Memo

| | |
|---|---|
| From | Dr Scott Larned<br>Dr Cathy Kilroy |
| To | Dr Jennifer Price<br>Ministry for the Environment<br>23 Kate Sheppard Place Wellington 6143 |
| Date | 28 January 2020 |
| Subject | Assessments of MfE (2019) and DairyNZ (2019) reports |
| Reviewed by | Dr Clive Howard Williams |
| Approved by | Dr Bryce Cooper |


**Summary**

MfE requested technical reviews of specific matters in each of two reports: *Ministry for the Environment. 2019. Essential Freshwater: Impact of existing periphyton and proposed dissolved inorganic nitrogen bottom lines* (hereafter, MfE (2019)), and DairyNZ 2019. *DairyNZ review of periphyton baseline to underpin economic modelling of Essential Freshwater policies* (hereafter, DairyNZ (2019)).

For MfE (2019) NIWA was requested to:

- A1. Evaluate the suitability of the methods in MfE (2019) to estimate, at a national scale, the nutrient reductions required to meet the periphyton bottom line set out in the existing National Policy Statement for Freshwater Management (NPS-FM 2017).
- A2. Outline the implications of choosing different spatial exceedance criteria for predicting nutrient load reductions required, considering their suitability for predicting, at a national scale, the actions needed to meet the current periphyton attribute.



National Institute
of Water & Atmospheric
Research Ltd (NIWA)

10 Kyle Street
Riccarton
Christchurch 8011
PO Box 8602
Riccarton
Christchurch 8440

P: +64 3 348 8987
enquiries@niwa.co.nz
www.niwa.co.nz

For DairyNZ (2019), NIWA was requested to:

- B1. For Section 2 (DairyNZ analysis of current periphyton state) and Section 3 (DairyNZ review of measured versus modelled periphyton biomass), review the methods employed and conclusions reached.
- B2. Evaluate the accuracy of the statements in Section 5 (Summary of aligned positions).

Specific comments on sections and statements in these two reports are given in the main body of the report. In this section, we summarise our assessments.

Regarding MfE (2019):

1. We consider the methods in MfE (2019) to be generally clear and logical. We have set out some suggestions regarding these methods below. First, we suggest that the modelling chain used in MfE (2019) is explained with greater clarity, using a flow chart.
2. In general, the performance of the models in the modelling chain was as good as can be expected, given data limitations, the large number of environmental factors that influence the response variables, and inherent variability in those environmental factors. The evaluation metrics for the nitrogen concentration and load models indicated satisfactory performance. The explanatory power of the regression model for predicting periphyton weighted composite cover (WCC) was relatively low ($R^2$ = 0.40), and moderate for the model for predicting periphyton biomass from WCC ($R^2$ = 0.65).
3. Despite the limitations, the MfE (2019) approach is the best available tool for estimating nitrogen reductions to achieve the periphyton bottom line at the national scale. Depending on the specific application and the level of confidence required, we recommend using the MfE (2019) approach.
4. There may be potential to improve the performance of the models used to predict river TN concentrations and loads and periphyton biomass by increasing the quantity and spatial extent of observational data.
5. We note that spatial exceedance criteria for periphyton are not stipulated in the NPS-FM 2017 or in the Essential Freshwater package. In contrast, temporal exceedance is built into the current periphyton attribute, which requires that the numeric attribute state in each band is "*exceeded no more than 8% of samples*" at a given site. The footnote to the

periphyton attribute in the NPS-FM 2017 does require setting exceedance criteria for DIN and DRP within freshwater management units, but does not specify whether they are spatial or temporal exceedance criteria. While they are not required by the NPS-FM, MfE (2019) has used spatial exceedance to address the uncertainty in the model that predicts periphyton biomass from TN across many river segments.

6. We note that spatial exceedance criteria are a logical way to account for model uncertainty when setting nutrient concentration targets that apply to many sites. For cases in which regression models are used to derive nutrient concentration targets based on statistical relationship with response variables (i.e., dose-response relationships), 50% of observations at or below the target concentration are expected to exceed the desired response and 50% to be less than the desired response. In other words, the model-based prediction corresponds to the 50% exceedance level. As used in MfE (2019), spatial exceedance criteria have the effect of reducing the target TN concentration from the predicted concentration corresponding to the periphyton bottom line (i.e., the 50% site exceedance level) to TN concentrations at which 10% or 20% of sites are expected to exceed the bottom line. Therefore, the spatial exceedance criteria used in MfE (2019) increased the stringency of TN concentration targets from those predicted directly from models for each river class. We note that decisions to use spatial exceedance criteria or not, and the choice of specific exceedance criteria are policy/regulatory decisions, which can be informed by technical analyses.

Regarding DairyNZ (2019), Sections 2, 3 and 5:

1. This report is intended to set out reasons for using 'current water quality state' as 'the baseline'. However, the definition and role of 'the baseline' is not explained, and the report refers to five baselines (periphyton baseline, NPS-FM baseline, nutrient baseline, MfE baseline, DairyNZ baseline). These may be different baselines or some or all may be synonyms. Statement 1.4 of the report: "DairyNZ used the current water quality state as the baseline…" suggests that 'DairyNZ baseline' refers to current river nitrogen or phosphorus concentrations. However, it is not clear from DairyNZ (2019) how current river concentrations alone, or in combination with other data, are to be used to estimate

required nutrient reductions to achieve the periphyton bottom line. These omissions limited our ability to assess the conclusions of the report.

2. The modelling approach in MfE (2019) is criticised in DairyNZ (2019), and some statements regarding model performance are accurate. However, the modelling approach in MfE (2019) and model performance are two separate issues, and the conclusion in DairyNZ (2019) that poor model performance indicates that the approach is wrong is not logical. As noted above, the MfE (2019) method is the best currently available for national-scale estimates of nutrient reductions required to achieve the periphyton bottom line. The data used to run the models could be improved, but the approach is sound.
3. If the aim of DairyNZ (2019) was to set out an alternative approach to MfE (2019) for estimating required nutrient reductions for achieving the periphyton bottom line, that approach was not clearly identified in the report.
4. Some of the statements in Sections 2, 3 and 5 of DairyNZ (2019) are factually correct. Others are incorrect; both are identified in the text below.

## 1. Introduction

The Ministry for the Environment (MfE) has requested technical reviews of two reports concerning nitrogen reductions required to meet the current National Objective Framework (NOF) bottom line for periphyton biomass:

A. Ministry for the Environment. 2019. Essential Freshwater: Impact of existing periphyton and proposed dissolved inorganic nitrogen bottom lines. Wellington: Ministry for the Environment. Hereafter "MfE (2019)".

B. DairyNZ. 2019. DairyNZ review of periphyton baseline to underpin economic modelling of Essential Freshwater policies. Hereafter "DairyNZ (2019)".

MfE requested assessments of specific matters in each of the two reports.

For MfE (2019):

- A1. Evaluate the suitability of the methods in MfE (2019) to estimate, at a national scale, the nutrient reductions required to meet the periphyton bottom line set out in the existing National Policy Statement for Freshwater Management (NPS-FM 2017).
- A2. Outline the implications of choosing different spatial exceedance criteria for predicting nutrient load reductions required, considering their suitability for predicting, at a national scale, the actions needed to meet the current periphyton attribute.

