43	-0.64	-0.85	-0.88	-0.90	0.37	0.22	0.15	0.47	0.31	-0.54	
46	makotuku_SH49	85	all data	0.06	0.04	0.07	0.12	0.14	0.31	0.25	0.40	0.49	0.52	-0.22	co-N
		56	accrual	-0.16	-0.11	0.10	-0.03	0.02	0.36	0.30	0.43	0.61	0.63	-0.43	
		37	< median	-0.36	-0.15	0.17	-0.06	0.00	0.41	0.31	0.45	0.65	0.65	-0.53	

HRCn	site	n	dataset	DIN variable DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP variable DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m	DIN : DRP	Limit.
		21	<0.5med.	-0.46	-0.07	0.05	-0.08	-0.05	0.36	0.52	0.58	0.76	0.74	-0.64	
50	mangawhero_pakihi	76	all data	-0.12	-0.14	-0.35	-0.32	-0.36	-0.01	-0.05	-0.06	-0.08	-0.14	-0.10	co-N
		71	accrual	-0.01	-0.09	-0.31	-0.27	-0.37	0.03	0.00	-0.08	-0.08	-0.15	-0.02	
		45	< median	0.13	0.06	-0.31	-0.18	-0.29	-0.05	0.18	-0.07	-0.01	-0.04	0.14	
		18	<0.5med.	-0.13	-0.28	-0.48	-0.48	-0.54	-0.07	0.25	-0.05	-0.11	-0.02	-0.06	
53	whakapapa_ds_gen	55	all data	-0.10	-0.15	-0.43	-0.36	-0.34	0.26	0.29	0.14	0.29	0.36	-0.16	N
		47	accrual	0.08	-0.10	-0.53	-0.36	-0.47	0.30	0.34	0.19	0.34	0.37	-0.02	
		28	< median	0.23	-0.18	-0.38	-0.24	-0.39	0.16	0.07	0.06	0.14	0.24	0.17	
		0	<0.5med.												
59	waikawa_nmr	85	all data	0.05	0.05	-0.13	-0.10	-0.21	-0.08	-0.03	-0.06	-0.09	-0.15	0.08	co
		15	accrual	0.27	0.37	-0.07	0.17	0.23	0.51	0.08	0.27	0.54	0.48	0.09	
		14	< median	0.25	0.36	-0.09	0.14	0.21	0.50	0.07	0.27	0.53	0.47	0.06	
		8	<0.5med.	-0.10	0.26	0.13	0.07	0.13	0.40	-0.16	0.38	0.51	0.31	-0.18	
60	ohau_gladstone	85	all data	0.00	0.04	-0.23	-0.14	-0.16	-0.14	-0.10	-0.10	-0.17	-0.23	0.04	co
		54	accrual	-0.02	0.07	-0.19	-0.15	-0.17	-0.09	-0.18	-0.06	-0.12	-0.29	0.00	
		44	< median	-0.06	0.03	-0.34	-0.24	-0.28	-0.20	-0.18	-0.03	-0.17	-0.27	0.00	
		17	<0.5med.	-0.36	-0.18	-0.30	-0.38	-0.41	0.04	-0.41	0.23	0.25	0.03	-0.44	
61	ohau_sh1	55	all data	-0.06	-0.09	-0.13	-0.19	-0.23	-0.27	-0.14	-0.10	-0.34	-0.18	0.18	co
		15	accrual	0.30	0.08	0.26	0.37	0.43	-0.20	0.03	0.21	0.18	0.26	0.32	

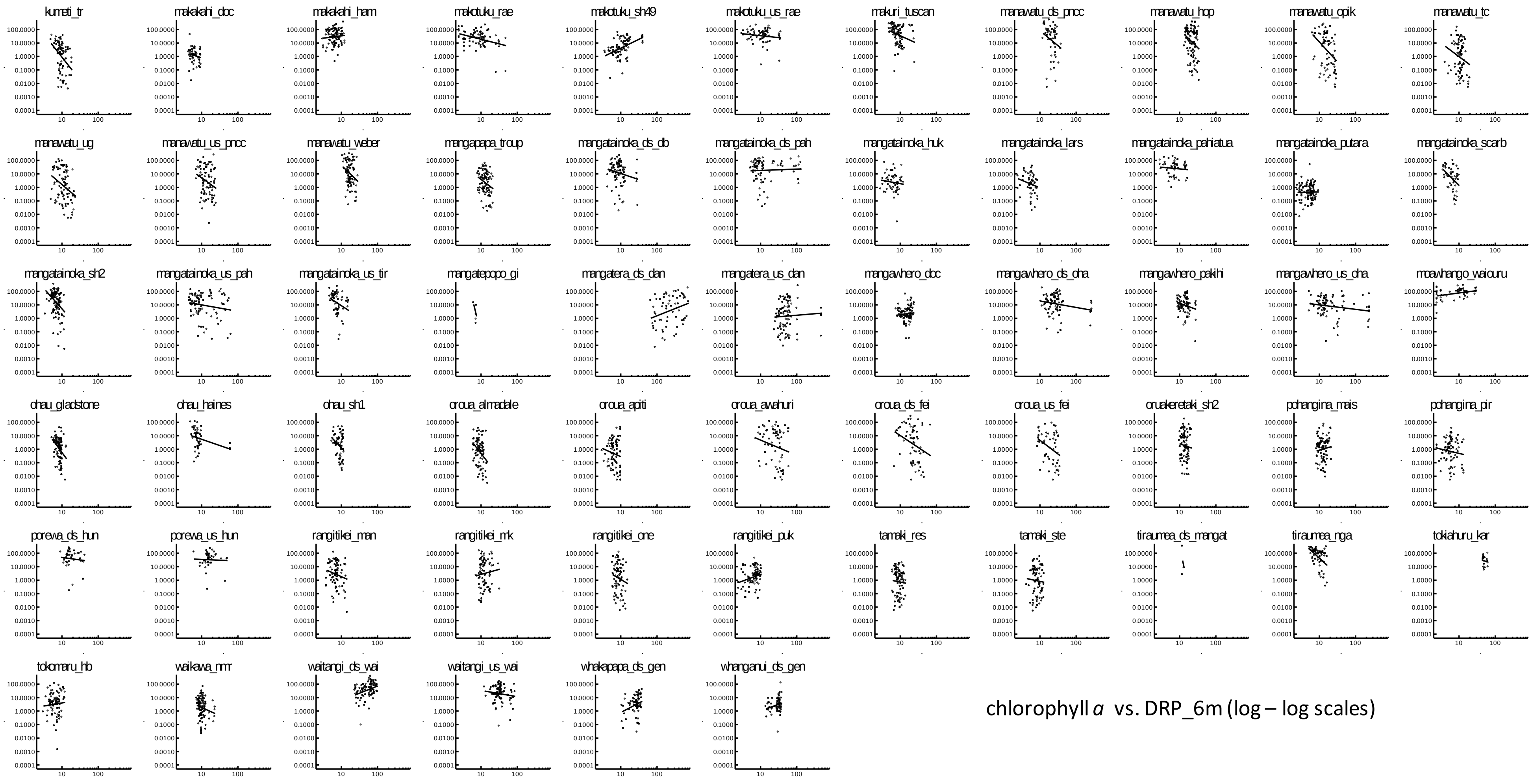
HRCn	site	n	dataset	DIN variable DIN	DIN_1m lag	DIN_2m lag	DIN_4m	DIN_6m	DRP variable DRP	DRP_1m lag	DRP_2m lag	DRP_4m	DRP_6m	DIN : DRP	Limit.
		14	< median	0.30	0.09	0.29	0.38	0.45	-0.18	0.07	0.20	0.21	0.35	0.31	
		7	<0.5med.	0.40	-0.15	0.58	0.61	0.73	-0.42	-0.01	-0.07	-0.17	0.06	0.66	
62	ohau_haines	36	all data	-0.26	-0.23	-0.11	-0.33	-0.34	-0.48	-0.17	-0.21	-0.39	-0.33	0.16	co
		17	accrual	-0.39	-0.24	0.07	-0.25	-0.18	-0.36	-0.09	-0.45	-0.43	-0.38	0.03	
		15	< median	-0.42	-0.30	0.04	-0.32	-0.21	-0.39	-0.10	-0.44	-0.47	-0.40	0.04	
		6	<0.5med.	-0.43	0.18	0.34	0.01	0.25	-0.20	-0.15	-0.42	-0.30	-0.42	-0.01	

Appendix H Plots of chlorophyll *a* vs. DIN or DRP (6-month means) and temperature vs. DIN at each site

