Review of 'Strategic assessment of New Zealand's freshwaters from a human health perspective' by Joanne Clapcott and Roger Young.

Our understanding of the report:

This report describes statistical modelling of *E. coli*, water quality variables and cyanobacterial biovolumes in relation to human health in New Zealand waterbodies. A robust *E.coli* dataset from 753 stream and river sites has been examined in relation to site flow and month of year at time of sample collection. Based on a lack of relationship with these variables, *E. coli* metrics were calculated from annual datasets and regressed against environmental variables using a flexible model technique to predict *E. coli* measures for all rivers and streams in New Zealand. Modelled *E. coli* measures include the median and 95th percentile *E. coli* concentration and the percentage of samples that exceeded 260 and 540 *E.coli* 100 mL^{-1}. Model performance statistics showed satisfactory to good model performance and model predictions showed broad spatial patterns in *E.coli* metrics. The proportion of river sites within NOF bands are not reported. The representativeness of sample sites in relation to all sites is explored.

Ten water quality measures were collated for 99 lakes and regressed against environmental variables to predict water quality in all New Zealand lakes greater than 1 hectare. The relationship between water quality and cyanobacterial biovolume observed at 37 lakes was used to predict cyanobacterial biovolume for all New Zealand lakes greater than 1 hectare. Based on the predictions, the proportion of lakes that meet the NPS-FM attribute bands for human health for recreation are reported.

Our critique and suggestions for content changes (order does not reflect importance):

1. Suggest adding 'for recreation' to the title for consistency with the NPS-FM, e.g. 'Strategic assessment of New Zealand's freshwaters from a human health for recreation perspective'
2. Further rationale on why the relationship between *E.coli* in rivers and flow and sample month were explored is required. It seems to be to justify the inclusion of annual datasets to calculate *E.coli* metrics. The lack of a relationship with flow is important to discuss because there are references to show that flow does indeed affect *E.coli* concentrations, e.g. Davies-Colley 2013, Wilkinson et al 2011. Perhaps the fact that a lot of samples were collected at base flow following regional council protocols may help inform this discussion. An analysis of *E. coli* associated with increasing versus decreasing flow is likely to produce a better relationship with flow. On P9, what proportion of the sites had measured compared to predicted flow measurements?
3. Also, it is not clear what statistical analysis was used to show there were no significant monthly trends in *E. coli* data? A visual inspection suggests strong sigmoidal patterns for many river classes.
4. The data and statistical approach used are robust and well described. Although we do suggest clarification of some aspects. The log-transformations used result in uneven error dependent on the predicted value. It would be useful to illustrate this for the reader.

Perhaps with two graphs, one showing non-transformed error and how this is consistent across the prediction gradient, and a second graph showing the back-transformed predictions and error around predictions. If error is asymmetric for transformed models, an additional column in Tables showing the error central tendency in original units may further enlighten the reader.

5. Greater discussion of model error is required. Currently the use of terms like 'reliable' and 'good' model performance give the impression that these models are correct and not simply the best approximation. For example, when error is taken into account what are the likelihoods of incorrect assignment to a NPS-FM band? Section 6.2 states the usefulness of these models, but could be balanced with an outline of the issues associated with them. For example, high levels of *E. coli* predicted on the West Coast are likely due to environmental gradient present rather than actual high values, as outlined in 6.3 for lakes
6. It would be beneficial to report the proportion of river length that falls within NPS-FM *E.coli* bands for rivers, as is reported for cyanobacterial bands for lakes.
7. No point sources are included as predictors in models of *E. coli* in rivers or water quality in lakes. The absence of predictors which account for known sources of faecal contamination need to be addressed; perhaps when discussing model limitations. This may be a particular concern in the future if the model is used to assess the effectiveness of response actions – which might include better treatment of point source discharges.
8. Why were *E. coli* data for lakes not explored? This seems to be a major omission. If this is an incomplete assessment of human health for recreation then there needs to be acknowledgement of the fact.
9. Further rationale on why water quality variables were explored for human health in lakes is required. If it was purely to link to cyanobacterial biovolume then why not regress biovolume with water quality first to determine primary water quality predictors and then only model those predictors, e.g. TP, Chl-a, Secchi. If for another reason, it is not clear why. Also, the selection of lakes with > 7 samples seems subjective. If so, a statement along the lines of 'a need to maximize dataset size while minimize error due to unrepresentativeness' is required. It is isn't immediately clear that sites > 12 samples were used for spatial modelling of water quality variables, but sites > 7 samples were used to regress cyanobacterial biovolume with water quality. A better description of how the regression was used to inform cyanobacterial biovolumes is needed. This wasn't modelled, it was estimated based on a regression etc.
10. Further discussion of the partial plots would be beneficial. For example, does it makes sense that *E. coli* would be greater in stable flows (i.e. low rainfall variability in Figure 8)?
11. On P11 there is a sentence 'Second, less than 50% of the values for a variable were censored' – what does this mean?
12. On P20 it is not clear what the following sentence means, perhaps reword? ' ...than the mean of the observed data respectively'.
13. The last paragraph in section 6.5 requires some thought. The message is not clear. The regression equation used to inform national scale predictions was biased towards summer values and hence may under/overestimate cyanobacterial biovolumes?