For DairyNZ (2019):

- B1. For Section 2 (DairyNZ analysis of current periphyton state) and Section 3 (DairyNZ review of measured versus modelled periphyton biomass), review the methods employed and conclusions reached.
- B2. Evaluate the accuracy of the statements in Section 5 (Summary of aligned positions). Note that Section 4 (DairyNZ comments on the proposed DIN ecosystem health thresholds) was out of scope.

Each report was taken as read, without assessing all of the previous reports cited in DairyNZ (2019) and MfE (2019). We assumed that information in previous reports was used and referenced appropriately. There was one exception, Snelder (2018), which we did include in our assessment, because the methods and results in this report were extensively used and cited in both DairyNZ (2019) and MfE (2019).

## 2. MfE (2019)

In this report, a multi-step process was used to estimate the nitrogen load reductions required in rivers to meet the NOF periphyton bottom line (200 mg Chl *a* $m^{-2}$ as a $92^{nd}$ percentile calculated from at least 3 years of monthly monitoring data). The rationale for considering nitrogen load reductions to achieve the periphyton bottom lines (and nitrate toxicity bottom lines in waterways that do not support and could not support conspicuous periphyton) is set out in the footnote to the periphyton attribute table in the NPS-FM 2017: *"Regional councils must use the following process, in the following order, to determine instream nitrogen and phosphorus criteria in a freshwater management unit:*

*a) either –*

> *i) if the freshwater management unit supports, or could support, conspicuous periphyton, derive instream concentrations and exceedance criteria for DIN and DRP to achieve a periphyton objective for the freshwater management unit; or*
>
> *ii) if the freshwater management unit does not support, and could not support, conspicuous periphyton, consider the nitrogen and phosphorus criteria (instream concentrations or instream loads) needed to achieve any other freshwater objectives:*

*b) if there are nutrient sensitive downstream environments, for example, a lake and/or estuary, derive relevant nitrogen and phosphorus criteria (instream concentrations or instream loads) needed to achieve the outcomes sought for those sensitive downstream environments:*

*c) compare all nitrogen and phosphorus criteria derived in steps (a) – (b) and adopt those necessary to achieve the freshwater objectives for the freshwater management unit and outcomes sought for the nutrient sensitive downstream environments.*

Only Clause (a)1 above concerns the periphyton bottom line. In MfE (2019), the requirement in Clause (b) to develop instream nitrogen and phosphorus criteria for nutrient-sensitive downstream environments is addressed only in terms of downstream river segments, not lakes or estuaries.

In MfE (2019), required reductions were also estimated for the NOF nitrate toxicity bottom line (6.9 mg $NO_3$-N $L^{-1}$), and the DIN bottom line proposed in the Essential Freshwater package (1 mg N $L^{-1}$). However, MfE has specifically requested that we assess the methods in MfE (2019) as they concern the periphyton bottom line. Therefore, we have not assessed the methods as they concern the toxicity or DIN bottom lines. The required load reductions to achieve the periphyton bottom line were estimated in terms of total nitrogen (TN).

The issues we were asked to assess in MfE (2019) concern methods for national-scale estimations. Therefore, we have not assessed the results section of MfE (2019), which consists of national maps of compliance and catchment excess loads and yields, or the case studies of impacted catchments.

### 2.1. Analysis steps in MfE (2019)

The major steps used in MfE (2019) to estimate nitrogen criteria and load reductions required to meet bottom lines are set out below. Note that these are general steps, most of which comprise several substeps that are not specified here. The list of general steps is included to help readers follow our assessments of MfE (2019) and DairyNZ (2019). This list also brings together steps that are currently located in two separate reports, Snelder (2018) and MfE (2019). For clarity in the context of this technical review, we have grouped these steps into seven 'Parts' as described below.

#### Part 1. Predict nitrogen concentrations and loads

1. Current TN and $NO_3$-N concentrations were estimated for all river segments in the New Zealand digital river network using random forest models. The methods and predictor variables followed those in Whitehead (2018), except that predicted livestock units were added as predictor variables to represent land-use intensity.

2. Current TN loads and yields were estimated for all river segments by rating curves (based on river flow and TN concentration data) and a random forest model, using the methods in Snelder et al. (2018).

**Part 2. Identify segments that support, or could support, conspicuous periphyton**

3. The substrate size index in FENZ was used to distinguish segments in the digital river network that are dominated by fine sediments (substrate size index values < 3) or coarse sediments (values ≥3). Only segments predicted to be dominated by coarse sediments were used to estimate TN concentration criteria for the periphyton bottom line. Note that there appears to be an error in the first paragraph of MfE (2019) Section 2.3: "*River segments with coarse and fine bed substrates were discriminated using substrate size index values of <3 and ≥3 respectively*". According to Leathwick et al. (2010, 2011): the substrate size index is based on "*weighted average of proportional cover of bed sediment using categories of 1 – mud; 2 – sand; 3 – fine gravel; 4 – coarse gravel; 5 – cobble; 6 – boulder; 7 – bedrock".* So <3 refers to the fine categories, not the coarse categories.

**Part 3. Derive TN concentration criteria to achieve the periphyton bottom line to apply to all river segments that support, or could support, conspicuous periphyton**

4. TN concentration criteria for the periphyton bottom line were derived and associated uncertainty quantified for coarse-substrate segments within each of 21 REC source-of-flow classes. This step was based on: (1) an underlying model for predicting 92$^{nd}$ percentile chlorophyll *a* densities and (2) Monte Carlo simulations. Detailed methods are in Snelder (2018); the major substeps are:
    A. River segments of Strahler order 4 or greater, in each of the 21 REC classes, were drawn at random from the digital river network in domains of 25, 100 and 1000 segments. Each set of random samples was termed a "realisation".
    B. Data from 78 training sites in the National River Water Quality Network (NRWQN) were used to develop a multiple linear regression model (hereafter, 'NRWQN model') that predicted the 92$^{nd}$ percentile of periphyton weighted composite cover

(WCC) as a function of eight predictor variables (three hydrological metrics, water temperature, PAR, DIN:DRP, and TN concentration).

C. 92nd percentile values of √WCC were predicted for all river segments in each of the 21 REC classes using the hydrological metrics, water temperature, and PAR predictor variables and holding the TN concentration constant.
D. Predicted WCC values were converted to periphyton biomass (mg Chl *a* $m^{-2}$) (the biomass attribute unit specified in the NOF) using a regression based on paired WCC and biomass measurements from 154 regional council monitoring sites.
E. Steps 4C and D were repeated 20 times while increasing the TN concentrations in 20 equal increments that covered the range of observed TN in the dataset).
F. The TN concentrations for which 10% or 20% of river segments exceeded five chlorophyll *a* thresholds (50, 120, 200, 400 and 800 mg $m^{-2}$) were interpolated from the output for each realisation. The 10% or 20% of segments exceeding the thresholds were used as 'spatial exceedance criteria', discussed below.
G. Steps 4A-4F were repeated for 1000 realisations.
H. For each domain size (25, 100 and 1000 segments), the mean values of the 1,000 realisations at which 10% or 20% of segments exceeded the periphyton bottom line were retained as the nutrient targets.
I. The standard deviations of the values over the 1,000 realizations were used to calculate standard errors and 95% confidence intervals for the nutrient targets.