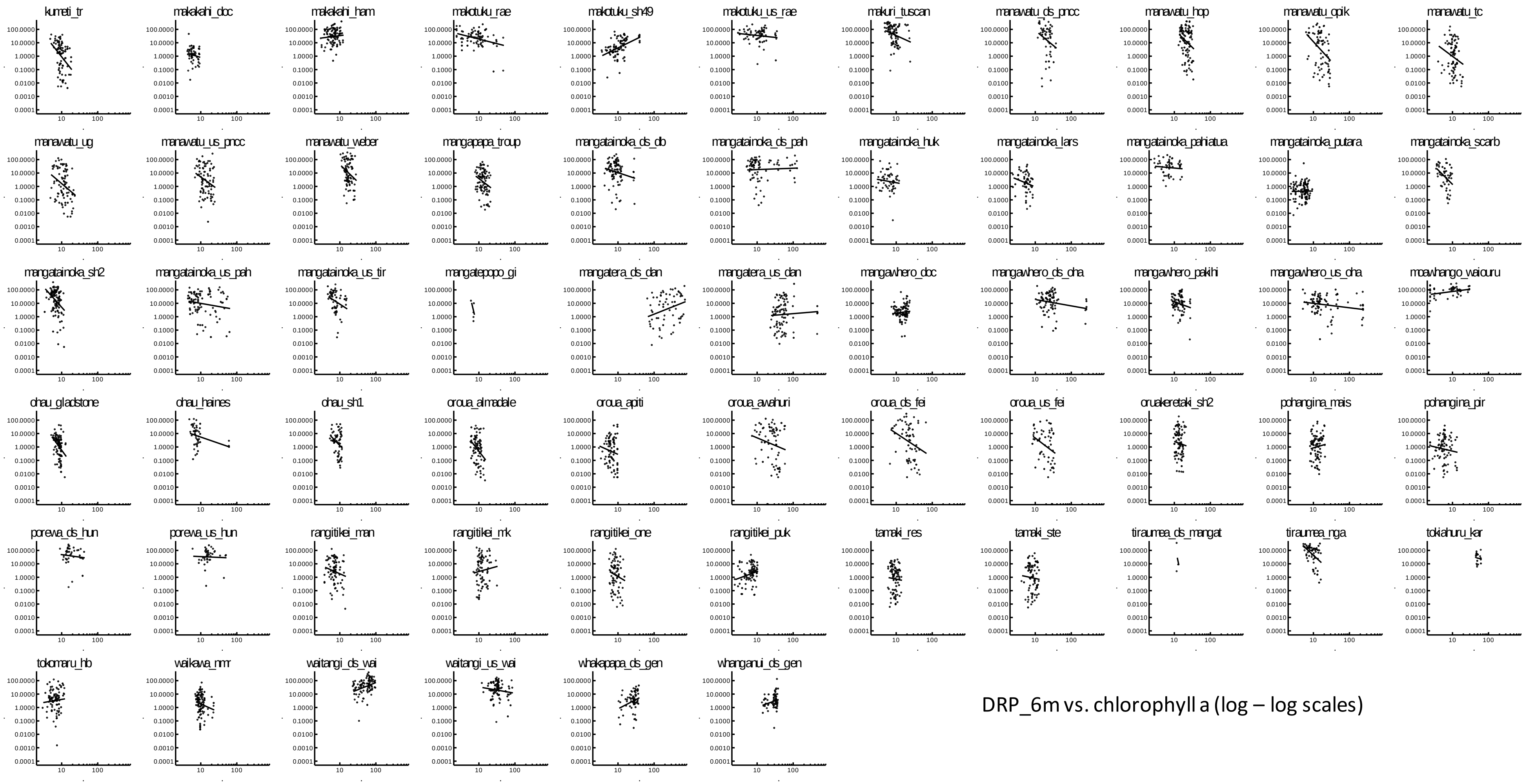
The data shown are unfiltered (i.e., all chlorophyll *a* sampling occasions at all flows). Plots are in alphabetical order of site abbrevations.



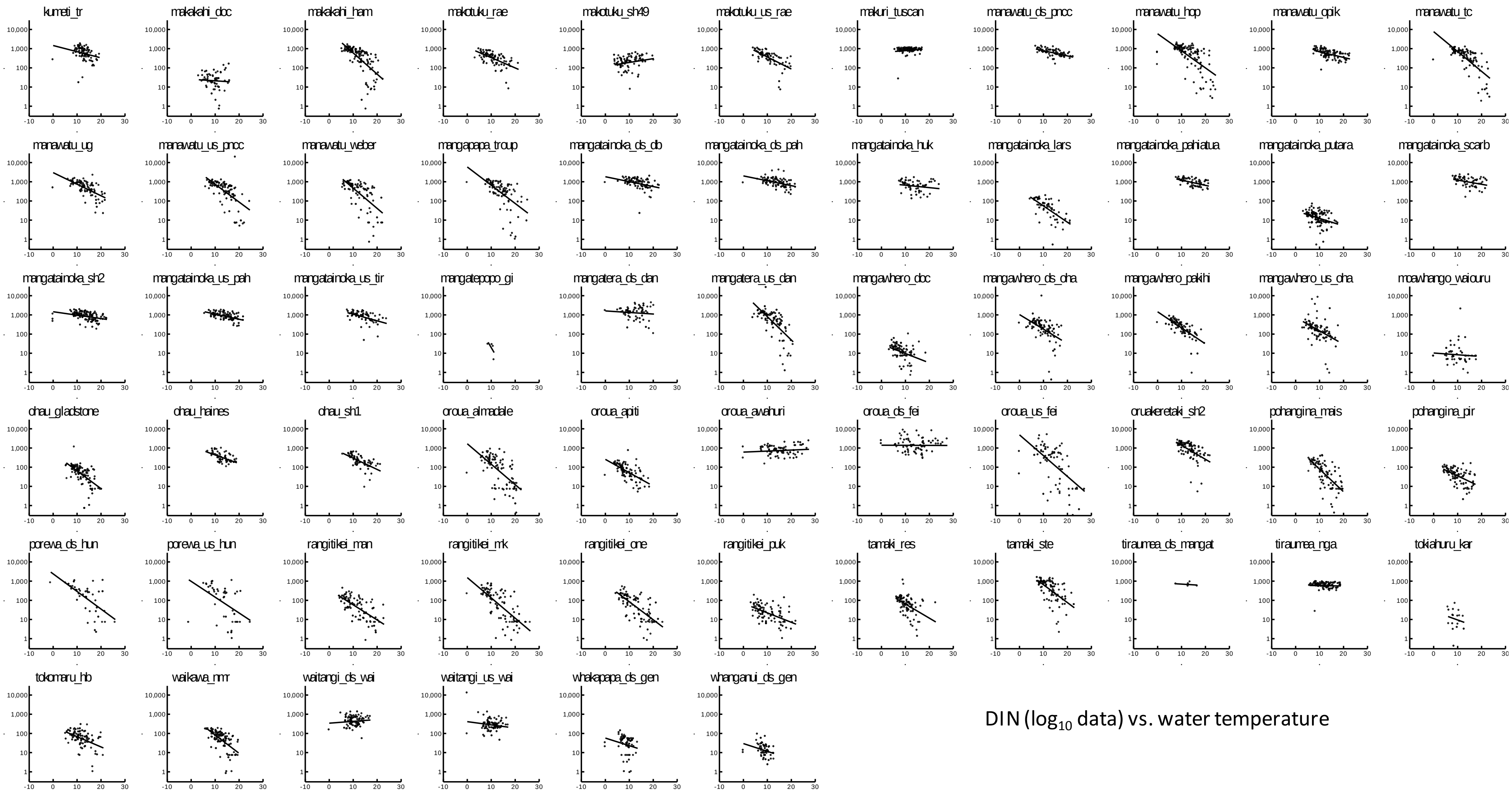
chlorophyll *a* vs. DIN_6m (log – log scales)



chlorophyll *a* vs. DRP_6m (log – log scales)

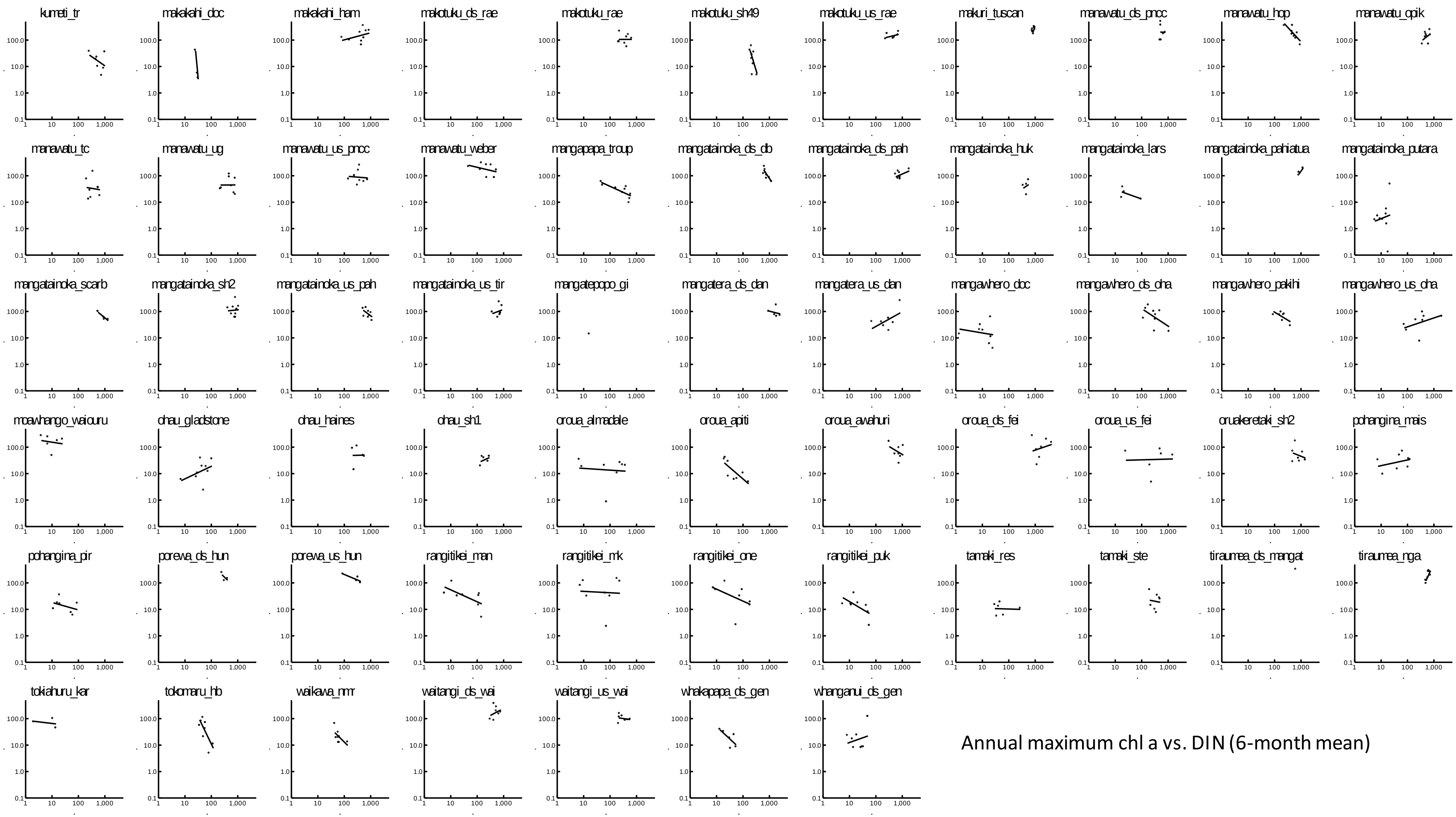


DRP_6m vs. chlorophyll a (log – log scales)

