NIWA
Taihoro Nukurangi

Inter-laboratory comparison of analysis of chlorophyll *a* from periphyton samples

Prepared for Envirolink (Tasman District Council)

November 2020

Climate, Freshwater & Ocean Science

Prepared by:
Cathy Kilroy
Oonagh Daly

For any information regarding this report please contact:
Cathy Kilroy
Freshwater Ecologist
Freshwater Ecology
+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: 2020321CH
Report date: November 2020
NIWA Project: ELF21503

Revision	Description	Date
Version 1.0	Final report	26 January 2021

Quality Assurance Statement		
	Reviewed by:	Juliet Milne
	Formatting checked by:	Rachel Wright
	Approved for release by:	Helen Rouse

Contents

Tables

Figures

Executive summary

A draft National Environmental Monitoring Standard (NEMS) for periphyton monitoring in wadable rivers and streams has been prepared and seeks to improve national consistency in State of the Environment monitoring as well as monitoring under the National Policy Statement for Freshwater Management (NPS-FM) and resource consents.

The NEMS *Periphyton* includes methods for the collection and processing of samples for biomass analysis, including chlorophyll *a*, which is the primary measure used to represent periphyton biomass in streams for comparison with the NPS-FM and other standards. The laboratory analysis method specified in the NEMS *Periphyton* (for a QC 600 rating) is a method described in earlier guidelines, which used 90% ethanol as an extractant.

Development of the NEMS *Periphyton* identified variation across regional council laboratory providers in both the method used for quantifying chlorophyll *a* in samples and in how the method prescribed in the NEMS is followed. The most significant of these variations was use of different chlorophyll *a* extraction solvents.

To establish the effect of using different extraction solvents and other potential variations in procedures, and implications for environmental reporting, the NEMS Steering Group, via Tasman District Council, obtained an MBIE Envirolink Small Advice Grant for NIWA to coordinate and report on an interlaboratory comparison across six of the main providers of chlorophyll *a* analyses in New Zealand. The comparison addressed the stages of chlorophyll *a* analysis following subsampling of field samples in the laboratory (i.e., pigment extraction, and instrument reading of absorbances with an acidification step).

Triplicate sets of pre-prepared filters (subsamples) from 19 periphyton samples (covering a wide range of chlorophyll *a* concentration) were shipped to the six laboratories on the same day and analysed over the same short time frame. Results for each filter were reported as mg/sample.

Laboratory names were anonymised (I to VI) for reporting. Laboratories I to IV used 90% ethanol as the extractant. Laboratories V and VI used acetone as the extractant.

The main results were:

- Chlorophyll *a* estimates (as means of the triplicate subsamples) from each of the six laboratories were closely correlated ($R > 0.97$) with one another.
- With the exception of laboratory pairs I vs II and I vs IV, there was evidence for some bias in measurement towards higher or lower values in the different laboratories. The order from lowest to highest values (on average) was laboratory VI, IV, I, II, III and V.
- Percentage differences from the overall mean value (all laboratories) ranged from -31% to 35%.
- Laboratories I to IV (90% ethanol extraction) showed modest discrepancies among their results (mean difference of 7.8 ± 3.5% between laboratories).
- Results from laboratories V and VI (acetone extraction) showed larger discrepancies with those from laboratories I to IV (mean difference of 16.6 ± 6.6% between laboratories) and also diverged from each other (mean difference of 25%).

- Despite the differences, individual results from all six laboratories explained at least 97% of the variance in the overall mean chlorophyll *a* (across all six laboratories) (linear regressions). This suggests that the overall mean value may be predictable in all laboratories using laboratory-specific relationships.

Recommendations from the comparison were:

- In view of both the inconsistency in results between laboratories V and VI (using acetone), and the generally larger discrepancies between these laboratories and the laboratory I to IV results (using 90% ethanol), the QC 600 specification for ethanol extraction in the NEMS *Periphyton* should be retained.
- Chlorophyll *a* measurements performed by laboratories using acetone should continue to be eligible for a maximum quality code rating of QC 500.
- Of the laboratories using ethanol as an extractant, results from laboratories III and IV deviated farthest from the mean and reviews of method details in those laboratories may be useful.
- A separate comparison of method details in laboratories V and VI is required to identify why the results differed so markedly.

Regarding achieving consistency of chlorophyll *a* results over time and across laboratories, it is suggested that:

- If a council wished to retain its current laboratory chlorophyll *a* extraction method (for consistency over time), then checks on the potential effect on use of the data (e.g., for comparison with NPS-FM band thresholds) could be made by applying the appropriate laboratory-specific equation (derived from the test data) to obtain an estimate of the laboratory average value.
- Laboratories using acetone extraction may opt to change their extraction method to align with NEMS QC 600 requirements, and the usual practice of carrying out parallel analyses using both extraction methods should be applied.

1 Introduction

A new National Environmental Monitoring Standard (NEMS) has been drafted for periphyton monitoring in wadable rivers and streams. The NEMS *Periphyton* (in prep.) seeks to improve national consistency in State of the Environment monitoring and monitoring under the National Policy Statement for Freshwater Management (NPS-FM) 2020 (NZ Government 2020) or resource consents.

The NEMS *Periphyton* includes methods for the collection and processing of samples for biomass analysis, including chlorophyll *a*. Chlorophyll *a* is an internationally accepted measure of periphyton biomass because all algae in periphyton, including cyanobacteria, contain this pigment. Chlorophyll *a* is specified in the periphyton attribute of the NPS-FM 2020 to represent periphyton biomass as an indicator of ecosystem health. Objectives or outcomes for periphyton are also specified in terms of chlorophyll *a* in some regional plans (e.g., Horizons Regional Council, Environment Canterbury).

Determination of chlorophyll *a* in periphyton requires collection of quantitative samples (i.e., from a known area of river bed) from appropriate river sites, appropriate storage of those samples to maintain the integrity of the chlorophyll *a*, and processing of subsamples in a laboratory to obtain an estimate of chlorophyll *a* in the whole sample. These estimates are converted to a quantity per unit area of riverbed for comparison with standards.

The draft NEMS *Periphyton* currently specifies use of the laboratory method described in Biggs and Kilroy (2000) for data to be eligible for the highest quality rating (QC 600). However, development of the NEMS *Periphyton* identified variation across regional council laboratory providers in both the method used for quantifying chlorophyll *a* in samples and in how the method prescribed in the NEMS is followed. The most significant of these variations is use of different chlorophyll *a* extraction solvents from that recommended (90% ethanol).

Preliminary work carried out in-house by NIWA has suggested that method differences could lead to different results, with implications for regional councils (e.g., in terms of maintaining a consistent record through time when laboratory providers change, or in reporting against NPS-FM objectives). It is also possible that the currently prescribed QC 600 test method is overly restrictive, potentially precluding the use of acetone extraction solvents use by some laboratory providers.

To address these issues, NIWA received an Envirolink Small Advice Grant via Tasman District Council, on behalf of the NEMS Steering Group, to coordinate and report on an interlaboratory comparison across six of the main providers of chlorophyll *a* analyses in New Zealand. The comparison specifically addressed the stages of chlorophyll *a* analysis following subsampling of field-collected periphyton samples in the laboratory. The steps are:

A. extracting chlorophyll *a* from a subsample concentrated onto a filter, and

B. reading chlorophyll *a* concentrations in the extractant (e.g., on a spectrophotometer), including an acidification step to exclude phaeophytins[1] from the results.

In this report we describe the procedure followed in the comparison and set out the results. Recommendations are made on the method to assist with finalisation of the NEMS *Periphyton*. Suggestions are also made on potential actions for improving consistency in chlorophyll *a* data across New Zealand.

[1] Phaeophytins are degradation products of pigments, which have a spectral peak very close to that of chlorophyll *a*. Not accounting for phaeopigments leads to overestimates of chlorophyll *a* concentrations in samples.

2 Methods

2.1 Participating laboratories

Six laboratories participated in the comparison: Bay of Plenty Regional Council, Cawthron, Hill Laboratories, NIWA Christchurch, NIWA Hamilton, and Ryder Environmental. In this report, all the results have been anonymised, and laboratories are hereafter referred to by randomised Roman numerals I to VI. Laboratories I to IV followed the chlorophyll *a* extraction and measurement method specified in the current draft NEMS *Periphyton* (i.e., the method described in Biggs and Kilroy 2000), which uses 90% ethanol as the extractant. Laboratories V and VI followed the chlorophyll *a* determination method described in APHA (2017) (with modifications), which uses acetone as the extractant.