**Part 4. Test the performance of the targets using independent data**

5. The independent data set consisted of periphyton biomass measurements from 173 regional council sites (or 'independent sites'). The 92nd percentile periphyton biomass and standard error at each independent site was estimated from the mean biomass, based on an exponential distribution. These values were referred to as 'observed values'. The 92nd percentile periphyton biomass and standard error at each of the independent sites were then predicted using the TN targets for each REC class and both spatial exceedance criteria (10%, 20%). These values were referred to as 'predicted values'. The observed values were plotted against the predicted values in the two spatial exceedance levels to visually check

whether the 10% and 20% of the observed biomass values exceeded the predicted biomass values as intended.

6. The mean and standard deviation of the proportion of sites at which the observed 92nd percentile periphyton biomass exceeded the predicted value was estimated by Monte Carlo simulation.

**Part 5. Recalibrate the targets to align with the independent sites**

7. The results of Part 4 indicated that the proportions of independent sites exceeding the periphyton bottom lines were less than the expected 10% and 20% (viz. 2% and 9% respectively), indicating that the TN targets were too conservative (i.e., too stringent). The TN targets were then recalibrated so that the proportion of sites exceeding the periphyton bottom line more closely matched the spatial exceedance criteria. We note that detailed methods for recalibration are not provided in MfE (2019) or in Snelder (2018). The only reference to the methods is "*As suggested by Snelder (2018), the original TN targets were recalibrated to match the observations at the 173 river monitoring sites*" (page 10 of MfE (2019)).

**Part 6 Estimate load targets (maximum allowable load) and local excess load**

8. For each river segment that does support or could support conspicuous periphyton, the maximum allowable TN load (MAL) corresponding to the 10% and 20% spatial exceedance criteria were estimated from the predicted current TN concentration (Step 1 above) and TN load (Step 2 above) and the target TN concentration for each of the 21 REC source-of-flow classes.
9. For each segment in Step 8 at which the current TN load exceeded the MAL, the 'local excess load' was calculated as the difference between the current load and MAL.

**Part 7 Estimate catchment excess loads and required load reductions**

10. MfE (2019) defines catchment excess load the amount by which the current load would need to be reduced at a segment to comply with the concentration criteria at that segment

and all upstream segments. For each New Zealand catchment that drains to the sea, catchment excess TN loads were estimated in three substeps, 10A-10C.

10A. The local excess TN load was estimated for each segment in the catchment.

10B. The catchment excess TN loads for the most upstream network segments were equal to their local excess loads.

10C. Starting with the most upstream segments and moving downstream, the catchment excess load was compared to the local excess load of the next segment downstream. If the local excess load at the next segment was less than the catchment excess load of the upstream segment, the catchment excess load of the downstream segment was set equal to the catchment excess load of the upstream segment. If the reverse applied, the catchment excess load of the downstream segment was set equal to its local excess load.

11. Required TN load reductions were as a percentage of the current TN load by dividing catchment excess loads by the corresponding current load.

**2.2. Evaluate the suitability of the methods in MfE (2019) to estimate, at a national scale, the nutrient reductions required to meet the periphyton bottom line in the existing NPS-FM.**

We interpreted 'suitability' to refer to two aspects of the methods used in MfE (2019), logic and clarity (i.e., whether the methods comprise a sensible series of steps, clearly explained and whether the stated assumptions are valid),-and reliability (i.e., model performance and uncertainty of model predictions). We consider both aspects below.

**Logic and clarity**. We consider the methods in MfE (2019) to be generally clear and logical. The methods comprise a large number of steps that readers may find challenging to follow, and there may be alternative approaches for some steps, as discussed below. However, the basic elements (predicting concentrations and loads, predicting periphyton abundance, relating nitrogen to periphyton, deriving and testing criteria, estimating local and catchment excess nitrogen loads) are all needed to estimate required load reductions. The same basic elements have been used in recent publications about achieving bottom lines through nutrient load reductions (Elliott et al. 2019), and are likely to be used in the future as NPS-FM implementation proceeds.

The explanation of methods in MfE (2019) would benefit from some clarifications and additional information. In Section 2.3 of the report, on concentration criteria and current compliance, readers are referred to Snelder (2018) for the explanation of the steps used to derive TN concentration criteria to achieve the periphyton bottom line and test the criteria. Snelder (2018) sets out a 19-step process involving two separate Monte Carlo simulations. It would help readers to have a comprehensive set of methods in one document. A flow chart to accompany the written methods would also be helpful.

Several assumptions made in MfE (2019) and Snelder (2018) would benefit from a thorough critique. These include:

1. The substrate size index cut-off of 3 distinguishes coarse sediment-dominated river segments that can support conspicuous periphyton from fine-substrate segments that do not and cannot support conspicuous periphyton. This assumption has two parts: (1) that average sediment size is accurately estimated across the river network with low uncertainty; and (2) that periphyton do not occur in reaches where the average sediment grain size corresponds to mud or sand (substrate size index <3). If these assumptions have been verified, that information should be included.

2. The periphyton bottom line can be achieved purely by managing instream nutrient concentrations (implying that shading by riparian planting or other control methods are not effective or necessary means of achieving the periphyton bottom line). We agree with the points made that riparian shading is not applicable at all river segments, and that reducing periphyton by shading in some segments may inadvertently increase nutrient loads in downstream segments. Further, the TN targets were estimated for 4$^{th}$-order river segments and larger, which will limit the number of narrow streams that can develop riparian canopies. However, there is no information provided about how much of the river network (≥ 4$^{th}$-order) that currently supports periphyton can or cannot be managed through riparian shading. Land management strategies aimed at controlling periphyton may need to consider a combination of nutrient reduction, riparian shading, flow management and other methods.

3. The availability of nitrogen for periphyton can be accurately represented by TN concentrations and loads, rather than dissolved inorganic nitrogen (DIN) concentrations and loads. TN and various forms of DIN are both used by researchers to quantify relationships between nitrogen and periphyton abundance and growth. Some researchers recommend using TN due to greater data availability, greater explanatory power in statistical analyses, and high DIN turnover (i.e., rapid DIN uptake by periphyton can make DIN availability appear low). The counterargument for using DIN rather than TN is based in part on the fact that algal cells assimilate DIN from water, but not the particulate or large organic fractions of TN. Therefore, TN can overestimate DIN availability. The footnote to the periphyton attribute in the NPS-FM 2017 specifies DIN as the form of nitrogen to be managed: "*To achieve a freshwater objective for periphyton within a freshwater management unit, regional councils must at least set appropriate instream concentrations and exceedance criteria for dissolved inorganic nitrogen (DIN)".* In MfE (2019), TN was selected for predicting periphyton and setting nitrogen targets over DIN *"because TN produced higher $R^2$ values*". If there was evidence that TN concentrations and loads vary in fixed proportions to DIN concentrations and loads, then converting TN-based criteria to DIN-based criteria would be a simple task. This is assumed in MfE (2019): "*The underlying assumption therefore is that loads of NO3-N and DIN change in proportion to change in the TN load*". However, the difference in the performance of periphyton models based on TN and DIN concentrations suggests that TN and DIN may not be interchangeable. We suggest that data to justify the assumption are added.