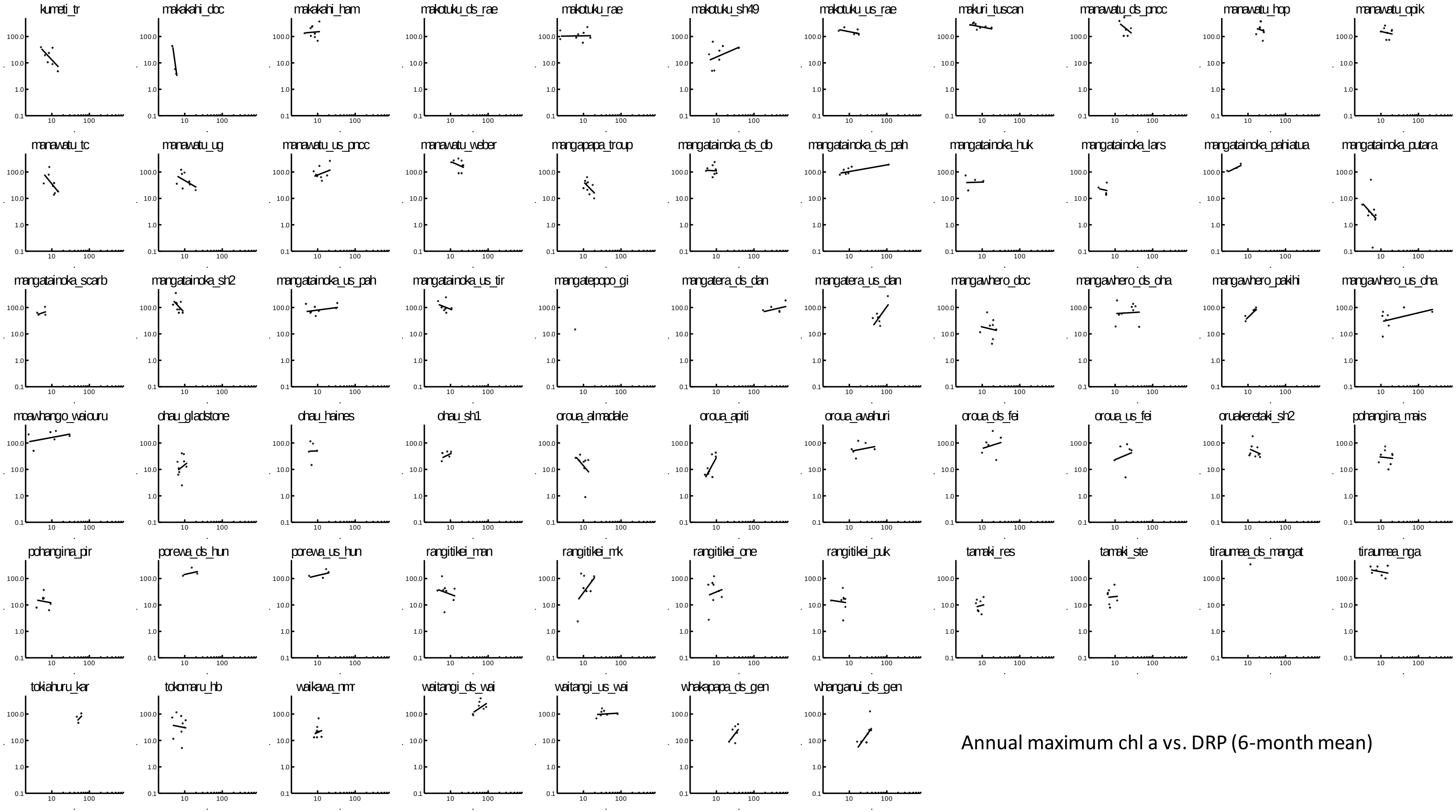


DIN (log_{10} data) vs. water temperature

Appendix I Plots of annual maximum chlorophyll *a* vs. DIN or DRP (6-month means) at each site



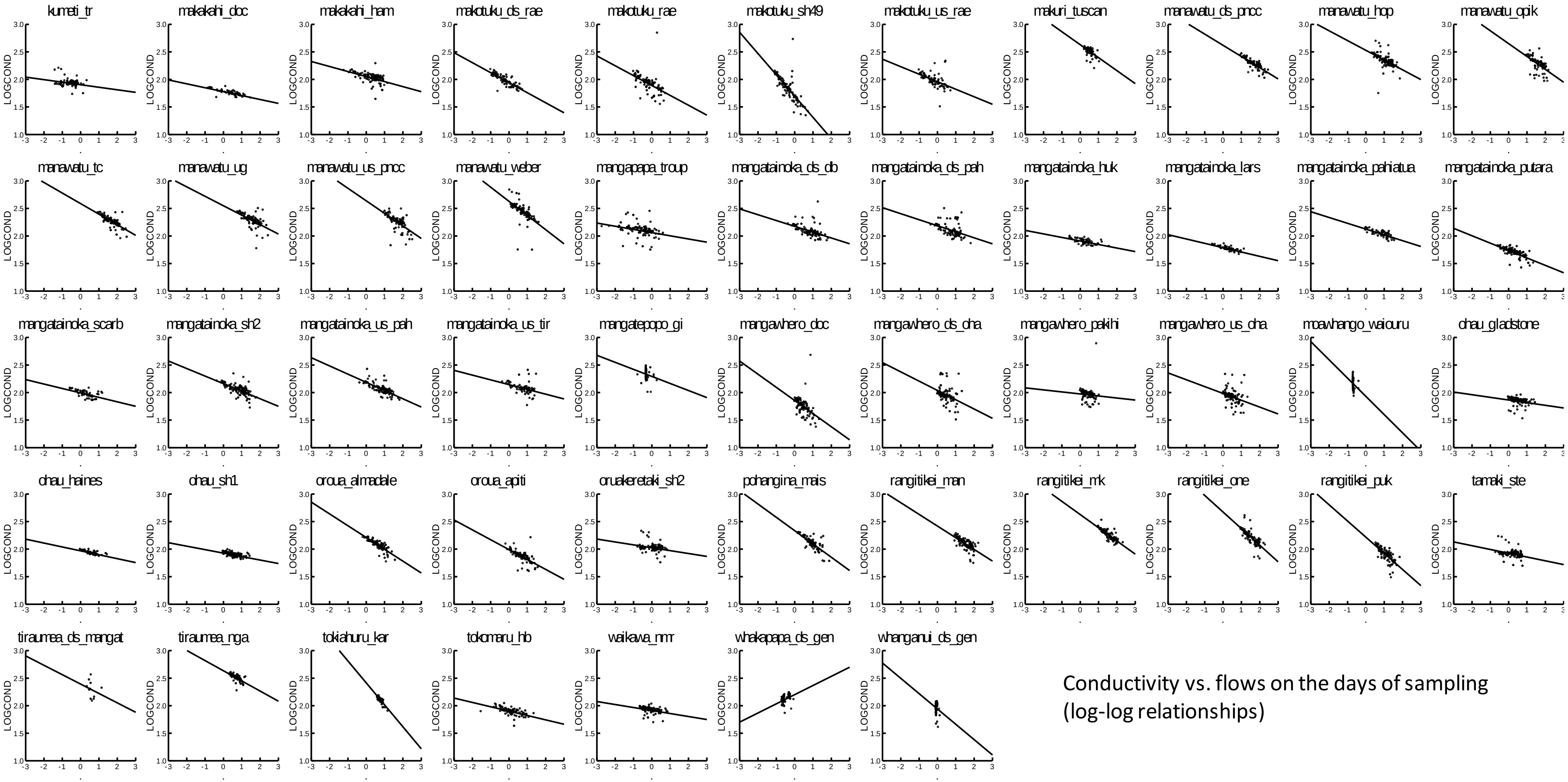
Annual maximum chl a vs. DIN (6-month mean)



Annual maximum chl a vs. DRP (6-month mean)

Appendix J Plots of conductivity vs. mean flow on the day of sampling, at sites with a flow record

The plots highlight strong negative relationships between conductivity and flow magnitude at most sites, driven by dilution during high flows.



Conductivity vs. flows on the days of sampling (log-log relationships)

Appendix K Site classifications based on relationships, water quality, habitat and catchment variables

Classifications are presented in the four tables below, covering within-site relationships (Table K-1), nutrients and water quality (Table K-2), site hydrological and physical characteristics (Table K-3) and catchment characteristics (Table K-4). In all tables sites are listed in order of their Horizons site number (HRCn).

Table K-1: Classification of all sites on the basis of within-site relationships between chlorophyll *a* and flow, nutrients and cover. The tables or sections of the report where each classification (and the groups within it) are shown on the top line. The number in the second line corresponds to the classification number in Table 10-1.