2.2 Sample selection and preparation

All preparation for the comparison was carried out at the NIWA Laboratory, Christchurch. Nineteen samples were selected from existing stored or recently collected material, aiming to cover the range of chlorophyll *a* readable on a spectrophotometer without requiring dilution of the extractant (i.e., up to an absorbance reading of about 1 on the unacidified sample, when using 90% ethanol).

Samples were homogenised and filtered following the procedure set out in the draft NEMS *Periphyton*. Filtration was carried out using a filtration manifold with a vacuum suction pump and subsamples were filtered onto 47 mm diameter GFC filters. Three 5 mL aliquots were pipetted onto each of 18 replicate filters per sample, providing three filters (subsamples) per laboratory. Multiple aliquots and multiple filters were required to try to minimise variability arising from the subsampling process.

After completion of filtering, each filter and its associated periphyton biomass was removed from its filter holder, folded in two (sample inwards), and placed individually into a Secol™ sleeve labelled with the sample code, then “a” to “r”, indicating the filtering sequence). All subsamples were immediately placed in a chilly bin cooled with ice and transferred to a freezer (-80°C) within 1 hour of filtering.

Triplicate sets of filters were assigned to laboratories randomly, with stratification to ensure that each set comprised samples well separated in the filtration sequence.

2.3 Procedure for sample dispatch and analysis

Filtration was carried out over 21 – 23 September 2020, and all prepared filters stored at -80 °C. With agreement from all the laboratories, their assigned sets of subsamples were packed on ice into chilly bins and shipped overnight on 28 September. Laboratories were instructed to place the subsamples in a freezer (-20 °C) immediately, and to process them on 1 – 2 October. The delay between receipt and processing was included to allow for the possibility that unexpected events might prevent commercial laboratories from processing the subsamples at the planned time. If processing was carried out later than 1 – 2 October, we reasoned that a few more days at -20 °C would have less effect on the subsamples than if some laboratories re-froze the subsamples and others did not.

2.4 Data reporting and analysis

2.4.1 Data units

Laboratories reported the results as either chlorophyll *a* per filter, or chlorophyll *a* per total sample (using the initial total and subsample volumes provided for each sample). All data were initially converted to milligrams of chlorophyll *a* per sample (range 0.2 to 2.4 mg/sample).

The usual unit for chlorophyll *a* is milligrams per square metre of riverbed. For data analysis (see Section 2.4.3 below) we therefore converted the per sample values to per m^2 by assuming that each sample was scraped from an area of 0.007 m^2 (or 10 individual rock scrapings from an area 3 cm in diameter). Note that this calculation was done simply to re-scale the data to a more familiar range (2.9 – 342 mg/m^2) to enable a comparison of reporting against standards.

2.4.2 Data screening

After establishing that the data generally showed congruence across laboratories, the following screening procedures were carried out.

1. The data were checked for outlier values that could have arisen during subsampling and preparation of the filters. Outliers were identified from box plots of all individual replicates from all laboratories (n = 18) as any value that fell outside 3 times the interquartile range of all data for that sample. Note that we expected variability in the dataset and any data removed from the dataset needed to be justified by an extreme difference from other replicates in a sample.

2. Variability within the three replicate subsamples used by each laboratory was calculated as the coefficient of variation as a percentage (%CV) to check for consistency (or not) across laboratories.

3. The raw data were checked against the replicate number to assess whether there was any sign of a trend in the subsamples due to order of subsampling (i.e., increasing or decreasing concentrations as subsampling proceeded). Laboratories received random replicate subsamples in each sample, to ensure that any such bias was averaged out. Two laboratories received six replicate subsamples of one of the samples instead of three (in error), and these series of subsamples also allowed an additional check on subsampling consistency.[2]

2.4.3 Data analysis

Once the screening procedure was completed, the chlorophyll *a* estimates obtained by each laboratory for each of the three replicate subsamples were averaged to obtain a mean value of chlorophyll *a* per sample, converted to mg/m^2 as noted above. We then compared chlorophyll *a* across laboratories using a correlation matrix (Spearman rank correlations, as the data were not normally distributed). Chlorophyll *a* concentration in each of the 19 (or 18) samples from the six laboratories was plotted against that from the other five (i.e., 15 plots), to visualise the relationships between laboratories. We used a non-parametric Wilcoxon Signed-Rank Test to determine the relative statistical significances of differences between results from each pair of laboratories.

[2] Laboratory IV received six replicates of sample AM132 instead of three; laboratory V did not receive any replicates of AM125; laboratory VI received six replicates of AM125 and none of AM132. This meant that laboratories V and VI processed 18 samples and laboratories I to IV processed 19 samples.

Absolute and percentage differences between all pairs of laboratories were calculated and plotted/tabulated to visualise patterns of differences between pairs of laboratories.

We calculated the percentage deviation of each chlorophyll *a* value in each laboratory from the overall mean for each sample (i.e., calculated from all 18 replicate subsamples) and compared these deviations across the six laboratories. Finally, we fitted linear regression relationships between the results from each laboratory and the overall mean value.

3 Results

3.1 Data screening

The complete dataset is presented in Appendix A. The screening procedure identified 10 data points (out of 342 subsamples (filters)) that met the criterion for defining outliers (Figure 3-1). Subsequent data analysis highlighted that: (a) differences between laboratories also accounted for much of the variability, and (b) provision of three filters per sample and use of the mean value evened out variability (as intended). Therefore, it was difficult to justify removal of selected datapoints and all the data were retained for data analysis.

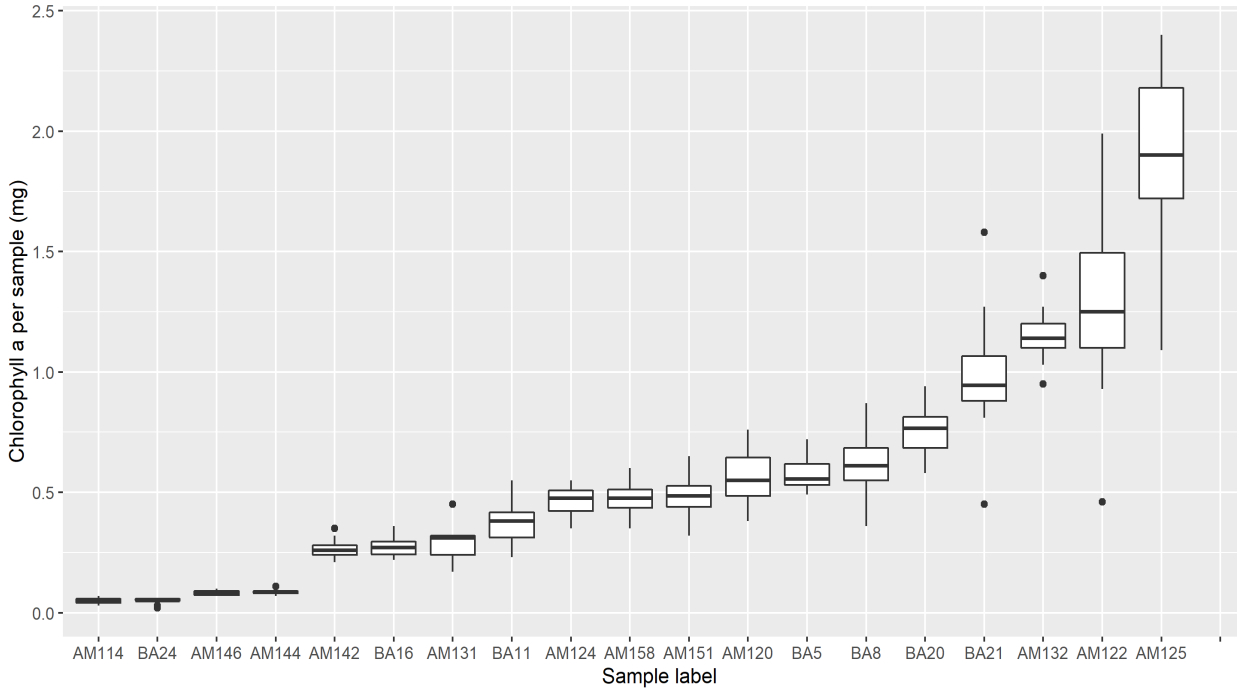

Figure 3-1: Box plot of data from individual filters (*n* = 18) from each of the 19 samples. Data from all laboratories combined, with samples arranged in order of median chlorophyll *a*. Outliers within each sample are shown by black dots (>3 x outside the interquartile range shown by the centre box).