**Reliability**

For most of the models in the modelling chain used in MfE (2019), model performance and uncertainty in the resulting estimates were quantified using evaluation metrics (e.g., Nash–Sutcliffe efficiency (NSE), $R^2$, bias, root mean square deviation (RMSD), standard errors, confidence intervals). The evaluation results are summarised here:

Predicting nitrogen concentrations and loads**.** The random forest models for concentrations had NSE values of 0.72 and 0.63 for TN and $NO_3$-N, respectively, and RMSDs 0.26 and 0.45 $_{log}10$ mg $m^{-3}$, for TN and $NO_3$-N, respectively. Bias was close to zero for both models. The random forest model for TN yield had an NSE value of 0.62, RMSD of 0.23 $_{log}$10 kg ha$^{-1}$ yr$^{-1}$) and bias of -0.14 %.

Predicting WCC as a function of TN, light, hydrological metrics and water temperature. The regression model for predicting WCC at each network segment had an $R^2$ of 0.40 and RMSD of 1.94 √mg m$^{-2}$.

Predicting periphyton biomass (as chlorophyll *a*) as a function of WCC. The regression model had an $R^2$ of 0.65 and RMSD of 0.319 $_{log10}$ mg m$^{-2}$.

Deriving TN targets to achieve the periphyton biomass bottom line within REC classes. Uncertainty in the estimates of TN concentration targets that correspond to 92$^{nd}$ percentile periphyton biomass are indicated by the error bars in Figure 4.3 of Snelder (2018). These error bars are not easily distinguished, but a related set of standard errors for the TN targets for each REC class (at 10% and 20% spatial exceedance levels) are listed in the supplementary data file for a published paper (Snelder et al. 2019), which is based on Snelder (2018). The supplementary data file is available at [https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2F1752-1688.12794&file=jawr12794-sup-0001-TableS1-S6.docx].

Predicting proportions of sites exceeding the NOF periphyton biomass bottom line. Table 4.3 in Snelder (2018) has 95% confidence intervals for estimated proportion of sites exceeding the NOF periphyton bottom line, in comparison with the 10% and 20% spatial exceedance criteria. For the 10% criterion, the 95% confidence interval (0-4%), excluded 10%. For the 20% criterion, the 95% confidence interval (4.8 - 11.8%) excluded 20%. Based on these observations, the criteria appear to be too restrictive for the 173 test sites. This observation was the reason for the recalibration, as discussed below.

In general, the performance of the models in the modelling chain was as good as can be expected, given data limitations, the large number of environmental factors that influence the response variables, and the inherent variability in those environmental factors. The evaluation metrics for the nitrogen concentration and load models indicated satisfactory performance, based on comparisons with comparable models (Moriasi et al. 2007, 2015).

The supplementary data file in Snelder et al. (2019) indicate that the standard errors for the TN targets corresponding to the NOF periphyton bottom line were relatively low; less than 15% of the target concentrations for the largest domain size (1000 river segments). The standard errors were highest for some REC classes in extremely cold climate category. However, these classes are probably at low risk of periphyton proliferations, as indicated by their high TN targets[1]. The greatest uncertainty in the chain of models appears to be in the regression used to predict WCC as a function of TN, light and hydrology. Comparable, multiple regression models that relate periphyton abundance across large, heterogeneous regions to nutrient concentrations and other environmental predictors have had similar levels of uncertainty (e.g., Dodds et al. 2002, Snelder et al. 2014).

With regard to deriving TN concentration targets to achieve the periphyton bottom line, our understanding of this procedure is summarised in Part 3 of Section 2.1 above, but we may have misunderstood some steps. It would be helpful if the method was explained as clearly as possible in MfE (2019). As we understand, thousands of predictions of periphyton biomass were made using the regression model based on the 78 NRWQN sites, holding all predictor values constant at each segment while changing the TN concentration in 20 steps (within the range of the NRWQN data). The targets were then interpolated from the predicted chlorophyll *a* values across the range of TN concentrations.

A limitation of all regression models is that predictions outside the range of the training dataset are unreliable. For the periphyton regression model used here, unreliability would apply to predicting outside the range of the combinations of values of all seven predictor variables in the model. It is not clear what effect this approach had on the resulting TN targets, but potential

[1] Note that the target TN concentration targets in the supplementary data file for Snelder et al. (2019) differ from those in MfE 2019. This difference is presumably due to the recalibration step, as discussed in Section 2.3 below.

effects should be acknowledged in the report. A possible rationale for using the regression and interpolation method, despite the potential problem of predicting outside the range of observations, is that problem would be most pronounced at low observed TN concentrations (e.g., predicting periphyton biomass responses to high TN concentrations in river segments where TN concentrations and periphyton biomass are typically low, such as in the REC class CXM). These segments may have unrealistically high TN concentration targets. However, observed TN at such sites would never exceed the targets and there would be no effect on the assessment.

There may be potential to improve the performance of the models used to predict TN concentrations and loads, and periphyton biomass used in MfE (2019) by increasing the quantity and spatial extent of observational data and trialling alternative approaches for deriving nitrogen targets. Most of the data used in the models in MfE (2019) come from national-scale observational datasets and spatial models that characterise river water quality and flows, topography, climate, land cover and other environmental factors. These datasets have relatively high spatial coverage and density. The periphyton dataset from the 173 test sites is an exception. These data came from periphyton monitoring projects run by six regional councils (Northland, Horizons, Bay of Plenty, Greater Wellington, Canterbury, Southland; see Kilroy et al. (2019)). These monitoring sites cover a limited range of environmental space and increasing this range could improve model performance.

A review of alternative approaches to estimate nutrient load reductions required to achieve objectives in aquatic systems (e.g., NOF attribute bottom lines) is out of the scope of this review. We note that multiple alternatives have been used in New Zealand and overseas. All of these approaches have some common elements. They all need to predict spatial (and in some cases temporal) variation in nutrient loads and concentrations, and in periphyton or other response variables, and all require predictive nutrient-response relationships (e.g., periphyton responses to variation in TN concentration). The approaches to carry out these predictive steps range from statistical modelling to process-based modelling, and model structure (e.g., choices of predictor variables) varies widely (e.g., Grizzetti et al. 2005, Mohamoud and Zhang 2018).

**2.3 Outline the implications of choosing different spatial exceedance criteria for predicting nutrient load reductions required, considering their suitability for predicting, at a national scale, the actions needed to meet the current periphyton attribute.**

We start by noting that spatial exceedance criteria for periphyton are not stipulated in the NPS-FM 2017 or in the Essential Freshwater package. This point is acknowledged in MfE (2019) in Section 2.3: "*This approach to defining the TN targets is a means to managing the uncertainty involved in specifying nutrient criteria to achieve periphyton biomass objectives. We have not considered if nutrient criteria that allow for a degree of risk of non-compliance are consistent with the requirements of the NPS-FM (see Section 4 for further discussion)*". In contrast, temporal exceedance is built into the current periphyton attribute, which requires that the numeric attribute state in each band is "*exceeded no more than 8% of samples*" at a given site. The footnote to the periphyton attribute in the NPS-FM 2017 does stipulate exceedances for DIN and DRP, but not for periphyton itself.

MfE (2019) has used spatial exceedance to address the uncertainty in the model that predicts periphyton biomass from TN in many river segments. As stated in Section 4: "*The spatial exceedance criteria represent risks that a site drawn at random will fail to achieve the bottom line. The risk of non-achievement reflects the uncertainties associated with defining criteria for nutrients. Although this risk is not generally explicitly acknowledged in objective setting processes, any criterion that is obtained by nominating a response and using a regression line to define the associated level of a stressor implicitly accepts that approximately 50% of cases (assuming normal residuals) will fail to achieve the specified response value*".