		Table 3.1		Table 7-4				Table 7-4				Table 9-3	
			1		2		5	6	3		4		8
HRCn	Site	Effective flow	Flow group	NSE, chla vs. flow	Str_accrual	Variables	n_incl	p_incl	NSE_all	Str_all	Chl_cover R2	Str_cover	Vars, cover
1	makakahi_doc	11	C	0	vweak	f	x	x	0.03	vweak	0.25	low	FM
2	mangatainoka_putara	10	C	0.09	vweak	f	x	x	0.11	vweak	0.09	vlow	FM
3	mangatainoka_lars	14	B	0.41	mod	fn-t	n-	x	0.47	mod	0.48	low	FGf
4	tamaki_res										0.55	mod	FGfM
5	mangatera_us_dan										0.39	low	FGfM
6	mangatera_ds_dan										0.59	mod	FGfM
7	mangatainoka_huk	10	B	0.21	weak	fn-pt	n-	p	0.47	mod	0.44	low	GfM
8	kumeti_tr	1.5	A	0.44	mod	fp-	x	p-	0.49	mod	0.63	mod	FGfM
9	manawatu_weber	5	D	0.35	mod	fn-	n-	x	0.43	mod	0.57	mod	FGfM
10	makakahi_ham	13	C	0.12	vweak	fn-p	n-	p	0.2	weak	0.42	low	GfM
11	oroua_apiti	3	A	0.36	mod	fn-	n-	x	0.46	mod	0.67	high	FGfM
12	tamaki_ste	3	A	0.41	mod	fn-p-	n-	p-	0.48	mod	0.56	mod	FGfM
13	oruakeretaki_sh2	2	A	0.45	mod	fn-p-	n-	p-	0.61	strong	0.7	high	FGfM
14	makuri_tuscan	3	A	0.28	weak	fnp-	n-	p-	0.34	weak	0.08	vlow	GfM
15	pohangina_pir										0.59	mod	FGfM
16	mangatainoka_scarb	15	B	0.1	vweak	fn-	n-	x	0.18	weak	0.54	mod	GfM
17	tiraumea_nga	4	A	0.5	strong	fp-	x	p-	0.58	strong	0.57	mod	FGfM
18	mangatainoka_pahiatua	4	D								0.38	low	FGfM

		Table 3.1		Table 7-4				Table 7-4				Table 9-3	
			1		2		5	6	3		4		8
HRCn	Site	Effective flow	Flow group	NSE, chla vs. flow	Str_accrual	Variables	n_incl	p_incl	NSE_all	Str_all	Chl_cover R2	Str_cover	Vars, cover
19	mangatainoka_sh2	10	B	0.47	mod	fn-	n-	x	0.5	mod	0.67	high	FGfM
20	mangatainoka_ds_db	11	B	0.3	weak	fn-	n-	x	0.39	mod	0.73	high	FGfM
21	mangatainoka_us_pah	10	B	0.43	mod	fn-	n-	x	0.47	mod	0.59	mod	FGfM
22	mangatainoka_ds_pah	11	B	0.35	mod	fn-p	n-	p	0.43	mod	0.69	high	FGfM
23	manawatu_hop	1.5	A	0.43	mod	f	x	x	0.45	mod	0.72	high	FGfM
24	mangatainoka_us_tir	11	B	0.33	weak	fn-	n-	x	0.38	mod	0.7	high	FGfM
26	mangapapa_troup	10	B	0.26	weak	fn-t	n-	x	0.42	mod	0.65	mod	FGfM
27	pohangina_mais	4	A	0.25	weak	fpt	x	p	0.28	weak	0.43	low	FGfM
28	manawatu_ug	3	A	0.38	mod	f	x	x	0.46	mod	0.74	high	FGfM
29	oroua_almadale	3	A	0.17	vweak	fn	n	x	0.19	weak	0.47	low	FGfM
30	oroua_us_fei										0.62	mod	FGfM
31	oroua_ds_fei										0.79	high	FGfM
32	oroua_awahuri										0.69	high	FGfM
33	manawatu_tc	2	A	0.41	mod	ft	x	x	0.47	mod	0.62	mod	FGf
34	manawatu_us_pncc	3	A	0.49	mod	fp	x	p	0.58	strong	0.74	high	FGfM
35	manawatu_ds_pncc	4	A	0.41	mod	f	x	x	0.45	mod	0.75	high	FGfM
36	manawatu_opik	2	A	0.48	mod	fnp-t	n	p-	0.57	strong	0.8	high	FGfM
37	tokomaru_hb		C								0.56	mod	FGfM
38	rangitikei_puk	4	A	0.28	weak	fn-	n-	x	0.27	weak	0.5	mod	FGfM
39	moawhango_waiouru	5	0	0.14	vweak	f	x	x	0.17	weak	0		
40	rangitikei_man	4	A	0.26	weak	fn-	n-	x	0.23	weak	0.28	low	GfM
41	porewa_us_hun										0.02	vlow	M
42	porewa_ds_hun										0.23	vlow	M

		Table 3.1		Table 7-4				Table 7-4				Table 9-3	
			1		2		5	6	3		4		8
HRCn	Site	Effective flow	Flow group	NSE, chla vs. flow	Str_accrual	Variables	n_incl	p_incl	NSE_all	Str_all	Chl_cover R2	Str_cover	Vars, cover
43	rangitikei_one	4	A	0.29	weak	fn-p	n-	p	0.41	mod	0.45	low	FGfM
44	rangitikei_mk	4	A	0.35	mod	fn-p	n-	p	0.41	mod	0.49	low	GfM
45	mangawhero_doc	0	C								0.17	vlow	Gf
46	makotuku_sh49	8	C	0.08	vweak	fnp	n	p	0.28	weak	0.21	vlow	FGfM
47	mangawhero_us_oha		C								0.46	low	FGfM
48	mangawhero_ds_oha		C								0.36	low	FGfM
49	makotuku_rae		C								0.31	low	FGfM
50	mangawhero_pakihi	8	A	0.28	weak	fn-p-	n-	p-	0.31	weak	0.21	vlow	FGfM
51	mangatepopo_gi	15	B								0.21	vlow	GfM
52	whanganui_ds_gen										0.15	vlow	FM
53	whakapapa_ds_gen	3		0.29	weak	fpt	x	p	0.3	weak	0.5	mod	FGfM
54	waitangi_us_wai										0.18	vlow	FGfM
55	waitangi_ds_wai										0.28	low	GfM
56	tokiahuru_kar	5									0.31	low	FM
57	makotuku_us_rae		C								0.25	low	GfM
58	makotuku_ds_rae		C								0.24	vlow	GfM
59	waikawa_nmr	2	A	0.17	vweak	fn-pt	n-	p	0.21	weak	0.43	low	FGfM
60	ohau_gladstone	6	C	0.18	vweak	fn-pt	n-	p	0.2	weak	0.31	low	FGf
61	ohau_sh1	3	A	0.33	weak	f	x	x	0.38	mod	0.57	mod	FGfM
62	ohau_haines	5	A	0.27	weak	f	x	x	0.28	weak	0.64	mod	FGf

Table K-2: Classification of all sites on the basis of nutrient limitation and nutrient / water quality variables. The tables or sections of the report where each classification (and the groups within it) are shown on the top line. The number in the second line corresponds to the classification number in Table 10-1. DIN and DRP are geometric means calculaed from all available data (up to 7+ years) at each site; mean conductivity and chlorophyll *a* (chla) are the long-term arithmetic means; 92nd percentile of chlorophyll *a* was calcualted from all available data at each site.

		Nutrient limitation, Table 4-2				DIN, Section 2		DRP, Section 2		Conductivity		Mean chla, Section 2		92nd percentile chla , Section 2	
		9	10	11	12		13		14		15		16		17
HRCn	Site	Classical, all	Classical, low flows	Conc, all	Conc., low flows	DIN_ ppb	DIN group	DRP_ppb	DRP group	Cond. µS/cm	Cond. group	Mean chla	Chla group	Chla 92nd	Chla 92nd group
1	makakahi_doc	N	N	co	co	37	vlow	6.7	low	56	low	2.2	vlow	5	vlow
2	mangatainoka_putara	N	N	co	co	17	vlow	5.3	low	50	low	0.9	vlow	2.227	vlow
3	mangatainoka_lars	co	N	co	co	60	low	6.2	low	57	low	5.9	low	16.41	low
4	tamaki_res	N		co		86	low	9.6	mod	69	low	3.2	vlow	10.97	vlow
5	mangatera_us_dan	N		none		646	vhigh	51.6	xhigh	153	high	11.4	low	35.86	low
6	mangatera_ds_dan	co		none		1530	xhigh	322.5	xhigh	187	high	25.4	mod	75.1	mod
7	mangatainoka_huk	P	P	P	P	667	vhigh	6.9	low	77	low	7.2	low	21.4	low
8	kumeti_tr	P	P	P	P	667	vhigh	10.1	high	83	low	5.3	low	18.44	low
9	manawatu_weber	co	co	none-co	co-none	429	high	18.2	vhigh	269	vhigh	43.0	high	161.7	high
10	makakahi_ham	P	P	P	co-P	519	high	6.8	low	106	mod	48.5	high	117.1	mod
11	oroua_apiti	co	N	co	co	80	low	6.8	low	73	low	2.6	vlow	8.17	vlow
12	tamaki_ste	P	P	P-co	co-P	518	high	8.8	mod	79	low	5.2	low	13.63	vlow
13	oruakeretaki_sh2	P	P	P	none-P	1020	xhigh	14.3	high	101	mod	11.6	low	38	low
14	makuri_tuscan	P	P	P	P	862	xhigh	9.4	mod	321	vhigh	89.1	vhigh	245.1	vhigh
15	pohangina_pir	N		co		46	vlow	6.0	low	70	low	3.3	vlow	9.84	vlow
16	mangatainoka_scarb	P	P	P	P	1061	xhigh	6.0	low	92	mod	16.5	mod	51	mod
17	tiraumea_nga	P	P	P	P	604	vhigh	10.6	high	297	vhigh	97.4	vhigh	207.7	vhigh
18	mangatainoka_pahiatua	P	P	P	P	964	xhigh	9.0	mod	108	mod	46.1	high	135.2	high
19	mangatainoka_sh2	P	P	P	P	887	xhigh	7.1	mod	112	mod	41.7	high	112.6	mod