%CV varied across sets of samples but was generally consistent across laboratories I to V (average of 10.6 to 12.9%, Table 3-1). Laboratory VI received samples that were more variable (average of 19.4%). Even excluding the extremely variable AM122 sample set, mean variability remained high (16.6%).

The data did not suggest any trends in biomass linked to order of filtration.

Table 3-1: Coefficients of variation (%) calculated for each set of triplicate subsamples (filters) sent to each laboratory. Samples arranged in order of increasing mean chlorophyll *a*. NA = no data, laboratory did not receive filters for this sample.

Sample	Coefficients of variation of sample replicates (%CV) by laboratory						Mean %CV
	I	II	III	IV	V	VI	
BA24	12.4	0.0	0.0	32.0	9.2	36.0	14.9
AM114	12.4	15.6	10.0	24.5	19.3	9.9	15.3
AM146	6.7	11.5	9.4	6.7	16.5	6.2	9.5
AM144	19.2	14.5	7.1	4.2	3.5	10.3	9.8
AM142	8.2	7.8	16.4	9.8	18.0	13.8	12.3
BA16	10.6	15.9	19.3	6.9	7.1	5.1	10.8
AM131	21.7	0.9	14.6	12.8	14.3	14.2	13.1
BA11	28.8	18.3	7.9	13.7	14.9	12.4	16.0
AM124	11.9	3.9	11.4	21.3	6.7	14.0	11.5
AM158	20.8	9.5	16.9	18.8	7.0	16.7	15.0
AM151	9.5	7.4	17.6	23.8	12.4	33.3	17.3
AM120	5.6	14.4	7.7	14.3	8.0	10.8	10.1
BA5	6.0	6.9	6.7	2.9	5.1	4.5	5.4
BA8	5.4	9.2	4.0	8.0	3.4	21.6	8.6
BA20	3.3	9.8	3.6	10.4	10.9	15.7	9.0
BA21	5.9	18.0	12.8	8.5	19.8	33.3	16.4
AM132	8.7	9.2	2.8	0.0	7.7	NA	5.7
AM122	10.0	29.3	29.7	9.9	21.2	64.2	27.4
AM125	2.5	4.9	2.9	15.8	NA	25.0	10.2
Mean %CV	**11.0**	**10.9**	**10.6**	**12.9**	**11.4**	**19.3**	

3.2 Correlation and correspondence between laboratories

Chlorophyll *a* was strongly correlated across all pairs of laboratories, with a minimum Spearman R of 0.97 (Table 3-2). The strong correlations were reflected by closely fitting linear model lines on scatterplots of data from all pairs of laboratories (Figure 3-2). However, divergence of the model lines from 1 : 1 lines was also evident on some plots, especially those including laboratories V and VI.

Table 3-2: Spearman rank correlation matrix showing correlations between chlorophyll *a* measured on equivalent samples at six laboratories (I to VI). Correlations were calculated between the means of three subsamples (filters) for each sample.

	I	II	III	IV	V
II	0.992	1			
III	0.988	0.983	1		
IV	0.996	0.996	0.985	1	
V	0.988	0.991	0.992	0.991	1
VI	0.971	0.977	0.971	0.972	0.97

All laboratories showed differences (as indicated by low P-values in a Wilcoxon Signed-Rank test) from at least two other laboratories (orange and green cells in Table 3-3). Only two laboratory pairs (I vs. II and 1 vs IV) returned P-values indicating a non-significant difference (P > 0.05).

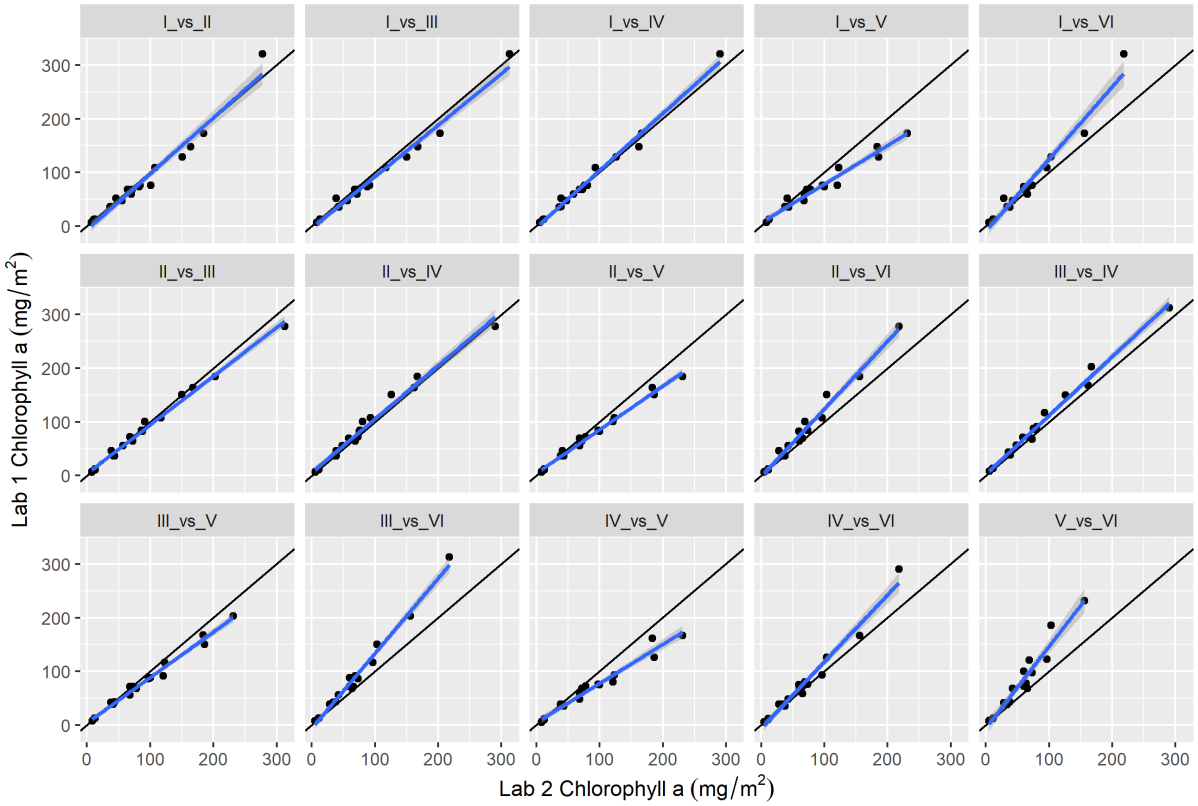

Figure 3-2: Scatter plots of chlorophyll *a* (scaled to mg/m^2) measured by all six laboratories (I to VI) against that by all other laboratories. *n* = 19 (I to IV) or 18 (V, VI). Chlorophyll *a* is the mean of three replicates. Lab 1 and Lab 2 are, respectively, the first and second laboratory in the label at the top of each plot. The blue lines are the fitted linear model for each relationship. The shaded areas (barely visible on most plots) show the 95% confidence interval on the fitted values (the line). The black lines show the 1:1 relationship that reflects complete agreement between two laboratories.

Table 3-3: P-values from a Wilcoxon Signed-Rank test between all pairs of laboratories (I to VI) on chlorophyll *a* measured on equivalent samples. The comparisons were made between the means of three subsamples (filters) for each sample Uncorrected P-values are shown, shaded to illustrate the strength of difference from orange (strongest differences, lowest P) through green, blue and lilac to grey (least difference, highest P). Orange and green cells indicate significant bias in measurements between laboratories.

	I	II	III	IV	V
II	0.099	1			
III	0.005	0.030	1		
IV	0.379	0.016	0.001	1	
V	0.002	0.002	0.012	<0.001	1
VI	0.002	<0.001	<0.001	0.010	<0.001

3.3 Magnitude of differences between laboratories

The magnitude of differences between laboratories shown in Figure 3-2 is seen more clearly in Figure 3-3, in which the *y*-axis is the difference between laboratories. There was generally close overall correspondence between laboratories I, II, III and IV, although laboratory III results tended to be higher than those from laboratories II and IV, especially at higher concentrations. Steeper slopes for the relationships between laboratories I vs IV and V vs VI (either above or below the y = 0 line) indicated increasing discrepancies in concentrations as chlorophyll *a* increased.