We agree with the logic of the statements quoted above from Sections 2.3 and 4 of MfE (2019). In general, setting limits on resource use to achieve freshwater objectives (including NOF bottom lines) requires the use of statistical relationships between stressors (e.g., nutrients, fine sediment) and responses (e.g., periphyton, phytoplankton). These relationships always have associated prediction uncertainty, and spatial exceedance criteria are an accepted way to set stressor criteria that accounts for prediction uncertainty. We use the term 'accepted' to indicate that there are precedents for the use of spatial exceedance criteria in developing and assessing compliance with water-quality criteria (including nutrients and chlorophyll, comparable to the 'nutrient targets' in MfE (2019)). For example, the US Environmental Protection Agency has assessed compliance with water-quality criteria with spatial and temporal exceedance criteria in rivers and estuaries (e.g., Tango et al. 2013, Zhang et al. 2018, Irby and Friedriches 2019). In most of these cases, the derivation of water-quality criteria and the selection of exceedance criteria were separate processes. The water-quality criteria were generally based on contaminant-response models or reference conditions. The exceedance criteria were based on fixed proportions of sites or samples, or observations of exceedances across reference sites, or on the statistical error in the contaminant-response models (Glibert et al. 2010). We have not seen previous cases where spatial exceedance criteria were incorporated in the derivation of a water-quality target, as in MfE (2019). One reason given for incorporating spatial exceedance criteria in the derivation of a water-quality target as in MfE (2019) is given in Snelder (2018): "*by allowing a spatial exceedance, the uncertainty of the nutrient targets is changed from the error associated with the site-specific prediction of biomass to the more general prediction of what proportion of sites will exceed a given threshold*".

As the statement quoted above from Section 4 of MfE (2019) indicates, when a regression model relating TN concentration to periphyton biomass is used to estimate the TN concentration corresponding to the periphyton bottom line, approximately 50% of the observations (i.e., sites) at or below that TN concentration are expected to have periphyton biomass in excess of the bottom line, and 50% to have periphyton biomass below the bottom line. Therefore, the TN target predicted from the model defines the 50% exceedance level. In MfE (2019), spatial exceedance criteria below 50% were used to ensure that more than 50% of sites in a given REC class achieved the bottom line. The direct implications of this approach are that the target TN concentrations decrease and nutrient reductions required to meet these target increase as the spatial exceedance criteria decrease. In other words, target concentrations and load reduction requirements become more stringent as spatial exceedance criteria decrease. This point is illustrated in MfE (2019) in Table 1 and by comparing the left (10% exceedance) and right (20% exceedance) maps in Figure 6. It is also shown in Figure 4 of Snelder et al. (2019). Note that the 10% and 20% exceedance criteria are arbitrary, and lower (more stringent) or higher (more lenient) exceedances could be applied.

NIWA Taihoro Nukurangi

As noted above, spatial exceedance criteria are not stipulated in the NPS-FM 2017 or in the Essential Freshwater package. They have not been applied to other NOF attributes to date. Decisions to use spatial exceedance criteria (or not), and the choice of specific exceedance criteria are policy/regulatory decisions, which can be informed by technical analysis.

As noted in the summary of methods above (Section 2.1, Part 5), the estimated TN targets were recalibrated in MfE (2019), to make the proportions of test sites exceeding the periphyton bottom line more closely match the 10% and 20% spatial exceedance criteria. The effect of recalibration was to increase the TN targets over those in Snelder (2018), for each REC class (see Table A). On average, recalibration led to a 4-fold increase in the TN targets for both exceedance criteria.

**Table A.** TN concentration targets for achieving the periphyton bottom line of 200 mg chl *a* $m^{-2}$. In each of 21 REC classes. The concentrations in Table 4.1 of Snelder (2018) were not recalibrated. The concentrations in Table 1 of MfE (2019) were recalibrated. The concentrations in Snelder (2018) were reported in mg TN $m^{-3}$. They have been converted here to mg TN $L^{-1}$ to make them comparable to the values in MfE (2019).

| REC class | Snelder (2018) | | MfE (2019) | |
|---|---|---|---|---|
| | 10% spatial exceedance | 20% spatial exceedance | 10% spatial exceedance | 20% spatial exceedance |
| CX/GM | 0.216 | 0.752 | 1.0 | 2.9 |
| CX/M | 0.369 | 1.334 | 1.7 | 3.8 |
| CX/H | 0.385 | 1.381 | 1.7 | 3.7 |
| CX/L | 0.288 | 0.985 | 1.3 | 3.4 |
| CX/Lk | 0.091 | 0.301 | 0.4 | 1.2 |
| CW/GM | 0.112 | 0.351 | 0.5 | 1.3 |
| CW/M | 0.119 | 0.384 | 0.5 | 1.6 |
| CW/H | 0.122 | 0.417 | 0.6 | 1.8 |
| CW/L | 0.097 | 0.323 | 0.5 | 1.3 |
| CW/Lk | 0.066 | 0.21 | 0.3 | 0.8 |
| CD/M | 0.071 | 0.223 | 0.3 | 0.9 |
| CD/H | 0.063 | 0.2 | 0.3 | 0.8 |
| CD/L | 0.066 | 0.209 | 0.3 | 0.8 |
| CD/Lk | 0.056 | 0.178 | 0.2 | 0.7 |
| WX/L | 0.108 | 0.365 | 0.5 | 1.5 |
| WX/H | 0.115 | 0.403 | 0.6 | 1.7 |
| WW/H | 0.168 | 0.576 | 0.8 | 2.5 |
| WW/L | 0.066 | 0.216 | 0.3 | 0.9 |
| WW/Lk | 0.065 | 0.209 | 0.3 | 0.8 |
| WD/L | 0.033 | 0.108 | 0.1 | 0.4 |
| WD/Lk | 0.075 | 0.245 | 0.3 | 1.0 |

.

## 3. DairyNZ (2019)

**B1. For Section 2 (DairyNZ analysis of current periphyton state) and Section 3 (DairyNZ review of measured versus modelled periphyton biomass), review the methods employed and conclusions reached.**

### 3.1. Section 2. DairyNZ analysis of current periphyton state

For clarity, the dataset used in Section 2 of DairyNZ (2019) was the one used to develop the models in Kilroy et al. (2019). Chlorophyll *a* data shown in Figure 1 of DairyNZ (2019) are 92$^{nd}$ percentiles for periphyton biomass, calculated from all data available at each site (data collected over 11 to 105 months, median 40 months), with data adjusted to account for months in which samples were not collected because of floods by assuming a nominal chlorophyll *a* of the 5$^{th}$ percentile of observations in those months. In the following text, we assess each of the four statements in Section 2.

Statement 2.1. "*Based on current available state of the environment (SoE) data of 196 sites from 6 regions, most catchments across all regions are already meeting the current periphyton national bottom line (Figure 1)*".

The dataset of chlorophyll *a* from 194 sites shows that 182 sites in the dataset met the current periphyton bottom line. The statement applies only to the dataset, in which six regions in New Zealand were represented. There is no rationale provided for the extrapolation in Statement 2.1 from sites in 6 regions to *"most catchments across all regions".* A predictive model is required to make prediction about periphyton in most catchments across all regions.

Statement 2.2. "*The relationship between measured periphyton biomass and measured DIN for this analysis is poor (regression explaining <2% of the observed variation)*".