		Nutrient limitation, Table 4-2				DIN, Section 2		DRP, Section 2		Conductivity		Mean chla, Section 2		92nd percentile chla , Section 2	
		9	10	11	12		13		14		15		16		17
HRCn	Site	Classical, all	Classical, low flows	Conc, all	Conc., low flows	DIN_ ppb	DIN group	DRP_ppb	DRP group	Cond. µS/cm	Cond. group	Mean chla	Chla group	Chla 92nd	Chla 92nd group
20	mangatainoka_ds_db	P	P	P	P	924	xhigh	10.0	high	119	mod	33.5	high	105	mod
21	mangatainoka_us_pah	P	P	P	P	896	xhigh	13.1	high	113	mod	29.1	mod	70	mod
22	mangatainoka_ds_pah	P	P	P	P	1018	xhigh	20.3	vhigh	121	mod	45.4	high	102.5	mod
23	manawatu_hop	P	co	none	none-N	637	vhigh	23.3	vhigh	211	vhigh	57.7	vhigh	168.4	high
24	mangatainoka_us_tir	P	P	P	P	854	xhigh	8.2	mod	120	mod	34.1	high	85.4	mod
26	mangapapa_troup	P	co	P-co	co-P	447	high	13.6	high	122	mod	7.8	low	29.66	low
27	pohangina_mais	N	N	co	co	87	low	13.4	high	129	mod	5.0	vlow	15.05	low
28	manawatu_ug	P	P	P-co	P-co	596	vhigh	11.6	high	186	high	12.6	low	41.5	low
29	oroua_almadale	N	N	co	co	160	mod	9.5	mod	115	mod	3.5	vlow	15.56	low
30	oroua_us_fei	co		none-co		565	vhigh	20.5	vhigh	141	high	9.5	low	40.29	low
31	oroua_ds_fei	P		none-P		1854	xhigh	22.2	vhigh	171	high	27.3	mod	94.9	mod
32	oroua_awahuri	P		none		856	xhigh	27.0	vhigh	164	high	17.4	mod	55	mod
33	manawatu_tc	P	P	co-P	co	439	high	11.1	high	180	high	8.7	low	31.25	low
34	manawatu_us_pncc	P	P	co-none-P	co-P	697	vhigh	14.7	high	173	high	20.0	mod	70	mod
35	manawatu_ds_pncc	P	P	none	none-P	635	vhigh	20.0	vhigh	185	high	65.6	vhigh	253.1	vhigh
36	manawatu_opik	P	P	none-P	none-P	577	vhigh	16.7	vhigh	173	high	32.2	high	120.8	high
37	tokomaru_hb	co	N	co	co	74	low	6.9	low	78	low	10.7	low	31.79	low
38	rangitikei_puk	N	N	co	co	30	vlow	6.8	low	78	low	4.6	vlow	13.87	vlow
39	moawhango_waiouru	N	N	co	co	49	vlow	11.1	high	142	high	87.9	vhigh	177.6	high
40	rangitikei_man	N	N	co	co	67	low	7.4	mod	124	mod	10.5	low	32.75	low
41	porewa_us_hun	N		N-co		208	mod	19.6	vhigh	269	vhigh	51.8	vhigh	124.2	high
42	porewa_ds_hun	N		N-co		274	mod	21.6	vhigh	272	vhigh	67.9	vhigh	145	high
43	rangitikei_one	N	N	co	co	88	low	9.6	mod	156	high	9.8	low	40.14	low

		Nutrient limitation, Table 4-2				DIN, Section 2		DRP, Section 2		Conductivity		Mean chla, Section 2		92nd percentile chla , Section 2	
		9	10	11	12		13		14		15		16		17
HRCn	Site	Classical, all	Classical, low flows	Conc, all	Conc., low flows	DIN_ ppb	DIN group	DRP_ppb	DRP group	Cond. µS/cm	Cond. group	Mean chla	Chla group	Chla 92nd	Chla 92nd group
44	rangitikei_mk	N	N	co	co	137	mod	14.8	high	171	high	16.3	mod	58.2	mod
45	mangawhero_doc	N	N	co-N	N	14	vlow	16.0	vhigh	61	low	4.8	vlow	11.43	vlow
46	makotuku_sh49	P	P	co	co-N	236	mod	12.9	high	77	low	10.3	low	33.72	low
47	mangawhero_us_oha	co	N	co	co	415	high	49.4	xhigh	86	low	17.4	mod	48.51	low
48	mangawhero_ds_oha	co	N	N-co	N-co	351	high	41.9	xhigh	92	mod	27.5	mod	70.3	mod
49	makotuku_rae	P	P	P-co	co-P	361	high	10.2	high	92	mod	39.7	high	96.2	mod
50	mangawhero_pakihi	P	co	co-N	co-N	262	mod	13.8	high	96	mod	20.5	mod	69.1	mod
51	mangatepopo_gi					20	vlow	6.9	low	213	vhigh	4.8	vlow	12.83	vlow
52	whanganui_ds_gen	N	N	N	N	15	vlow	29.4	vhigh	91	mod	6.2	low	15.28	low
53	whakapapa_ds_gen	N	N	N	N	36	vlow	26.3	vhigh	130	mod	7.4	low	19.85	low
54	waitangi_us_wai	co		N		449	high	34.3	vhigh	168	high	37.2	high	94.3	mod
55	waitangi_ds_wai	co		none		498	high	57.3	xhigh	179	high	66.6	vhigh	171.5	high
56	tokiahuru_kar					13	vlow	51.8	xhigh	126	mod	18.0	mod	49.19	low
57	makotuku_us_rae	P	P	P-co	P-co	405	high	12.4	high	98	mod	62.2	vhigh	131.6	high
58	makotuku_ds_rae									93	mod	109.1	vhigh	217.9	vhigh
59	waikawa_nmr	N	N	co	co	67	low	10.8	high	82	low	5.1	low	13.44	vlow
60	ohau_gladstone	N	N	co	co	57	low	8.6	mod	69	low	2.8	vlow	6.5	vlow
61	ohau_sh1	P	P	co	co	242	mod	115.9	xhigh	78	low	5.6	low	21.06	low
62	ohau_haines	P	P	co	co	341	high	7.5	mod	85	low	18.6	mod	71.6	mod

Table K-3: Classification of all sites on the basis of site-specific hydrological and physical variables. The tables or sections of the report where each classification (and the groups within it) are shown on the top line. The number in the second line corresponds to the classification number in Table 10-1.

						Hydrological variables (see Section 2)							Physical variables (see Section 2)			
		22		18		19		20		21		23		24		25
HRCn	Site	Stream order	Mean flow m^3/s	Size group	FREeff	FREeff group	FRE3	FRE3 group	FRE10	FRE10 group	% coarse	Coarse group	% fine	Fine group	Altitude m. asl	Alt. group
1	makakahi_doc	3	6.3	medium	4.0	low	12.4	vhigh	5.2	low	25.8	vhigh	7.3	low	397	hill
2	mangatainoka_putara	4	5.0	medium	8.9	high	16.0	vhigh	9.7	high	28.3	vhigh	8.5	mod	388	hill
3	mangatainoka_lars	4	5.0	medium	5.2	low	16.0	vhigh	9.7	high	19.7	high	8.8	mod	302	hill
4	tamaki_res	3									14.1	high	6.9	low	400	hill
5	mangatera_us_dan	4									0.1	vlow	20.7	vhigh	181	low
6	mangatera_ds_dan	4									1.6	vlow	22.6	vhigh	181	low
7	mangatainoka_huk	2	5.0	medium	8.9	high	16.0	vhigh	9.7	high	11.6	high	10.8	high	197	low
8	kumeti_tr	3	0.5	small	6.6	mod	7.0	mod	2.0	low	2.4	low	10.5	high	275	hill
9	manawatu_weber	6	14.0	large	6.6	mod	8.8	high	4.6	low	10.1	high	7.9	mod	176	low
10	makakahi_ham	1	6.3	medium	3.1	low	12.4	vhigh	5.2	low	5.4	mod	6.5	low	205	hill
11	oroua_apiti	4	9.3	medium	9.2	high	9.9	high	2.0	low	7.0	mod	11.0	high	461	hill
12	tamaki_ste	4	3.3	small	8.3	high	9.0	high	1.8	vlow	7.5	mod	8.1	mod	168	low
13	oruakeretaki_sh2	4	2.1	small	9.9	high	10.1	high	2.3	low	6.1	mod	10.7	high	160	low
14	makuri_tuscan	5	5.1	medium	9.3	high	10.0	high	1.8	vlow	27.0	vhigh	12.1	high	146	low
15	pohangina_pir	1									21.1	vhigh	11.0	high	287	hill
16	mangatainoka_scarb	5	5.0	medium	4.4	low	16.0	vhigh	9.7	high	9.6	mod	11.3	high	125	low
17	tiraumea_nga	6	16.1	large	9.4	high	9.7	high	5.2	low	8.6	mod	8.1	mod	100	low
18	mangatainoka_pahiatua	6	18.0	large	10.1	high	12.5	vhigh	4.6	low	8.2	mod	10.6	high	112	low
19	mangatainoka_sh2	6	18.0	large	4.3	low	12.5	vhigh	4.6	low	10.4	high	6.3	low	93	low
20	mangatainoka_ds_db	6	18.0	large	3.6	low	12.5	vhigh	4.6	low	9.1	mod	7.0	low	93	low
21	mangatainoka_us_pah	6	18.0	large	4.3	low	12.5	vhigh	4.6	low	10.4	high	7.0	low	100	low