The largest discrepancies were seen between laboratories V and VI, as indicated by the steep slope of the fitted line on the bottom right-hand plot in Figure 3-3. The highest concentrations measured in laboratory V (around 200 mg/m^2) were 75 – 80 mg/m^2 higher than those measured in laboratory VI on the same samples. The significance of these differences is discussed in Section 4.4.

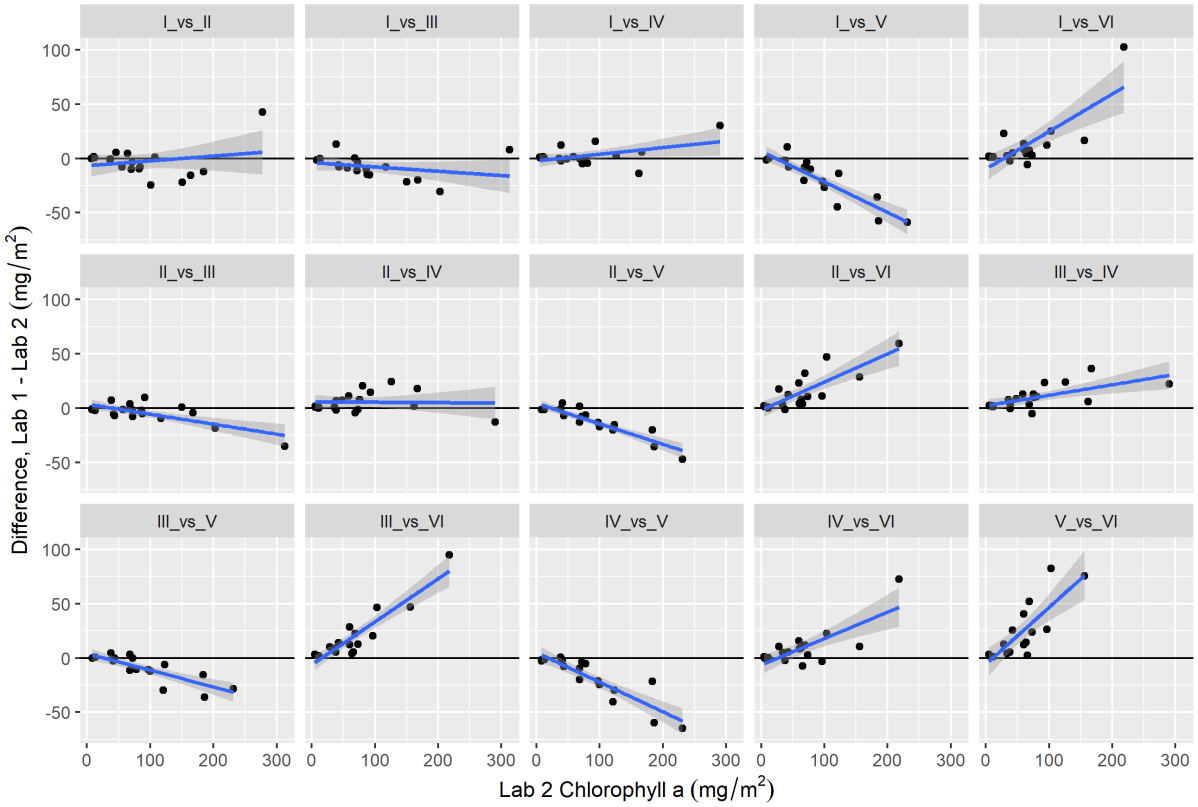

Figure 3-3: Differences in chlorophyll *a* (scaled to mg/m^2) between pairs of laboratories (I to VI) plotted against chlorophyll *a* in the second laboratory. Chlorophyll *a* is the mean of three replicates for each sample. Lab 1 and Lab 2 are, respectively, the first and second laboratory in the label at the top of each plot. The blue lines are the fitted linear model for each relationship. The shaded areas show the 95% confidence interval on the fitted values (the blue line). The black lines are plotted at y = 0.

Mean percentage differences (ignoring positive and negative signs) between each pair of laboratories ranged from 4% (I vs IV) to almost 25% (V vs VI) (Table 3-4). Excluding laboratories V and VI, the overall average difference between laboratories was 7.8 ± 3.5%. Comparisons with laboratories V and VI included had a mean percentage difference of 16.6 ± 6.6%. In both laboratory groups (I to IV and V & VI), percentage differences were generally consistent across the range of chlorophyll *a* concentration.

Table 3-4: Mean percentage difference in chlorophyll *a* between all pairs of laboratories. Differences are the laboratory on the horizonal heading minus the laboratory on the vertical heading. A negative sign means that chlorophyll *a* measured in the first laboratory was smaller than in the second. The shaded panel highlights comparisons between laboratories I to IV (90% ethanol) only.

Laboratory	I	II	III	IV	V
II	-4.8	1			
III	-10.9	-6.2	1		
IV	4.0	7.9	12.9	1	
V	-20.1	-13.0	-7.4	-24.8	1
VI	12.9	16.3	21.3	8.5	25.0

An alternative comparison is to look at percentage differences from the overall mean of chlorophyll *a*, calculated from the combined results from all six laboratories. Laboratory II was closest to the overall mean with only two samples deviating by more than 10% (outliers on Figure 3-4). Results from laboratory II were on average <2% above the mean, followed by laboratories I, IV, III, V and VI (Figure 3-4, Table 3-5). Results from laboratories V and VI were more variable and divergent than laboratories I to IV, with 56 (laboratory V) and 50% (laboratory VI) of samples within 15% of the overall mean, compared to 84 to 100% of samples for laboratories I to IV (Table 3-5).

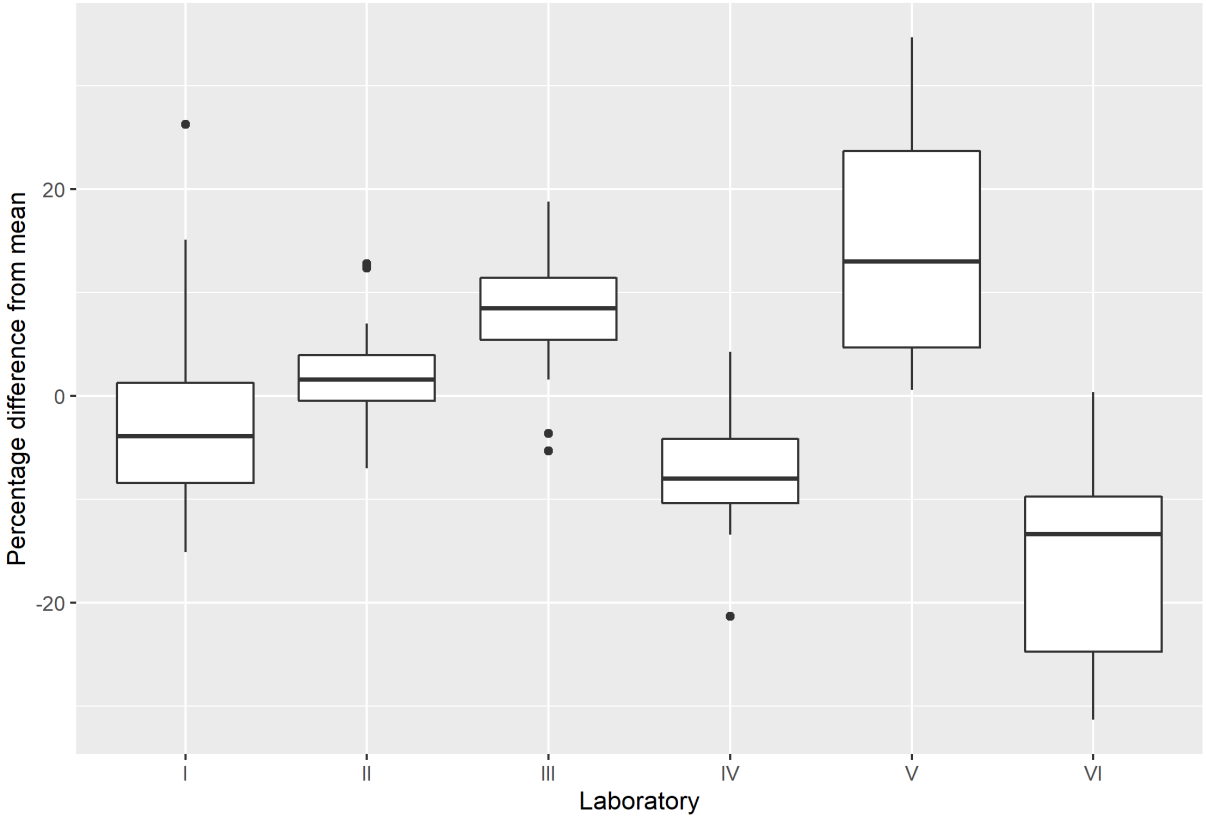

Figure 3-4: Summary of percentage differences from the overall mean of chlorophyll *a* in the results reported from the six laboratories (I to VI). *n* = 19 (I to IV) or 18 (V, VI). Outliers (black dots) are defined as in Figure 3-1.