The statement is correct based on the data in Figure 1 of DairyNZ (2019). The weak relationship is not unexpected given that multiple environmental factors influence periphyton biomass. Figure A below shows the same data after log-transformation, to evaluate the linear relationship between periphyton biomass and DIN. This relationship is stronger than for the untransformed data ($R^2$ = 0.176) shown in Figure 1 of DairyNZ (2019), but there is still substantial unexplained variability.

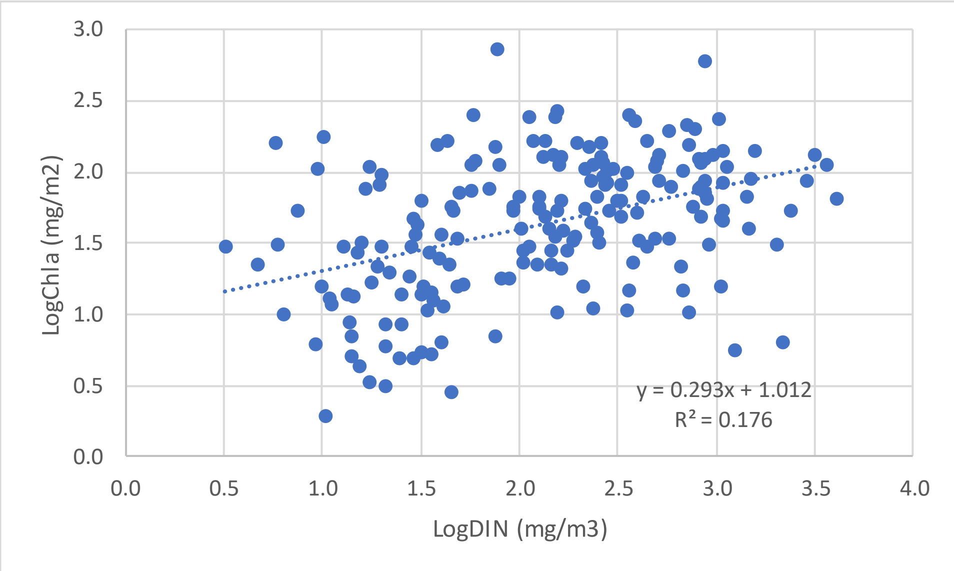


**Figure A.** The data in Figure 1 of DairyNZ (2019) replotted on log - log axes. After transformation, there is a weak linear relationship between DIN concentration and 92nd percentile chlorophyll *a*.

Statement 2.3. *"Only 6% (12 of 194) sites exceed the periphyton national bottom-line"*.

The statement is correct based on the data in Figure 1. See response above to Statement 2.1.

Statement 2.4. *"The monitored data indicates that DIN concentrations of <1 g/m3 are not necessarily required to meet the periphyton bottom-line".*

This statement only applies to the data shown in Figure 1. There is no rationale given for extrapolating beyond that dataset, or for making predictions about any sites not included in the regional council dataset. A predictive model will be required to do that.

### 3.2. Section 3: DairyNZ review of measured versus modelled periphyton biomass.

Statement 3.1. *"The assessment used by MfE to determine exceedance of the existing periphyton national bottom line and nutrient reduction targets is based on an existing NRWQN periphyton-nutrient model (refer to method in Snelder 2018, Snelder et al. 2019, but recalibrated in MfE 2019). This model was initially developed to quantify uncertainties, not to set regulation".*

The first part of the first sentence is correct: the procedure used to estimate nutrient reduction requirements was based on the NRWQN model. However,"… recalibrated in MfE 2019" is not correct. The outputs from the model (the targets derived for river segments in21 REC classes) were recalibrated, not the model itself (see Section 2.1, Part 5 above).

To clarify the second sentence in Statement 3.1: the NRWQN model was initially developed and reported in Appendix B of Larned et al. (2015). The intention was to use the NRWQN model as a first attempt to derive nutrient concentration thresholds specific to different classes of the REC. Larned et al. (2015) discussed deficiencies of the model and presented tables of targets as a "provisional guide". The NRWQN model was applied in the analyses reported in Snelder (2018) and Snelder et al. (2019), followed by a series of steps (See Section 2.1 above) to enable quantification of uncertainty around the thresholds, including multiple spatial exceedance criteria.

Statement 3.2. "*This work (i.e. model and derived TN targets) represents a "data mining" approach which required 'pushing' the NRWQN data beyond what it was collected for".*
It is not clear what "data mining approach" refers to, and whether using a data mining approach is a problem. At the time the NRWQN model was developed, the NRWQN dataset was the only one available for trialling such models. Given the limitations of the dataset (as acknowledged in Larned et al. (2015)), model fit was reasonable. Note, that most NRWQN sites are on large rivers, whereas the 196 regional council sites used to test the NRWQN model comprise a wide range of river sizes, including small streams. This and other differences between the NRWQN and RC datasets may have contributed to the poor performance of the NRWQN model across the test sites. Converting periphyton cover data from NRWQN data to biomass (as chlorophyll *a*) also increased uncertainty.

Statement 3.3*. "The approach was not recommended to define default thresholds in regulation, as the model was not able to adequately predict periphyton biomass at sites in a regional test dataset (Larned et al. 2015)".*
The reference to Larned et al. (2015) in this statement is misleading, as the statement does not appear in the 2015 report. The source of the recommendation about default thresholds in Statement 3.3 is not identified, but it was not Larned et al. (2015). Regardless of the source, Statement 3.3 is not logical as written. An 'approach' may be suitable, while the output from a particular model run using that approach proves to be unsuitable for use in regulation. That does

not mean that the approach should not be recommended. Further, there is no indication in this statement of how adequacy is to be determined. We note that models with various levels of complexity and model output at various levels of uncertainty have different roles for policy development and guidance. In the case of the NRWQN model, Larned et al. (2015) stated that the output "*should be treated with caution and only be used as a provisional guide*", but the model itself served as an example of the regression-based approach for developing nutrient criteria in the draft technical guidance (MfE 2018). As noted above, the limitations of the NRWQN model were summarised in Larned et al. (2015). However, the general approach used in MfE (2019) to identify concentration targets for achieving periphyton objectives is appropriate. Improvements in data quality and spatial coverage may be more useful than a major change in the modelling approach.

Statement 3.4. *"The most relevant test to assess the performance of the underlying periphyton model is comparing predicted vs measured periphyton biomass (as described in Larned et al. 2015)".*

Agreed. Comparing observed data from independent sites (i.e., not at sites used to develop a model) versus predictions for the same sites using the model would provide a robust test of the model.

*Statement 3.5. "A DairyNZ assessment of the dataset provided by Dr Ton Snelder shows that:*

> *I. The underlying NRWQN model is not able to predict periphyton biomass in the test dataset comprising >150 subcatchment sites across 6 regions. (Figure 2).*
>
> *II. Modelled data explained only 8% of the variation in observed periphyton biomass in the test dataset. The NRWQN periphyton model predicts 97 of 156 sites (i.e. 62%) of sites exceed the national bottom-line 92nd percentile threshold of 200 mg/m2 (i.e. D-band); compared with only 13 of 156 (around 8%) sites classified as D-band using observed (measured) data.*
>
> *III. The NRWQN model significantly over-predicts the proportion of sites that exceed the periphyton bottom line (by a factor of around 8)".*

Regarding Statement 3.5 (I): The predicted chlorophyll-*a* values in Figure 2 of DairyNZ (2019) are not the predicted value at each site based on substituting site-specific values of seven environmental variables into the NRWQN model and calculating chlorophyll *a*. Therefore, Statement 3.5 (I) "*The underlying NRWQN model is not able to predict periphyton biomass in the test dataset…"* and the caption of Figure 2 "*Measured (observed) vs predicted (modelled) periphyton biomass (as Chl-a) for sites within the test dataset (157 sites, 6 regions*" are both misleading as they imply that the predictions come directly from the NRWQN model.