		Hydrological variables (see Section 2)									Physical variables (see Section 2)					
		22		18		19		20		21		23		24		25
HRCn	Site	Stream order	Mean flow m³/s	Size group	FREeff	FREeff group	FRE3	FRE3 group	FRE10	FRE10 group	% coarse	Coarse group	% fine	Fine group	Altitude m. asl	Alt. group
22	mangatainoka_ds_pah	6	18.0	large	3.6	low	12.5	vhigh	4.6	low	12.9	high	6.4	low	105	low
23	manawatu_hop	6	26.9	large	8.2	high	8.7	high	3.7	low	2.6	low	8.0	mod	101	low
24	mangatainoka_us_tir	6	18.0	large	3.6	low	12.5	vhigh	4.6	low	12.3	high	7.5	low	82	low
26	mangapapa_troup	4	0.7	small	4.6	low	10.8	vhigh	5.0	low	1.3	vlow	16.0	vhigh	68	low
27	pohangina_mais	2	16.9	large	8.8	high	11.4	vhigh	2.8	low	11.8	high	14.1	high	200	hill
28	manawatu_ug	5	87.2	vlarge	11.1	vhigh	12.0	vhigh	1.9	vlow	2.3	low	18.0	vhigh	60	low
29	oroua_almadale	5	9.3	medium	9.2	high	9.9	high	2.0	low	7.9	mod	9.6	mod	150	low
30	oroua_us_fei	6									1.4	vlow	13.2	high	53	low
31	oroua_ds_fei	6									1.7	vlow	10.6	high	47	low
32	oroua_awahuri	6									2.8	low	11.0	high	38	low
33	manawatu_tc	7	107.0	vlarge	10.6	vhigh	11.4	vhigh	2.0	low	4.3	low	8.9	mod	83	low
34	manawatu_us_pncc	7	107.0	vlarge	10.6	vhigh	11.4	vhigh	2.0	low	2.3	low	8.7	mod	38	low
35	manawatu_ds_pncc	7	107.0	vlarge	8.4	high	11.4	vhigh	2.0	low	4.8	low	6.8	low	19	low
36	manawatu_opik	1	107.0	vlarge	10.6	vhigh	11.4	vhigh	2.0	low	0.0	vlow	18.7	vhigh	10	low
37	tokomaru_hb	4	2.2	small			10.8	vhigh	2.8	low	19.4	high	11.7	high	88	low
38	rangitikei_puk	5	24.1	large	6.6	mod	9.4	high	1.2	vlow	26.1	vhigh	10.5	high	467	hill
39	moawhango_waiouru	1	0.2	small	0.5	vlow	1.0	vlow	0.1	vlow	5.9	mod	5.0	low	775	upland
40	rangitikei_man	7	64.2	vlarge	5.0	low	8.0	mod	1.1	vlow	23.3	vhigh	5.6	low	260	hill
41	porewa_us_hun	4									4.6	low	10.3	high	272	hill
42	porewa_ds_hun	4									3.5	low	9.8	mod	260	hill
43	rangitikei_one	2	68.3	vlarge	6.1	mod	8.2	mod	1.0	vlow	5.3	mod	9.5	mod	95	low
44	rangitikei_mk	3	73.0	vlarge	6.4	mod	8.6	high	1.5	vlow	3.4	low	9.6	mod	29	low
45	mangawhero_doc	3	4.7	medium			7.8	mod	0.8	vlow	34.3	vhigh	6.2	low	680	upland

					Hydrological variables (see Section 2)								Physical variables (see Section 2)			
		22		18		19		20		21		23		24		25
HRCn	Site	Stream order	Mean flow m^3/s	Size group	FREeff	FREeff group	FRE3	FRE3 group	FRE10	FRE10 group	% coarse	Coarse group	% fine	Fine group	Altitude m. asl	Alt. group
46	makotuku_sh49	1	0.9	small	7.1	mod	13.3	vhigh	5.4	low	13.2	high	18.4	vhigh	629	upland
47	mangawhero_us_oha	1	4.7	medium			7.8	mod	0.8	vlow	12.7	high	9.7	mod	569	hill
48	mangawhero_ds_oha	4	4.7	medium			7.8	mod	0.8	vlow	19.8	high	9.3	mod	560	hill
49	makotuku_rae	4	1.7	small			11.0	vhigh	5.3	low	26.0	vhigh	7.9	mod	534	hill
50	mangawhero_pakihi	5	4.7	medium	1.7	vlow	7.8	mod	0.8	vlow	30.2	vhigh	15.1	vhigh	537	hill
51	mangatepopo_gi	3	0.4	small	1.7	vlow	4.4	low	2.2	low	35.5	vhigh	8.8	mod	740	upland
52	whanganui_ds_gen	3	1.0	small			0.1	vlow	0.1	vlow	24.5	vhigh	15.4	vhigh	646	upland
53	whakapapa_ds_gen	5	0.4	small	3.7	low	4.0	low	0.1	vlow	38.2	vhigh	10.6	high	739	upland
54	waitangi_us_wai	4									9.1	mod	21.8	vhigh	755	upland
55	waitangi_ds_wai	3									7.7	mod	21.6	vhigh	755	upland
56	tokiahuru_kar	4	6.6	medium	0.4	vlow	1.7	vlow	0.1	vlow	3.1	low	20.3	vhigh	622	upland
57	makotuku_us_rae	4	1.7	small			11.0	vhigh	5.3	low	9.3	mod	9.5	mod	510	hill
58	makotuku_ds_rae	4	1.7	small			11.0	vhigh	5.3	low	9.1	mod	14.3	high	510	hill
59	waikawa_nmr	4	1.4	small	13.6	vhigh	12.0	vhigh	3.0	low	15.7	high	14.5	high	66	low
60	ohau_gladstone	1	6.5	medium	6.8	mod	13.4	vhigh	3.0	low	12.0	high	14.9	high	154	low
61	ohau_sh1	5	6.5	medium	12.4	vhigh	13.4	vhigh	3.0	low	9.3	mod	12.5	high	38	low
62	ohau_haines	5	6.5	medium	8.8	high	13.4	vhigh	3.0	low	6.3	mod	16.0	vhigh	19	low

Table K-4: Classification of all sites on the basis of catchment-scale variables. Top Rock and Base Rock are Horizons classifications: AL (alluvium), HS (hard sedimentary), LO (loess), LI (limestone), SS (soft sedimentary), VA (volcanic acidic). Landcover variables were extracted from the LCDB3 database (see Section 2.4). LSC is the Horizons life-supporting capacity classification (derived from geology and source of flow). ds_stp indicates sites that are downstream of waste-water treatment plants or other discharges.

		Dominant geology			Land cover								
		26	27	28		29		30		31	33	32	34
HRCn	Site	Top Rock	Base Rock	REC climate	% farm	Farm group	%lo_grass	Grass group	%indig_for	Forest group	LSC class	Subregion	ds_stp
1	makakahi_doc	HS	HS	CX	0.0	vlow	0.28	low	84.2	xhigh	HM	Mana_8d	no
2	mangatainoka_putara	HS	HS	CX	0.0	vlow	0.00	vlow	79.4	xhigh	UHS	Mana_8a	no
3	mangatainoka_lars	HS	HS	CX	31.5	high	0.94	low	49.0	high	UHS	Mana_8a	no
4	tamaki_res	HS	HS	CW	3.4	low	0.00	vlow	3.8	low	UHS	Mana_3	no
5	mangatera_us_dan	LO	SS	CW	91.9	xhigh	0.05	low	0.9	vlow	HM	Mana_2b	no
6	mangatera_ds_dan	LO	SS	CW	91.9	xhigh	0.05	low	0.9	vlow	HM	Mana_2b	yes
7	mangatainoka_huk	HS	HS	CX	56.8	vhigh	0.44	low	31.0	high	HM	Mana_8b	no
8	kumeti_tr	HS	HS	CW	34.7	high	0.00	vlow	0.1	vlow	UHS	Mana_4	no
9	manawatu_weber	SS	SS	CW	88.5	xhigh	0.38	low	3.6	low	HM	Mana_1a	no
10	makakahi_ham	SS	SS	CW	78.6	xhigh	0.14	low	10.8	mod	HM	Mana_8d	no
11	oroua_apiti	SS	SS	CW	10.7	mod	0.54	low	55.6	vhigh	HM	Mana_12a	no
12	tamaki_ste	HS	HS	CW	58.7	vhigh	3.57	mod	2.0	vlow	HM	Mana_5b	no
13	oruakeretaki_sh2	HS	AL	CW	67.7	vhigh	0.00	vlow	0.5	vlow	HM	Mana_5d	no
14	makuri_tuscan	HS	LI	CW	78.8	xhigh	0.10	low	7.0	low	ULi	Mana_7d	no
15	pohangina_pir	HS	HS	CW	14.1	mod	0.31	low	20.3	high	UHS	Mana_10b	no
16	mangatainoka_scarb	AL	AL	CW	68.6	vhigh	0.31	low	21.7	high	HM	Mana_8b	no
17	tiraumea_nga	SS	SS	CW	82.7	xhigh	0.09	low	3.2	low	HSS	Mana_7b	no
18	mangatainoka_pahiatua	AL	AL	CW	76.2	xhigh	0.20	low	14.2	mod	HM	Mana_8c	no
19	mangatainoka_sh2	AL	AL	CW	76.3	xhigh	0.19	low	13.8	mod	HM	Mana_8c	no
20	mangatainoka_ds_db	AL	AL	CW	76.3	xhigh	0.19	low	13.8	mod	HM	Mana_8c	yes
21	mangatainoka_us_pah	AL	AL	CW	76.1	xhigh	0.19	low	14.0	mod	HM	Mana_8c	no