Table 3-5: Basic statistics for each laboratory for percentage differences from the overall mean for each sample. The bottom row shows the percentage of samples analysed by each laboratory with a result within 15% of the overall mean. The 15% 'threshold' is somewhat arbitrary but differences of less than 15% would generally be ecologically insignificant (see Discussion, section 4.3).

	Percentage difference from mean chlorophyll *a*					
	I	**II**	**III**	**IV**	**V**	**VI**
Mean	-2.1	1.8	7.7	-6.7	15.1	-15.7
Standard deviation	9.8	5.3	6.1	6.4	11.0	9.9
Minimum	-15.1	-7.0	-5.3	-21.3	0.6	-31.3
Maximum	26.3	12.8	18.8	4.3	34.7	0.4
% Samples within 15% of mean	**84**	**100**	**89**	**95**	**56**	**50**

A plot of the raw data used to generate Figure 3-4 (Figure 3-5) allows a review of the consistency of results across samples (i.e., were measurements consistently higher or lower at the same laboratories, possibly with a shift in relative measurements along the gradient of overall mean chlorophyll *a*?).

The only markedly atypical sample was AM131, for which laboratory I (red symbols) returned mean chlorophyll *a* > 25% higher than the overall mean (compared to slightly lower than the mean, on average, i.e., the only red symbol above the 15% line).

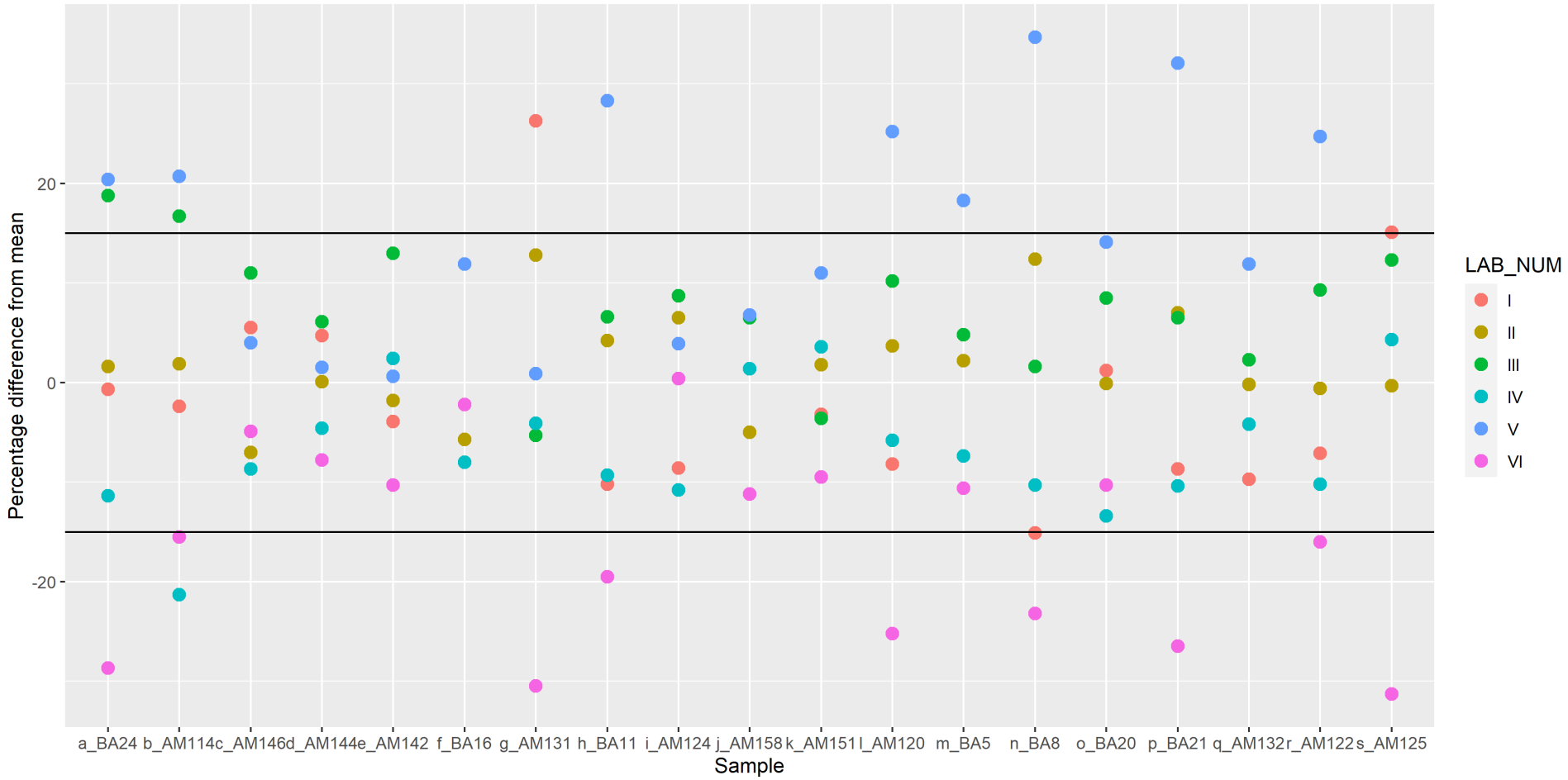

Figure 3-5: Percentage difference from overall mean chlorophyll *a* for each sample analysed by the six laboratories. Samples are arranged in order from lowest to highest overall mean chlorophyll *a*. Seventy-nine percent of chlorophyll *a* estimates were within ± 15% of the overall mean (shown by the black lines). Some samples show data from fewer than six laboratories. This was due to either (a) two laboratories not receiving triplicates of AM125 (laboratory V) and AM132 (laboratory VI), or (b) some labs returning identical results for the same samples (e.g., for sample BA16, results from laboratories I and IV & III and V were identical; for sample BA5 & AM158, results from laboratories I and IV were identical).

3.4 Aligning laboratory results

The true chlorophyll *a* content of the 19 samples analysed is unknown. Given that 79% of results lay within 15% of the mean and 100% within 35% of the mean (see Figure 3-5), a reasonable assumption is that the true means can be approximated by the overall mean values from the six laboratories. Chlorophyll *a* estimates from each of the six laboratories explained at least 97% of the variance in the overall mean value in linear regressions (Table 3-6).

The regression statistics in Table 3-6 apply to chlorophyll *a* per sample. To apply to chlorophyll *a* per square metre, first scale the *y*-axis intercept value of the regression equation by the area (m^2) from which the samples were collected. The slope of the regression line does not change. If different areas apply to different samples, then the equations should be used to convert the per sample value to the overall mean value prior to conversion to mg/m^2.

Table 3-6: Linear regression relationships between the overall mean value for chlorophyll *a* and the results from each laboratory. The regressions were derived using chlorophyll *a* per sample, (mean value from three replicate subsamples, which were the results reported by the laboratories). SE = standard error. Data were not transformed.

	Regression statistics			Equation for predicting mean chlorophyll *a*	
Laboratory	**R**	**R^2**	**SE, estimate**	**Intercept**	**Slope**
I	0.987	0.975	0.080	0.052	0.926
II	0.999	0.997	0.027	-0.008	0.995
III	0.998	0.997	0.029	0.016	0.906
IV	0.995	0.990	0.051	0.029	1.000
V	0.993	0.986	0.046	0.029	0.796
VI	0.987	0.973	0.081	-0.051	1.353

4 Discussion

4.1 Integrity of samples

All laboratories were able to process the samples within or close to the planned timeframe and we did not expect that any differences in results were attributable to differences in sample handling.

The sets of triplicate samples received by laboratory VI were, on average, more variable in terms of estimated chlorophyll *a* than those received by the other laboratories. There is no obvious explanation and the difference probably arose by chance over the relatively small sample numbers (*n* = 18 for laboratory VI). Five sets of triplicates processed by laboratory VI had %CV ≥ 25 compared to only four sets across other laboratories. Removal of the five samples (AM122, AM125, AM151, BA21, BA24) from the dataset resulted in similar means and ranges of %CV across all laboratories.