Instead, Figure 2 shows data from Figure 7 (left hand panel, middle row) of Snelder et al. (2019). In turn, Figure 7 in Snelder et al. (2019) does not show predictions from the NRWQN model, it shows the degree to which the TN targets derived from the model resulted in predicted spatial exceedances. The "predictions" in Figure 7 of Snelder et al. (2019) are chlorophyll *a* values interpolated from the relationship between chlorophyll *a* at five levels (50, 120, 200, 400, 800 mg chlorophyll *a* $m^{-2}$) and the estimated TN targets, in the REC class relevant to each site, required to achieve each chlorophyll *a* level, with the specified spatial exceedance (in this case, 20%). The TN targets were estimated from thousands of predictions from the NRWQN model in each REC class. There was low variability around the targets in some classes and high variability in others (CXL).

The actual procedure that DairyNZ used to calculate the 'predicted values' in Figure 2 is identified in Footnote 4 of DairyNZ (2019), but that may not stop readers from presuming that the predictions are directly from the NRWQN model, as suggested by Statement 3.5(I).

Furthermore, we disagree with the statement in Footnote 4: "*However, as this calibration involved the test data set, the data shown in Figure 2 is the most relevant assessment of the ability of the NRWQN model to predict meaningful periphyton biomass values at regional sites in the test dataset*". The most relevant assessment of the ability of the NRWQN model to predict periphyton is a plot of observed values versus predictions directly from the NRWQN model.

Regarding Statement 3.5 (II): The numbers are correct based on Figure 2, but Figure 2 does not accurately reflect the actual performance and purpose of the MfE (2019) model, as discussed above.

Regarding Statement 3.5(III): Again, Figure 2 does not accurately reflect the actual performance and purpose of the MfE (2019) model, as discussed above.

Statement 3.6*. The poor performance of the underlying NRWQN 'nutrient-WCC' model to predict periphyton biomass at sites in the regional dataset is not unexpected given that it is the same model reported in Larned et al. (2015, Appendix B) where the authors identified the following significant limitations (verbatim from Larned et al. 2015):*

> *I. "The NRWQN models were not able to predict the pattern of the observed 92nd percentile values of WCC at the Canterbury and the Whanganui-Manawatu sites."*
>
> *II. "There were some significant limitations associated with this analysis and the derived nutrient criteria should be treated with caution and only be used as a provisional guide."*
>
> *III. "The NRWQN data was not collected with the intention of deriving nutrient concentration criteria and this limitation probably contributes to the large uncertainties in the derived criteria."*
>
> *IV. "Finally, our regression models that underlie the method did not explain a lot of variation in periphyton abundance and predictions made using these models have high uncertainty."*

Agreed. Larned et al. (2015) was completely transparent regarding the limitations of the model.

## 3.3. DairyNZ Section 5: Review the accuracy of the statements in Section 5: summary of aligned positions.

Statement 5.1. *Greater emphasis should be placed on attributes that represent ecosystem outcomes (i.e. periphyton, macroinvertebrate) rather than attributes based on 'drivers' of these outcomes (e.g. sediment, nutrients).*

Response. The context for this statement is not clear. Many of the current NOF attributes act as both driver and response variables, and there are multiple feedbacks between them (e.g., TN affects lake cyanobacteria, which affects TN). The need to categorise attributes exclusively as drivers or outcomes is not clear. This statement may refer to Section 4 of DairyNZ (2019), which is out of scope.

Statement 5.2. *National 'one-size' fits all nutrient thresholds to 'protect' ecosystem health are not robust (i.e. based on overly simplistic relationships that do not account for the complexities of improving ecosystem health) and are unlikely to result in the improved outcomes sought by communities*.

Response. It is not clear whether this statement is referring to the national DIN and DRP bottom lines proposed in the Essential Freshwater package, or to the TN targets set out in MfE (2019). If the former, we agree that ecological responses to national-scale nutrient bottom line DIN and DRP concentrations are likely to vary widely across sites. If the latter, the TN targets are not 'one size-fits-all'; there are 21 TN targets corresponding to river classes.

Statement 5.3. *Periphyton is best managed at a regional level, consistent with NPS-FW 2017 periphyton attribute and draft technical guidance document (MfE 2018).*

Response. No information is provided in DairyNZ (2019) to support this statement, and it is not clear what "*best managed"* means in this sentence. Note that neither the NPS-FM nor the draft technical guidance state that "*periphyton is best managed at a regional level*". The NPS-FM requires councils to determine instream nitrogen and phosphorus criteria for periphyton management in their FMUs. That requirements does not preclude the use of national nutrient criteria per se, or the results of down-scaled national models. Several such models are summarised in the draft technical guidance, with recommendations for using them to set nutrient criteria. The draft technical guidance then "*shows process steps recommended <u>if development of a regional model is considered the best option</u> for deriving robust nutrient criteria*" (MfE 2018).

Statement 5.4. *Nutrients are an important driver of excessive periphyton growth but other factors like stream bed shading may be more relevant for managing excessive biomass, particularly in small streams in intensively developed catchments.*

Response. Agreed, there are a large number of environmental factors that affect periphyton growth and abundance (e.g., light, water temperature, herbivory, hydraulic conditions, nutrient supplies, pH). However, opportunities to use each of these factors to manage periphyton are limited, including stream shading. At catchment scales and larger, and in segments too wide for effective canopy shading, nutrient control may be the most practical management tool.

Statement 5.5. *The periphyton N (or phosphorus) targets should not be used as default thresholds (as recommended by the STAG), however, they may provide useful interim guidance for councils in setting instream nutrient criteria*.

Response. It is not clear what makes a nutrient target unsuitable for default thresholds, but suitable for interim guidance. This statement may be missing some information.

Statement 5.6. *The underlying NRWQN periphyton model pushes data beyond what it was originally intended for, and is not able to predict site-scale periphyton biomass. The N targets used to define the nutrient baseline therefore have very high uncertainty*.

Response. See the response to Statement 3.2 above.

Statement 5.7. *The spatial exceedance criterion concept is inconsistent with NPS-FM framework. That is, NPS-FM does not allow 10 or 20% (or even 50%) of sites to exceed bottom-lines. The expectation is that councils will monitor at representative and meaningful locations (e.g. sites that are sensitive to high periphyton biomass).*

Response. Agreed, spatial exceedance criteria for periphyton are not requirements in the NPS-FM 2017 or in the Essential Freshwater package. See the preceding assessment of MfE (2019) Section 2.3 above. However, it is not clear from this statement how using or not using spatial exceedance criteria is related to councils monitoring periphyton. Periphyton monitoring is required to evaluate compliance regardless of spatial exceedance criteria.

Statement 5.8 *Some catchments/FMUs are over-allocated with respect to periphyton; monitored data from 6 region indicates that approximately 6% (i.e. 12 of 194) of sites currently exceed the periphyton bottom-line, and may require nutrient reductions*.