		Dominant geology			Land cover								
		26	27	28		29		30		31	33	32	34
HRCn	Site	Top Rock	Base Rock	REC climate	% farm	Farm group	%lo_grass	Grass group	%indig_for	Forest group	LSC class	Subregion	ds_stp
22	mangatainoka_ds_pah	AL	AL	CW	76.1	xhigh	0.19	low	14.0	mod	HM	Mana_8c	yes
23	manawatu_hop	SS	SS	CW	85.6	xhigh	0.44	low	2.5	low	HM	Mana_5a	no
24	mangatainoka_us_tir	AL	AL	CW	78.2	xhigh	0.16	low	12.0	mod	HM	Mana_8c	no
26	mangapapa_troup	HS	HS	CW	68.4	vhigh	3.54	mod	6.8	low	HM	Mana_9b	no
27	pohangina_mais	SS	SS	CW	48.4	high	0.99	low	11.9	mod	HM	Mana_10c	no
28	manawatu_ug	SS	SS	CW	78.6	xhigh	0.33	low	7.5	low	HM	Mana_9a	no
29	oroua_almadale	SS	SS	CW	60.2	vhigh	0.52	low	18.5	mod	HM	Mana_12a	no
30	oroua_us_fei	SS	SS	CW	75.3	xhigh	0.71	low	10.1	mod	HM	Mana_12b	no
31	oroua_ds_fei	SS	SS	CW	75.3	xhigh	0.71	low	10.1	mod	HM	Mana_12b	yes
32	oroua_awahuri	LO	SS	CD	78.8	xhigh	0.57	low	8.1	low	LM	Mana_12c	no
33	manawatu_tc	SS	SS	CW	74.5	vhigh	0.46	low	7.8	low	HM	Mana_10a	no
34	manawatu_us_pncc	SS	SS	CW	73.9	vhigh	0.62	low	7.9	low	HM	Mana_11a	no
35	manawatu_ds_pncc	SS	SS	CW	73.2	vhigh	0.92	low	7.8	low	HM	Mana_11a	yes
36	manawatu_opik	SS	SS	CW	73.8	vhigh	0.87	low	7.4	low	HM	Mana_11a	no
37	tokomaru_hb	HS	HS	CW	0.2	vlow	11.16	high	38.3	high	LM	Mana_13c	no
38	rangitikei_puk	VA	HS	CW	30.3	high	1.50	mod	22.2	high	UHS	Rang_2a	no
39	moawhango_waiouru	VA	HS	CW	0.2	vlow	1.05	mod	11.3	mod	UVM	Rang_2d	no
40	rangitikei_man	VA	SS	CW	41.7	high	1.03	mod	20.0	mod	HM	Rang_3a	no
41	porewa_us_hun	SS	SS	CD	72.9	vhigh	0.02	vlow	3.4	low	HSS	Rang_4c	no
42	porewa_ds_hun	SS	SS	CD	73.2	vhigh	0.01	vlow	3.4	low	HSS	Rang_4c	yes
43	rangitikei_one	VA	SS	CW	49.4	high	0.90	low	16.8	mod	HM	Rang_3a	no
44	rangitikei_mk	SS	SS	CW	55.4	vhigh	0.80	low	14.4	mod	HM	Rang_4a	no
45	mangawhero_doc	VA	VA	CW	37.2	high	0.00	vlow	54.4	vhigh	UVA	Whau_3d	no

		Dominant geology			Land cover								
		26	27	28		29		30		31	33	32	34
HRCn	Site	Top Rock	Base Rock	REC climate	% farm	Farm group	%lo_grass	Grass group	%indig_for	Forest group	LSC class	Subregion	ds_stp
46	makotuku_sh49	VA	VA	CW	20.3	mod	0.64	low	62.7	vhigh	UVA	Whau_3b	no
47	mangawhero_us_oha	VA	VA	CW	27.7	high	1.31	mod	55.2	vhigh	UVA	Whau_3d	no
48	mangawhero_ds_oha	VA	VA	CW	29.2	high	1.27	mod	54.1	vhigh	UVA	Whau_3d	yes
49	makotuku_rae	VA	VA	CW	58.0	vhigh	0.68	low	30.5	high	UVA	Whau_3c	no
50	mangawhero_pakihi	VA	VA	CW	45.6	high	0.67	low	43.7	high	UVA	Whau_3d	no
51	mangatepopo_gi	VA	VA	CX	0.0	vlow	0.00	vlow	4.0	low	UVA	Whai_1	no
52	whanganui_ds_gen	VA	VA	CX	0.0	vlow	0.37	low	12.3	mod	UVA	Whai_1	no
53	whakapapa_ds_gen	VA	VA	CX	12.6	mod	0.00	vlow	21.6	high	UVA	Whai_2b	no
54	waitangi_us_wai	VA	VA	CD	27.5	high	0.00	vlow	0.2	vlow	UVM	Whau_1b	no
55	waitangi_ds_wai	VA	VA	CD	27.5	high	0.00	vlow	0.2	vlow	UVM	Whau_1b	yes
56	tokiahuru_kar	VA	VA	CW	2.8	low	12.35	high	24.0	high	UVA	Whau_1c	no
57	makotuku_us_rae	VA	VA	CW	61.8	vhigh	0.59	low	26.6	high	UVA	Whau_3c	no
58	makotuku_ds_rae	VA	VA	CW	61.8	vhigh	0.59	low	26.6	high	UVA	Whau_3c	yes
59	waikawa_nmr	HS	HS	CW	9.1	low	0.81	low	65.9	vhigh	HM	West_9a	no
60	ohau_gladstone	HS	HS	CW	10.5	mod	1.87	mod	65.8	vhigh	UHS	Ohau_1a	no
61	ohau_sh1	HS	HS	CW	20.4	mod	2.15	mod	54.5	vhigh	HM	Ohau_1b	no
62	ohau_haines	HS	HS	CW	27.5	high	1.94	mod	49.2	high	HM	Ohau_1b	no

Appendix L Example of the use chlorophyll *a* – environment relationships to predict chlorophyll *a* under different scenarios

The example below uses the equation derived from data collected from July 2012 to June 2015. The equation is:

$$\text{Log}_{10}\ 92^{nd} \text{ percentile chlorophyll a} = -1.94 + 0.323\ (\log_{10}\text{DIN}) + 0.33\ (\log_{10}\text{DRP}) + 0.106\ (\text{sqrt Cond}) + 0.04\ (\text{mean temp.}) + (0.524 \log_{10}\text{Da\_EFF})$$

R^2 of the regression was 0.78; R^2 of the cross-validation observed vs. expected was 0.75. The mean square error of the regression was 0.05. The uncertainty of the predictions is represented by the RMSD, which is 0.239 (in the same units as the predicted parameter).

There are five independent variables. The mean values and ranges of each independent variable in the intial dataset are shown below:

	92^{nd} percentile chlorophyll *a* (mg/m²)	**Conductivity (µS/cm)**	**DIN (mg/m³)**	**DRP (mg/m³)**	**Water temperature (°C)**	**Da_EFF (days)***
Mean	71.3	133.2	391.5	12.3	12.8	57.4
Minimum	2.7	51.6	10.8	6.2	8.6	12.6
Maximum	272.8	328.0	1085.8	51.8	15.1	280.3

- Note that an outlying value of 1124 days at moawhango_waiouru was excluded from this mean and range, but was included in the dataset.

Back-transformed predictions of the 92^{nd} percentile of chlorophyll *a* are shown. Because the predicted variable is log-transformed the errors around the mean are asymmetrical when back-transformed. Therefore, a correction was applied to the power to which the log-transformed prediction was raised, calculated as 0.5 x mean squared error of the regression (Dambolena et al. 2012). This has the effect of increasing the back-transformed value slightly compared to not applying the correction.

Predictions of the 92^{nd} percentile of chlorophyll *a* are shown for scenarios covering three levels of each of DRP, conductivity, temperature and Da_EFF, within the range of the original data (see table above). The range of DIN is set in 19 steps from 20 to 1000 mg/m³.

The predictions are colour-coded to indicate which of the bands of the NPS-FM periphyton attribute are met under each scenario (on average). The thresholds are: Band A ≤ 50 mg/m² (blue cells), Band B > 50 ≤ 120 mg/m² (green cells), Band C > 120 ≤ 200 mg/m² (amber cells), Band D > 200 mg/m² (red cells).

Some combinations of variables did not occur in the dataset used to derive the relationship. For example, sites with DIN >500 mg/m³ never had conductivity < 77 µS/cm. Therefore cells have been left unshaded.