4.2 Alternative statistical approaches

The objective of the present comparison was to evaluate results from six different providers of chlorophyll *a* analyses for regional councils in New Zealand. We used conventional statistical techniques to assess differences between laboratories and included a wide range of concentrations to (a) provide more certainty about generalising the results and (b) allow development of relationships between each laboratory and the overall mean value. Some specialised statistical techniques used in inter-laboratory comparisons were not strictly applicable for our objectives, but are briefly mentioned here, for reference.

Formal inter-laboratory comparisons frequently use two samples but a larger number of laboratories (e.g., *n* = 11, Schilling et al. 2006). Results can be evaluated using at least two methods. The first is to calculate Z-scores for each laboratory (e.g., Martin et al. 2017). The Z-score is the difference between the sample and the median value across all laboratories divided by the standard deviation of the error across all laboratories (i.e., the number of standard deviations from the median). Z-scores <2 are interpreted as satisfactory, 2 – 3 questionable, and > 3 unsatisfactory (Martin et al. 2017).

A second method is to plot the samples on a "Youden plot" (Youden 1959)[3], which is a scatterplot of the results of the two samples, with *n* = number of laboratories. Over-plots of the 95% (or 99%) confidence limits (as ellipses) allow identification of outliers and inconsistencies. Typically, Z-scores and Youden plots are used when two samples have similar expected values.

To illustrate the methods, we calculated Z-scores for laboratories I to VI for all pairs of consecutive samples, in order of increasing mean chlorophyll *a*. Complete results are shown Appendix B. All Z-scores were < 2 (range 0 to 1.82). Therefore, using the scale from Martin et al. (2017) the results for all laboratories showed satisfactory correspondence. An example Youden plot generated between samples AM158 and AM151 (median chlorophyll *a* of 69 and 70 mg/m^2, respectively) also highlighted good correspondence across all samples (i.e., all within the 99% confidence ellipse), especially laboratories I to IV, but some systematic error in laboratories V and VI (Figure 4-1).

Additional statistical methods in interlaboratory comparisons require a true value of the sample, either from use of reference materials, or prior analyses by "expert" laboratories (e.g., Schilling et al. 2006, Szewczak and Bonarzewski 2016). In this study we took the mean value across all laboratories as a surrogate for the true mean, which is common practice (Szewczak and Bonarzewski 2016).

[3] See also https://www.statisticshowto.com/youden-plot/ for a clear description of Youden plots and their application.

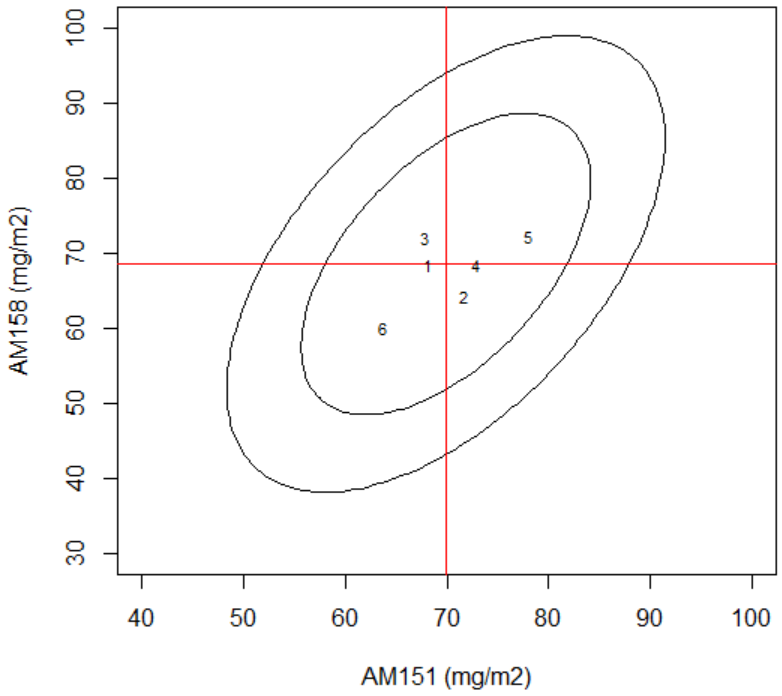

Figure 4-1: Youden plot showing chlorophyll *a* (mg/m^2) in samples AM151 and AM158. The values from laboratories I to IV are clustered around the intersection of the medians of both samples (red lines) and show good correspondence. Values for laboratories 5 and 6 are along a 45-degree line (not shown) and indicate systematic error originating in the laboratories. Nevertheless, all six results lie within the 99% confidence limit (inner ellipse). The outer ellipse shows the 95% confidence limit.

4.3 Effect of analysis method

Laboratories I to IV used the chlorophyll *a* analysis method specified in the draft NEMS *Periphyton* (i.e., extraction using boiling 90% ethanol, followed by reading on a spectrophotometer, with an acidification step, as described by Biggs and Kilroy 2000). The results from laboratories I and IV showed the closest correspondence to one another, followed by laboratories I and II. Laboratories I and II also returned results very close to the overall mean (within ~2% on average). Laboratory III consistently returned results higher than the overall mean and laboratory IV lower, by about the same amount (8% to -7% on average). Chlorophyll *a* measured in laboratory III was, on average, about 13% higher than that measured in laboratory IV.

Thus, there was some bias in chlorophyll *a* estimates across the four laboratories using 90% ethanol extraction, especially a tendency for laboratory III to return higher values than the other three. Nevertheless, discrepancies of less than 15% are relatively low, and represent lower variability than that associated with sample collection (e.g., Kilroy et al. 2013). The choice of threshold for identifying unimportant discrepancies is somewhat arbitrary and to some extent depends on the data. Thus, Kilroy et al. (2013) selected a higher threshold (20%) for acceptable departures from the mean when comparing inter-operator variability in chlorophyll *a* sample collection, on the basis that lower percentage differences were unlikely to be ecologically significant along an open-ended range. In the present comparison, the results from laboratories III and IV deviated farther from the mean than laboratories I and II and reviews of method details in laboratories III and IV may be useful.

Laboratories V and VI used acetone as an extractant and followed the same general method (based on APHA 2017). Some discrepancies between these two laboratories and laboratories I to IV were

anticipated, but the different directions of the discrepancies were unexpected. Chlorophyll *a* measured in laboratory V was, on average, almost 25% higher than that measured in laboratory VI, the largest difference observed across all pairs of laboratories. The discrepancy suggested that there were important differences in some specific aspect(s) of the methods used by these two laboratories.

It is beyond the scope of this report to investigate the reasons for the discrepancy between laboratories V and VI. A separate comparison of method details in laboratories V and VI would be required to identify why the results differed. Acetone has been reported to be a less efficient extractant of chlorophyll *a* than ethanol in several studies (e.g., Sartory and Grobbelaar 1984, Wasmund et al. 2006, Lan et al. 2011), consistent with the result from laboratory VI, compared to laboratories I to IV. In contrast, Webb et al. (1992) reported no difference between results using 90% ethanol and acetone (a 2: 3 mixture of dimethyl sulfoxide (DMSO) and acetone) as extractants, but strongly recommended 90% ethanol as an extractant for both safety and cost reasons.

4.4 Implications for application of chlorophyll *a* results

The comparison highlighted that consistent differences (i.e., bias) exist between some laboratories in chlorophyll *a* concentration returned from the same samples. The laboratory pairs most affected were: V vs VI; V and VI vs other laboratories (except III vs V, IV vs VI); III vs I; and III vs IV (refer to Table 3-4). The implication is that compliance with thresholds specified in the NPS-FM periphyton attribute or with objectives or outcomes specified in regional plans may vary depending on which laboratory processes periphyton samples.

As an illustration, Table 4-1 shows how many of the test samples (means of three replicate subsamples (filters) scaled to chlorophyll *a* per square metre, see Section 2.3.1) would have met thresholds specified in the NPS-FM 2020 periphyton attribute (and in some regional plans). The numbers of samples in each threshold group (i.e., attribute band) matched reasonably well. The numbers for laboratories I and II were identical, and one sample was assigned to a different band for laboratories III and IV. At least two samples were in different bands for laboratories V and VI. Unfortunately, the two samples not received in laboratories V and VI (one in each laboratory) were both high chlorophyll *a* samples. Therefore, the full extent of the differences is unknown.