Our suggestions for minor editorial issues:

1. There are two Table 3s. There is also inconsistent formatting of Table and Figure headers.
2. Table 2 description of variables do not line up with abbreviations.
3. There are numerous spelling errors and inconsistencies in usage; check in particular for 'health'[including in the title], 'cyanobacterial', 'chl-a' and all of its alternatives, 'R^2' and 'P' formatting, 'et al.' formatting, 'source-of-flow' consistency, 'DOBottom', 'log_{10}-transformed'
4. Additional errors include:

a. PviL4 – delete 'as'
b. PviL22 – 'fourth' not 'forth'
c. PviL39 – 'to estimate cyanobacterial biovolume for all New Zealand lakes'.
d. PviiL35 – 'scale patterns of cyanobacterial biovolume'
e. P8L6 – 'Larned et al. (2015) provided…'
f. P9L8 – delete 'defines'
g. P9L12 – delete 'data'
h. P9L13 – delete 'until'
i. P9L22 – 'analyses'
j. P9L33 – 'recorder and provided by the…'
k. P13L8 – 'Secchi'
l. P15L2 – '(REC; Snelder and Biggs, 2002).'
m. P16L11 – 'predictions for all lakes,'
n. P18L8 – delete 'and'
o. P18L19 – 'detection limits, half the'
p. P19L24 – unseen data (Breiman, 2001)).
q. P20L20 – delete 'for each'
r. P28L2 – 'in Figure 9'?
s. P31L3 'by Larned et al. (2015).'
t. P35Fig13 – delete 'NO3N'
u. P44L15 – 'temperate lakes'
v. P44L19 – 'associated with'
w. P44L37 – 'For the reasons discussed'
x. P44L42 – 'programmes do not include'
y. P45L4 – 'related to the planktonic cyanobacteria attributes defined by the NPS-FM.'
z. P35L15 – 'the relationship was established during periods'
aa. P35L18 – 'datasets are'
bb. P46L6 – 'cell concentration data.'

5. On PviiL37, suggest you delete 'reliable'. As 'reliability' infers independent validation over time.
6. P12L2, suggest you delete 'when consumed or by contact'
7. P12L3, suggest you add 'for lakes' e.g. '…NPS-FM attribute for lakes based on…'
8. P22L2 – suggest you include all classes in parentheses and order as shown in Figure 4, e.g. (CD/L, CW/L, CX/L, WD/L, WW/L, WX/L). Correct following sentences too.
9. P23Fig5 – 'and slope values for regression between *E.coli* and flow for each REC source-of-flow class'
10. P23L12 – 'The analysis revealed that exceedances associated with a threshold of 260 *E.coli* 100mL^{-1} tended to occur…'
11. P31Fig11 – the panels are not ordered or labelled as they are referred to in the text on this page or the accompanying figure header.
12. P32Table7 – does not include DOBottom. See also previous comment about this section.
13. P34L2 – suggest edit ' that had satisfactory to good performance'
14. P36L3 – Delete sentence starting 'Figure 14 indicates…' this is a repeat of previous sentence
15. P43L3 – suggest 'are likely included in the models'
16. P44L30 – reword sentence starting 'When potentially toxic'. It is hard to follow.
17. References: Check Bryers for a better HTTP link. Larned et al 2016 is incomplete.
18. Appendices: The second table needs a header. Perhaps use Table A1 and Table A2.

References:

Davies-Colley JR 2013. River water quality in New Zealand: an introduction and overview. Ecosystem services in New Zealand: conditions and trends. Manaaki Whenua Press, Lincoln, pp.432-447.

Wilkinson RC, McKergow LA, Davies-Colley RJ, Ballantine DJ, Young RG 2011. Modelling storm-event E.coli pulses from the Motueka and Sherry Rivers in the South Island, New Zealand. New Zealand Journal of Marine and Freshwater Research 45: 369–393.