	DRP	cond	DaEFF	Temp	Gradient of DIN (mg/m^3)																		
run	mg/m^3	µS/cm	days	C	20	30	40	50	75	100	125	150	200	250	300	350	400	450	500	600	700	900	1000
1	6	70	15	10	5	5	6	6	7	8	8	9	10	11	11	12	12	13	13	14	15	16	17
2	6	70	15	12	6	6	7	8	9	9	10	11	12	13	13	14	15	15	16	17	18	19	20
3	6	70	15	14	7	8	8	9	10	11	12	13	14	15	16	17	18	18	19	20	21	23	24
4	6	70	40	10	8	9	10	10	12	13	14	15	16	18	19	20	21	21	22	23	25	27	28
5	6	70	40	12	9	11	12	13	14	16	17	18	20	21	23	24	25	26	27	28	30	32	33
6	6	70	40	14	11	13	14	15	17	19	20	22	24	26	27	28	30	31	32	34	36	39	40
7	6	70	75	10	11	12	14	15	17	18	20	21	23	25	26	27	29	30	31	33	34	37	38
8	6	70	75	12	13	15	16	18	20	22	24	25	27	30	31	33	34	36	37	39	41	45	46
9	6	70	75	14	16	18	20	21	24	26	28	30	33	35	38	40	41	43	44	47	49	54	56
10	6	120	15	10	9	10	11	12	13	15	16	17	18	20	21	22	23	24	25	26	28	30	31
11	6	120	15	12	11	12	13	14	16	18	19	20	22	24	25	27	28	29	30	32	33	36	37
12	6	120	15	14	13	14	16	17	19	21	23	24	27	29	30	32	33	35	36	38	40	43	45
13	6	120	40	10	15	17	18	20	22	25	27	28	31	33	35	37	39	40	41	44	46	50	52
14	6	120	40	12	18	20	22	24	27	30	32	34	37	40	42	44	46	48	50	53	56	60	62
15	6	120	40	14	21	24	27	29	33	36	38	41	45	48	51	53	56	58	60	64	67	73	75
16	6	120	75	10	20	23	26	27	31	34	37	39	43	46	49	51	54	56	58	61	64	70	72
17	6	120	75	12	25	28	31	33	38	41	44	47	52	55	59	62	65	67	69	74	77	84	87
18	6	120	75	14	30	34	37	40	45	50	53	57	62	67	71	74	78	81	83	89	93	101	104
19	6	180	15	10	16	18	20	21	25	27	29	31	34	36	38	40	42	44	45	48	50	55	57
20	6	180	15	12	19	22	24	26	29	32	35	37	40	43	46	48	51	53	54	58	61	66	68
21	6	180	15	14	23	26	29	31	35	39	42	44	49	52	55	58	61	63	65	69	73	79	82
22	6	180	40	10	27	30	33	36	41	45	48	51	56	60	64	67	70	73	76	80	84	91	95
23	6	180	40	12	32	37	40	43	49	54	58	62	68	73	77	81	85	88	91	96	101	110	114
24	6	180	40	14	39	44	48	52	59	65	70	74	81	87	93	98	102	106	109	116	122	132	137
25	6	180	75	10	37	42	47	50	57	63	67	71	78	84	89	94	98	102	105	112	117	127	132
26	6	180	75	12	45	51	56	60	69	75	81	86	94	101	107	113	118	122	126	134	141	153	158
27	6	180	75	14	54	61	67	72	82	90	97	103	113	122	129	136	142	147	152	161	170	184	190

	DRP	cond	DaEFF	Temp	Gradient of DIN (mg/m³)																		
run	mg/m³	µS/cm	days	C	20	30	40	50	75	100	125	150	200	250	300	350	400	450	500	600	700	900	1000
1	15	70	15	10	6	7	8	9	10	11	11	12	13	14	15	16	17	17	18	19	20	22	22
2	15	70	15	12	8	9	10	10	12	13	14	15	16	17	18	19	20	21	22	23	24	26	27
3	15	70	15	14	9	10	11	12	14	15	17	18	19	21	22	23	24	25	26	28	29	31	32
4	15	70	40	10	11	12	13	14	16	18	19	20	22	24	25	27	28	29	30	32	33	36	37
5	15	70	40	12	13	15	16	17	20	21	23	24	27	29	31	32	34	35	36	38	40	44	45
6	15	70	40	14	15	17	19	21	23	26	28	29	32	35	37	39	40	42	43	46	48	52	54
7	15	70	75	10	15	17	18	20	23	25	27	28	31	33	35	37	39	40	42	44	46	50	52
8	15	70	75	12	18	20	22	24	27	30	32	34	37	40	42	45	47	48	50	53	56	61	63
9	15	70	75	14	21	24	27	29	33	36	39	41	45	48	51	54	56	58	60	64	67	73	75
10	15	120	15	10	12	14	15	16	18	20	22	23	25	27	29	30	31	33	34	36	38	41	42
11	15	120	15	12	14	16	18	19	22	24	26	27	30	32	34	36	38	39	41	43	45	49	51
12	15	120	15	14	17	20	22	23	26	29	31	33	36	39	41	43	45	47	49	52	54	59	61
13	15	120	40	10	20	23	25	27	31	33	36	38	42	45	48	50	52	54	56	60	63	68	70
14	15	120	40	12	24	27	30	32	37	40	43	46	50	54	57	60	63	65	68	72	76	82	85
15	15	120	40	14	29	33	36	39	44	48	52	55	61	65	69	73	76	79	82	86	91	99	102
16	15	120	75	10	28	32	35	37	42	47	50	53	58	63	66	70	73	76	78	83	87	95	98
17	15	120	75	12	33	38	42	45	51	56	60	64	70	75	80	84	88	91	94	100	105	114	118
18	15	120	75	14	40	46	50	54	61	67	72	77	84	91	96	101	105	110	113	120	126	137	142
19	15	180	15	10	22	25	27	29	33	37	39	42	46	49	52	55	57	59	61	65	68	74	77
20	15	180	15	12	26	30	33	35	40	44	47	50	55	59	63	66	69	71	74	78	82	89	92
21	15	180	15	14	31	36	39	42	48	53	57	60	66	71	75	79	83	86	89	94	99	107	111
22	15	180	40	10	36	41	45	49	56	61	66	70	76	82	87	92	96	99	103	109	114	124	128
23	15	180	40	12	44	50	55	59	67	73	79	84	92	99	105	110	115	119	124	131	138	149	154
24	15	180	40	14	53	60	66	71	81	88	95	101	111	119	126	132	138	144	149	158	166	180	186
25	15	180	75	10	50	58	63	68	77	85	91	97	106	114	121	127	133	138	143	151	159	173	179
26	15	180	75	12	61	69	76	82	93	102	110	116	128	137	146	153	160	166	172	182	191	208	215
27	15	180	75	14	73	83	91	98	112	123	132	140	154	165	175	184	192	200	207	219	230	250	258

	DRP	cond	DaEFF	Temp	Gradient of DIN (mg/m^3)																		
run	mg/m^3	µS/cm	days	C	20	30	40	50	75	100	125	150	200	250	300	350	400	450	500	600	700	900	1000
1	25	70	15	10	8	9	9	10	12	13	14	14	16	17	18	19	20	21	21	23	24	26	27
2	25	70	15	12	9	10	11	12	14	15	16	17	19	20	22	23	24	25	26	27	28	31	32
3	25	70	15	14	11	12	14	15	17	18	20	21	23	25	26	27	29	30	31	33	34	37	38
4	25	70	40	10	13	14	16	17	19	21	23	24	26	28	30	32	33	34	36	38	40	43	44
5	25	70	40	12	15	17	19	20	23	25	27	29	32	34	36	38	40	41	43	45	48	52	53
6	25	70	40	14	18	21	23	24	28	31	33	35	38	41	44	46	48	50	51	55	57	62	64
7	25	70	75	10	17	20	22	23	27	29	32	33	37	39	42	44	46	48	49	52	55	60	62
8	25	70	75	12	21	24	26	28	32	35	38	40	44	48	50	53	55	57	59	63	66	72	74
9	25	70	75	14	25	29	32	34	39	43	46	48	53	57	61	64	67	69	71	76	80	86	89
10	25	120	15	10	14	16	18	19	22	24	26	27	30	32	34	36	37	39	40	42	45	48	50
11	25	120	15	12	17	19	21	23	26	29	31	33	36	38	41	43	45	46	48	51	54	58	60
12	25	120	15	14	20	23	26	27	31	34	37	39	43	46	49	52	54	56	58	61	64	70	72
13	25	120	40	10	24	27	30	32	36	40	43	45	50	53	57	60	62	65	67	71	74	81	84
14	25	120	40	12	28	32	36	38	44	48	51	54	60	64	68	72	75	78	80	85	90	97	101
15	25	120	40	14	34	39	43	46	52	57	62	66	72	77	82	86	90	93	97	103	108	117	121
16	25	120	75	10	33	37	41	44	50	55	59	63	69	74	79	83	86	90	93	99	104	112	116
17	25	120	75	12	40	45	49	53	61	66	71	76	83	89	95	100	104	108	112	118	125	135	140
18	25	120	75	14	48	54	59	64	73	80	86	91	100	107	114	120	125	130	134	143	150	162	168
19	25	180	15	10	26	29	32	35	39	43	47	49	54	58	62	65	68	70	73	77	81	88	91
20	25	180	15	12	31	35	39	42	47	52	56	59	65	70	74	78	81	85	88	93	98	106	110
21	25	180	15	14	37	42	47	50	57	63	67	71	78	84	89	94	98	102	105	112	117	127	132
22	25	180	40	10	43	49	54	58	66	72	78	83	91	97	103	109	113	118	122	129	136	147	152
23	25	180	40	12	52	59	65	70	79	87	94	99	109	117	124	131	136	142	146	155	163	177	183
24	25	180	40	14	62	71	78	84	95	105	113	119	131	141	149	157	164	170	176	187	196	213	220
25	25	180	75	10	60	68	75	80	92	101	108	115	126	135	144	151	158	164	169	180	189	205	212
26	25	180	75	12	72	82	90	97	110	121	130	138	151	163	173	181	189	197	204	216	227	246	255
27	25	180	75	14	87	99	108	116	133	146	157	166	182	196	208	218	228	237	245	260	273	296	306

horizons
regional council

11-15 Victoria Avenue
Private Bag 11 025
Manawatu Mail Centre
Palmerston North 4442

T 0508 800 800
F 06 952 2929
help@horizons.govt.nz
www.horizons.govt.nz