Table 4-1: Numbers of samples meeting or exceeding chlorophyll *a* thresholds (bands) following analysis in laboratories I to VI. Samples are the means of three subsamples (filters) . The concentrations per unit area were obtained by scaling up the amounts per sample by a constant area sampled. Therefore the results are ilustrative only. *Laboratories V and VI received only 18 samples. The missing samples were in the high biomass range (see Appendix A).

NPS-FM threshold/band (mg/m^2)	Numbers of samples meeting or exceeding chlorophyll *a* thresholds					
	I	II	III	IV	V	VI
< 50 (Band A)	7	7	7	8	7	8
50 – 120 (Band B)	8	8	8	7	6	8
120 – 200 (Band C)	3	3	2	3	4	2
> 200 (Band D)	1	1	2	1	1*	0*
Total number of samples:	19	19	19	19	18	18

4.5 Implications for the NEMS *Periphyton* and laboratory providers

The purpose of this inter-laboratory comparison was to obtain information about variability in chlorophyll *a* results to support, or expand on, the current specification in the draft NEMS *Periphyton* that laboratories must follow the laboratory method described in Biggs and Kilroy (2000) for data to be eligible for the highest quality rating (QC 600). Laboratory processing of samples for chlorophyll *a* content requires several steps including blending, sub-sampling, pigment extraction, and instrument reading of absorbances with an acidification step, all of which may affect the final measurement value. Our comparison included only extraction and instrument reading, for reasons outlined in Section 1.

Six laboratories took part in the comparison, four of which used the Biggs and Kilroy (2000) method (extraction of chlorophyll *a* using 90% ethanol, laboratories I to IV) and two of which used an APHA method using acetone (laboratories V and VI).

Considering that: (a) the results from laboratories V and VI (acetone) showed the largest discrepancies (25% difference, on average), and (b) the discrepancies between laboratory pairs including V or VI were larger than those between laboratory pairs including only laboratories I to IV (respectively 7.8% vs 16.6%, on average), retention of the ethanol extraction method (from Biggs and Kilroy 2000) for achieving QC 600 in the NEMS *Periphyton* is warranted.

If laboratories currently using acetone as an extractant wish to retain their current methodologies and agencies (e.g., regional councils) wish to retain their current laboratory providers (for consistency over time), the results of the present comparison could be used to check for the extent of deviations from an overall mean value across all laboratories, using the equations summarised in Table 3-6. For example, the agency could estimate the overall mean value for samples that are close to the chlorophyll *a* thresholds of interest (once scaled up to mg/m^2). Application of the equations in Table 3-6 in this way would be on the assumption that the samples analysed in the comparison were representative of periphyton samples in general. The agency would need to know which laboratory of those in the comparison (I to VI) is their provider.

Regardless of whether the check above is applied, chlorophyll *a* measurements performed by using acetone as the extractant should continue to be eligible for a maximum quality code rating of QC 500.

To align with QC 600 specifications in the NEMS *Periphyton*, laboratories using acetone extraction may opt to change their extraction method. In this case, the usual practice of carrying out parallel analyses using both methods should be applied (see draft NEMS *Periphyton*, Section 6.5 Managing Changes in Laboratory Methods).

It is also noted that IANZ accreditation (ISO 107 025) provides recognition of adherence to a high quality standard by laboratories to any nominated method. Accreditation status of laboratories can be checked on the IANZ website https://www.ianz.govt.nz/directory.

5 Conclusions and recommendations

The interlaboratory comparison showed:

- Estimates of chlorophyll *a* obtained by each laboratory for the 19 samples (or 18 samples for laboratories V and VI) were closely correlated with those from the other laboratories.
- Absolute discrepancies between laboratories varied and generally increased as chlorophyll *a* concentration increased.
- With the exception of laboratory pairs I vs II and I vs IV, differences in the measurements returned from all possible pairs of laboratories suggested some bias in measurement towards higher or lower values in the different laboratories. The order from lowest to highest values (on average) was laboratory VI, IV, I, II, III and V.
- Percentage differences from the overall mean value (all laboratories) were not clearly related to chlorophyll *a* concentration and ranged from -31% to 35% on individual samples.
- Laboratories I to IV followed the analysis method currently specified in the draft NEMS *Periphyton* (chlorophyll *a* extraction using boiling 90% ethanol). Discrepancies between these four laboratories were modest (mean difference of 7.8 ± 3.5% between laboratories).
- Laboratories V and VI followed an APHA method based on extraction of chlorophyll *a* in acetone. The results from both these laboratories diverged from those of laboratories I to IV (mean difference of 16.6 ± 6.6% between laboratories) and also diverged from each other (mean difference of 25%).
- The individual results from all six laboratories explained at least 97% of the variance in the overall mean chlorophyll *a* (across all six laboratories) (linear regressions), suggesting that the overall mean value may be predictable in all laboratories using laboratory-specific relationships.

The following recommendations resulted from the comparison:

- In view of (a) the inconsistency in results between laboratories V and VI (using acetone), and (b) generally larger discrepancies between V and VI and the laboratory I to IV results (using 90% ethanol) than discrepancies within I to IV, the QC 600 specification for ethanol extraction in the NEMS *Periphyton* should be retained.
- Chlorophyll *a* measurements performed by laboratories using acetone should continue to be eligible for a maximum quality code rating of QC 500.
- Of the laboratories using ethanol as an extractant, results from laboratories III and IV deviated farthest from the mean and reviews of method details in those laboratories may be useful.
- A separate comparison of method details in laboratories V and VI is required to identify why the results differed so markedly.

Regarding achieving consistency of chlorophyll *a* results over time and across laboratories, it is suggested that:

- If a council wishes to retain its current laboratory chlorophyll *a* extraction method (for consistency over time), then checks on the potential effect on use of the data (e.g., for comparison with NPS-FM band thresholds) could be made by applying the appropriate laboratory-specific equation (derived from the test data, see Table 3-6) to obtain an estimate of the laboratory average value.
- Laboratories using acetone extraction may opt to change their extraction method to align with NEMS QC 600 requirements, and the usual practice of carrying out parallel analyses using both extraction methods should be applied.

6 Acknowledgements

Organisation and reporting of this inter-laboratory comparison was funded by an MBIE Envirolink Small Advice Grant (No. 2117-TSDC177) to Tasman District Council. Participating laboratories are thanked for their contribution to the comparison. In particular we thank Adrian Spence (Bay of Plenty Regional Council), Catherine Moisan and Natsumi Nishikawa (Cawthron), Jane Sherrard and Thorne Pullan (Hill Labs), Karen Robinson (NIWA Christchurch), Mike Crump (NIWA Hamilton) and Ben Ludgate (Ryder Environmental Ltd.) for their assistance during the project. Juliet Milne (NIWA) is thanked for initial organisation of the project and for her review of the draft report.

7 References

American Public Health Association (APHA) American Water Works Association (AWWA), Water Environment Federation (WEF) (2017) Standard methods for the examination of water and wastewater, 23rd ed. Washington, DC: American Public Health Association.

Biggs, B.J.F., Kilroy, C (2000) Stream periphyton monitoring manual. NIWA: Christchurch, 226 p.

Hambrook Berkman, J.A.; Canova, M.G. (2007) Algal biomass indicators (ver. 1.0): U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chap. A7, section 7.4, August, available online only from http://pubs.water.usgs.gov/twri9A/.

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.

Lan, S., Wu, L., Zhang, D., Hu, C., Liu, Y (2011) Ethanol outperforms multiple solvents in the extraction of chlorophyll-a from biological soil crusts. Soil Biology and Biogeochemistry 43: 857-861.

Martín, J., Velázquez, N., Asuero, A.G. (2017) Youden two-sample method, quality control and assurance - an ancient Greek term re-mastered, Leo D. Kounis, IntechOpen, DOI: 10.5772/66397. Available from: https://www.intechopen.com/books/quality-control-and-assurance-an-ancient-greek-term-re-mastered/youden-two-sample-method

NEMS. In prep. Periphyton: Sampling and measuring periphyton in wadeable rivers and streams. National Environmental Monitoring Standards, draft, 30 October 2020.

New Zealand Government. 2020. National Policy Statement for Freshwater Management 2020.

Sand-Jensen K.A.J. (1976) A comparison of chlorophyll *a* determinations of unstored and stored plankton filters extracted by methanol and acetone. Vatten, 4, 337-341.