Response. Based on the data plotted in Figure 1, we agree that 12 of the 194 data points exceeded the 200 mg chl *a* $m^{-2}$ bottom line during the periods that the corresponding sites were monitored. This statement only applies to the sites in Figure 1.

Statement 5.9 *The regional test data set is biased towards impact (i.e. pastoral) catchments*.

Response. Agreed. The proportion of sites in the regional council periphyton dataset classified as pastoral landcover in the REC (64%) exceeds the proportion of river reaches in New Zealand classed as pastoral (42%).

Statement 5.10 *There is generally high uncertainty in national-scale periphyton-nutrient models, but much stronger regressions (and lower uncertainties) are possible with regional datasets. For example, Kilroy et al. (2019) reported 4-models with coefficient of determination values ranging from 0.74 to 0.87. This supports the Ministry's guidance (MfE 2018) for councils to set regionally-derived instream nutrient criteria to meet (provide for) periphyton objectives as part of the amended NPS-FM (2017).*

Response. The NRWQN model was weaker than regional-scale models developed for some but not all of the regions in Kilroy et al. (2019). Note that Kilroy et al. (2019) is not included in the reference list in DairyNZ (2019). The report that is cited, Kilroy (2019), describes relatively strong periphyton – environment models that were developed for a subset of river sites in the Horizons region. In contrast, Kilroy et al. (2019) has individual regional models for all six regions represented in the test dataset; three of these models performed better than the NRWQN model, and three performed worse. See Kilroy et al. (2019) for details.

Statement 5.11 *The NRWQN model is not able to predict periphyton biomass at individual sites – reflected in the poor regression between observed and predicted periphyton biomass (92nd percentile) for the 6-region test dataset.*

Response. Agreed.

Statement 5.12. *In some catchments where current state is exceeding the periphyton bottom line, the DairyNZ baseline may underestimate the extent of nutrient reductions required by the existing NPS-FM*.

Response. More information is needed to assess this statement than is provided in the report. DairyNZ (2019) refers to five different baselines (periphyton baseline, NPS-FM baseline, nutrient baseline, MfE baseline, DairyNZ baseline). These may be different baselines or some or all may be synonyms. The term baseline is not defined. Statement 1.4 of the report: "DairyNZ used the current water quality state as the baseline…" suggests that 'DairyNZ baseline' may refer to current river nitrogen or phosphorus concentrations. However, it is unclear how current river concentrations alone, or in combination with other data, are intended to estimate required nutrient reductions.

Statement 5.13 *The proposed DIN and DRP thresholds are based on overly simplistic relationships, have significant flaws and are unlikely to result in improved ecosystem health outcomes sought by communities across many catchments. Attempts to define a set of 'global' nutrient bottom-lines that are 'harmonised' to measurable ecosystem health outcomes (e.g. macroinvertebrate community health) are inconsistent with current scientific knowledge about multiple drivers influencing, and the inherent complexities of managing, ecosystem health*.

Response. This statement appears to refer to Section 4 in DairyNZ (2019) and is out of scope.

## 4. References

Dodds WK, Smith VH, Lohman K. 2002. Nitrogen and phosphorus relationships to benthic algal biomass in temperate streams. Canadian Journal of Fisheries and Aquatic Sciences 59: 865-874.

Elliott A, Snelder T, Muirhead R, Monaghan R, Whitehead A, Bermeo-Alvear S, Howarth C. 2019. A heuristic method for determining changes of source loads to comply with water quality limits in catchments. Environmental Management. DOI: 10.1007/s00267-019-01235.

Glibert P, Madden C, Dettmann E, Boynton W, Heil C, Nelson W, Lehrter J, Latimer J, Flemer D, Kennish M, Brown C. 2010. A framework for developing nutrient criteria. Glibert P., Madden C, Boynton W, Flemer D. (eds). Nutrients in estuaries, pp 43-71. US Environmental Protection Agency.

Grizzetti B, Bouraoui F, De Marsily G. 2005. Modelling nitrogen pressure in river basins: a comparison between a statistical approach and the physically-based SWAT model. Physics and Chemistry of the Earth 30: 508-517.

Irby ID, Friedrichs MA. 2019. Evaluating confidence in the impact of regulatory nutrient reduction on Chesapeake Bay water quality. Estuaries and Coasts 42: 16-32.

Kilroy C. 2019. Using empirical relationships to develop nutrient targets for periphyton management: a case study from the Horizons region. Horizons Regional Council. NIWA Client Report 2019092CH.

Kilroy C, Whitehead A, Howard S, Greenwood M. 2019. Modelling periphyton in New Zealand rivers. 1. An analysis of current data and development of national predictions. Ministry for the Environment. NIWA client report 2019002CH.

Larned ST, Snelder T, Unwin M, McBride G. Verburg P, McMillan H. 2015. Analysis of Water Quality in New Zealand Lakes and Rivers. Ministry for the Environment. NIWA Client Report CHC2015-033.

Leathwick J, West D, Gerbeaux P, Kelly D, Robertson H, Brown D, Chadderton W, Ausseil A. 2010. Freshwater ecosystems of New Zealand geodatabase. Department of Conservation.

Leathwick J, Snelder T, Chadderton W, Elith J, Julian K, Ferrier S. 2011. Use of generalised dissimilarity modelling to improve the biological discrimination of river and stream classifications. Freshwater Biology 56: 21-38.

Ministry for the Environment (MfE) 2018. A draft technical guide to the Periphyton Attribute Note. (https://www.mfe.govt.nz/sites/default/files/media/Fresh%20water/Periphyton%20note%20draft%20technical%20guidance%20_FINAL.pdf).

Mohamoud Y, Zhang H. 2018. Applications of linked and nonlinked complex models for TMDL development: approaches and challenges. Journal of Hydrologic Engineering 24: p.04018055.

Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50: 885-900.

Moriasi DN, Gita, MW, Pai N, Daggupati P. 2015. Hydrologic and water quality models: performance measures and evaluation criteria. Transactions of the ASABE 58: 1763-1785.

Snelder TH, Booker DJ, Quinn M, Kilroy C. 2014. Predicting periphyton cover frequency distributions across New Zealand's rivers. Journal of the American Water Resources Association 50: 111-127.

Snelder TH, Larned ST, McDowell RW. 2018. Anthropogenic increases of catchment nitrogen and phosphorus loads in New Zealand. New Zealand Journal of Marine and Freshwater Research 52: 336–361.

Snelder TH, Moore C, Kilroy C. 2019. Nutrient concentration targets to achieve periphyton biomass objectives incorporating uncertainties. Journal of the American Water Resources Association. DOI: org/10.1111/1752-1688.12794.

Tango PJ., Batiuk RA. 2013. Deriving Chesapeake Bay water quality standards. Journal of the American Water Resources Association 49: 1007-1024.

Whitehead A. 2018. Spatial modelling of river water-quality state, incorporating monitoring data from 2013 to 2017. Ministry for the Environment. NIWA Client Report 2018360CH.

Zhang Q, Tango PJ, Murphy RR, Forsyth M., Tian R, Keisman J, Trentacoste EM. 2018. Chesapeake Bay dissolved oxygen criterion attainment deficit: three decades of temporal and spatial patterns. Frontiers in Marine Science 5: 422.