Sartory, D.P. and Grobbelaar, J.U. (1984) Extraction of chlorophyll *a* from fresh-water phytoplankton for spectrophotometric analysis. Hydrobiologia 114(3), 177-187.

Szewczak, E., Bondarzewski, A. (2016) Is the assessment of interlaboratory comparison results for a small number of tests and limited number of participants reliable and rational? Accreditation and Quality Assurance 21: 91-100.

Wasmund, N., Topp, I., Schories, D (2006) Optimising the storage and extraction of chlorophyll samples. Oceanologia 48 (1) 125-144.

Webb, D., Burnison, B., Trimbee, A. and Prepas, E. (1992) Comparison of chlorophyll *a* extractions with ethanol and dimethyl sulfoxide/acetone, and a concern about spectrophotometric phaeopigment correction. Canadian Journal of Fisheries and Aquatic Sciences 49, 2331-2336.

Youden, W.J. (1959) Graphical diagnosis of interlaboratory test results. Ind. Qual. Control. 1959; 15(11):24–28.

Appendix A Raw results from all laboratories

The reported quantities are milligrams of chlorophyll *a* per sample. Amounts per filter were scaled up to amounts per sample using the initial volume of the sample and the volume filtered (15 mL in all cases) provided with the samples. The required chlorophyll *a* concentration range was achieved (spectrophotometer readings on the unacidified extracted sample ranging from <0.015 to >0.9).

	Chlorophyll *a* (mg per sample) reported from each laboratory					
Sample label	**I**	**II**	**III**	**IV**	**V**	**VI**
BA5	0.52	0.54	0.64	0.55	0.72	0.49
BA5	0.57	0.60	0.56	0.53	0.65	0.53
BA5	0.51	0.62	0.61	0.52	0.67	0.53
BA8	0.55	0.78	0.61	0.61	0.85	0.56
BA8	0.55	0.69	0.63	0.56	0.87	0.52
BA8	0.5	0.65	0.67	0.52	0.81	0.36
BA11	0.42	0.46	0.43	0.39	0.55	0.34
BA11	0.35	0.39	0.39	0.31	0.46	0.28
BA11	0.23	0.32	0.37	0.31	0.41	0.28
BA16	0.24	0.25	0.25	0.24	0.28	0.25
BA16	0.28	0.30	0.36	0.27	0.32	0.27
BA16	0.23	0.22	0.30	0.24	0.31	0.28
BA20	0.76	0.77	0.82	0.6	0.88	0.58
BA20	0.79	0.67	0.79	0.63	0.76	0.65
BA20	0.74	0.82	0.85	0.73	0.94	0.79
BA21	0.86	0.96	1.05	0.96	1.26	0.85
BA21	0.88	0.93	0.92	0.81	1.07	0.45
BA21	0.96	1.27	1.18	0.88	1.58	0.88
BA24	0.050	0.048	0.056	0.045	0.058	0.022
BA24	0.040	0.048	0.056	0.027	0.061	0.032
BA24	0.050	0.048	0.056	0.053	0.051	0.046
AM114	0.040	0.046	0.050	0.043	0.056	0.036
AM114	0.050	0.043	0.056	0.027	0.047	0.041
AM114	0.050	0.057	0.061	0.043	0.069	0.044
AM120	0.53	0.57	0.58	0.45	0.68	0.47
AM120	0.48	0.50	0.60	0.53	0.66	0.41
AM120	0.53	0.67	0.67	0.6	0.76	0.38
AM122	1.09	1.26	1.12	1.1	1.55	0.98
AM122	1.2	0.93	1.24	1.1	1.32	0.46
AM122	1.33	1.69	1.90	1.3	1.99	1.84
AM124	0.42	0.51	0.51	0.35	0.49	0.43
AM124	0.37	0.48	0.44	0.37	0.44	0.42
AM124	0.47	0.48	0.55	0.51	0.50	0.54
AM125	2.18	2.04	2.14	1.9	missing	1.72
AM125	2.26	1.94	missing	1.8	missing	1.09
AM125	2.29	1.85	2.23	2.4	missing	1.21
AM131	0.45	0.32	missing	0.31	0.31	0.17
AM131	0.31	0.32	0.30	0.27	0.31	0.19
AM131	0.32	0.32	0.24	0.24	0.24	0.23

	Chlorophyll *a* (mg per sample) reported from each laboratory					
Sample label	**I**	**II**	**III**	**IV**	**V**	**VI**
AM132	1.03	1.09	1.17	1.1	1.26	missing
AM132	1.13	1.08	1.21	1.1	1.20	missing
AM132	0.95	1.27	1.15	1.1	1.40	missing
AM142	0.27	0.24	0.28	0.24	0.23	0.23
AM142	0.26	0.26	0.26	0.28	0.24	0.21
AM142	0.23	0.28	0.35	0.29	0.32	0.27
AM144	0.08	0.08	0.08	0.08	0.09	0.07
AM144	0.08	0.08	0.09	0.08	0.09	0.07
AM144	0.11	0.10	0.09	0.086	0.09	0.09
AM146	0.08	0.07	0.09	0.075	0.09	0.08
AM146	0.09	0.07	0.08	0.07	0.07	0.07
AM146	0.09	0.09	0.10	0.08	0.10	0.08
AM151	0.43	0.46	0.44	0.44	0.53	0.32
AM151	0.48	0.52	0.41	0.44	0.49	0.41
AM151	0.52	0.52	0.57	0.65	0.62	0.61
AM158	0.38	0.40	0.46	0.39	0.46	0.43
AM158	0.48	0.47	0.45	0.48	0.53	0.35
AM158	0.58	0.48	0.60	0.57	0.52	0.49
Extra samples (laboratories IV and VI)						
AM132				1.2		
AM132				1.2		
AM132				1.1		
AM125						1.68
AM125						1.82
AM125						1.65

Appendix B Z-scores for pairs of samples with similar chlorophyll *a*

Z-scores (i.e., the number of standard deviations from the standardised mean) calculated for pairs of consecutive samples with samples arranged in order of increasing chlorophyll *a*. For example, Z-scores for the comparison between samples BA24 and AM114 are shown against sample AM114. Martin et al. (2017) suggested that Z-scores <2 indicate "satisfactory" correspondence.

Median concentrations of chlorophyll *a* were calculated from three replicate subsamples. Laboratories provided results as mg/sample, which was converted to mg/m^2 by dividing by an area of 0.007 m^2 (equivalent to 10 sub-samples of periphyton each scraped from an area of 3 cm diameter. Note that this calculation was performed simply to convert the raw results to a more familiar scale. The areas from which samples were scraped are not known.

NA = no sample.

Sample	Median concentration		Z-score for comparison of consecutive samples in each laboratory					
	mg/sample	mg/m^2	I	II	III	IV	V	VI
BA24	0.05	6.7						
AM114	0.05	6.8	0.1	0.1	0.8	0.9	0.9	0.7
AM146	0.08	11.7	0.5	0.5	1.0	0.7	0.4	0.4
AM144	0.09	12.4	0.4	0.1	0.6	0.6	0.1	0.9
BA16	0.26	37.3	0.4	0.2	1.6	0.4	1.6	0.2
AM142	0.26	37.4	0.3	0.1	1.1	0.2	0.1	0.8
AM131	0.28	40.1	1.5	0.8	0.2	0.1	0.1	1.5
BA11	0.36	51.7	0.4	0.3	0.4	0.3	1.5	0.8
AM124	0.47	67.1	0.7	0.3	0.4	0.8	0.1	0.1
AM158	0.48	68.6	0.0	0.6	0.5	0.0	0.5	1.2
AM151	0.49	69.9	0.3	0.3	0.3	0.4	1.2	0.9
AM120	0.55	79.0	0.4	0.3	0.7	0.3	1.5	1.4
BA5	0.56	80.1	0.3	0.3	0.4	0.3	1.1	0.4
BA8	0.60	85.8	0.5	0.8	0.3	0.3	1.8	0.9
BA20	0.76	108.4	0.0	0.0	0.4	0.7	0.7	0.6
BA21	0.97	139.3	0.4	0.4	0.4	0.5	1.7	1.3
AM132	1.15	163.7	0.3	0.0	0.1	0.1	0.3	NA
AM122	1.25	178.5	0.1	0.1	0.4	0.2	0.8	0.3
AM125	2.03	290.5	0.3	0.1	0.2	0.0	NA	0.9
		Mean Z score	**0.37**	**0.29**	**0.53**	**0.38**	**0.86**	**0.78**