IN THE MATTER of the Resource Management Act 1991

AND

IN THE MATTER of Water quality and stormwater – Hearing 4

AND

IN THE MATTER of the Proposed Natural Resources Plan for the Wellington Region

---

EXPERT CONFERENCING JOINT WITNESS STATEMENT TO THE HEARING PANEL

TOPIC: WATER QUALITY TRENDS

DATE: 19 FEBRUARY 2018

INTRODUCTION

1. This joint witness statement (JWS) was written following discussion between the experts to clarify for the hearings panel:
   - the issues/matters on which the expert witnesses agree;
   - the issues/matters on which they do not agree, and the reasons for their disagreement.

2. This joint witness statement relates to the conferencing topic of **water quality trends**.

3. A conferencing discussion was held on **12 February 2018** via teleconference. This witness statement was then circulated, refined and agreed by email between the participants.

4. Participants were:
   - Kate McArthur – Practice Leader Water, The Catalyst Group (for the Minister of Conservation and Rangitāne)
   - Dr Ton Snelder – Director, Land Water People (for GWRC)
   - Dr Adam Canning – Research Scientist, (for Wellington Fish and Game Council)

5. In preparing this statement, the experts have read and understood the Code of Conduct for Expert Witnesses as included in the Environment Court of New Zealand Practice Note 2014.

AREAS OF AGREEMENT / DISAGREEMENT

6. The experts agree that, based on Dr Snelder's analyses and further analyses (Appendix 1), that the rate of change in water quality is not rapid or ubiquitous over the ten-year period analysed, across the Wellington Region.

7. The experts agree that there is no evidence of *region-wide* degradation over the ten-year or five-year time periods tested by Dr Snelder.

8. The experts agree that, based on Dr Snelder's analyses, some physical, chemical and biological metrics at individual sites are improving, whilst other have uncertain trends overall across the region. There are individual sites where some metrics are degrading (Appendix 2).

9. Dr Snelder's high-level conclusion is: there is strong evidence of water quality improvement across the region over the past decade. Water quality has not improved everywhere, but degradation is isolated rather than occurring in a consistent and regional scale manner.

10. However, Dr Canning and Ms McArthur have identified limitations and have reservations about the applicability of this statement (paragraph 9 above) for management decision-making for the following reasons:

    a. Overall 'water quality' does not have a defined meaning – it is made up of a range of biophysical attributes which may behave differently to variations in natural and human-induced stressors;
    b. The sites may not truly represent the region;
    c. Some water quality attributes (e.g. turbidity and clarity or various analytes of nutrients) are highly correlated with each other, therefore including trends for all attributes in overall analysis may be considered 'double counting' and contribute to false discovery of trends;
    d. False discovery and pseudo replication may bias region-wide results;
    e. Dr Snelder's analyses included trends with low-confidence;
    f. Water quality trends can be influenced by natural variability and climate effects; and
    g. The state of water quality is also a key factor in deciding management actions, not just trend.

11. In discussing the limitations listed above, the experts agree that additional information may be useful to the hearing panel to clarify the issue of water quality trends for the purposes of decision-making and management responses. Their joint responses to the issues raised above are included in the following paragraphs.

Representativeness of sites

12. The experts agree that the sites used may not be adequately representative of the Wellington Region. Sites used in Dr Snelder's analyses are from the Wellington Region's SOE monitoring programme. Such programmes monitor water quality for multiple purposes. By their nature, SOE sites are not randomly selected. This introduces potential for underlying assumptions not to be met in statistical tests (e.g., binomial testing) or in assumptions on equal distribution of false positive or negative trends; in essence, this could provide a misleading representation of regionwide trends.

Correlated water quality attributes

13. The experts agree that there may be some influence on the overall water quality trend as a result of including all (sometimes correlated) water quality metrics in the analyses. This could have the effect of increasing the number of improving trends. This could be avoided by identifying primary attributes for aspects of water quality and only using those in the analyses, although this would mean less data was available for use. In Dr Snelder's opinion, this is unlikely to materially change the overall result.

14. Dr Snelder's opinion is based on there being a similar number of physical and chemical variables (which generally have the greatest proportion of improving trends) as biotic variables (invertebrates and periphyton, which have the greatest proportion of degrading trends) and variables within these two sets of variables (physical/chemical and biological) exhibit correlation.

Pseudo-replication and false discovery

15. Some rivers have multiple sample sites, where others only have one. The experts agree that this can result in pseudo-replication that may allow one river with multiple sites, each having the same trend direction, dominating the overall trend direction. In discussing the potential of pseudo-replication and false discovery with Dr Canning, Dr Snelder undertook further analysis (attached as Appendix 1) whereby trends were assessed using only one randomly selected site from rivers that had multiple sites. The sensitivity analysis supports the robustness of Dr Snelder's conclusions.

16. The experts agree that when multiple hypotheses are tested, the probability of getting a "false positive" or "false negative" increases. In Dr Snelder's further analysis, he corrected the test statistic p-values for false discovery. All designations of "regional trends" remained the same, except nitrate-nitrite-nitrogen (NNN) and total nitrogen over the five-year period changed from "improving" to "not significant." No degrading trends were observed across the entire region.

The inclusion of low-confidence trends

17. Dr Snelder's analysis includes trends that have low-confidence. The analysis assumes that false negative trends will be offset by false positive trends, this can affect the confidence of the overall results. Typically, as sample size (number of sites) increases, the confidence in overall findings also increases. It is uncertain whether the region contains sufficient sites for this assumption to have little influence on the confidence of the overall results. No further analyses has been undertaken to assess this.

Influence of natural variability

18. The experts agree that understanding water quality trends and the causes of them, requires scientists to be aware of potential trends resulting from wider climate patterns, such as those identified by Scarsbrook et al. (2003)[1]. Climate trends may explain why five-year trends do not necessarily correlate with ten-year trends.

19. The experts agree they cannot attribute the trends in water quality for the Wellington Region to particular causes, such as resource management decisions or climate, as this has not been specifically tested within the experimental design used by Dr Snelder.

Importance of water quality state

20. The experts agree that water quality state (current condition versus the objective) must also be considered alongside trends in water quality when determining management actions or the need for them.

Water quality degradation

21. The experts agree that some attributes of water quality are degrading at some sites. A summary of water quality trend direction, including degradation was compiled by Dr Snelder and is included as Appendix 2 to the JWS.

22. Filamentous periphyton, MCI and QMCI have degraded at more sites than they have improved over the five-year time-period although there are not significant regional trends for these variables.

23. Some sites have several water quality variables with degrading trends that are inconsistent with the general regional pattern of water quality improvement. This may reflect the effects of specific activities in the upstream catchments of the sites.

24. The sites with more than five trends that were likely to be degrading included: Mangaone Stream at Sims Road Bridge, Otaki River at Pukehinau, Pakuratahi River 50m Below Farm Creek and Taueru River at Castlehill.

---

[1] Scarsbrook MR, McBride CG, McBride GB, Bryers GG. 2003. Effects of climate variability on rivers: Consequences for long-term water quality analysis. *Journal of the American Water Resources Association 39(6)*: 1435-1447.

25. Five-year flow adjusted trends suggest that the dissolved nitrogen variables NO_3-N and NNN were at least as likely as not to be degrading at 61% and 54% of sites.

26. This result is inconsistent with that of the raw trends for which NO_3-N and NNN were at least as likely as not to be improving at 66% and 69% of sites. However, the regional-scale trends for NO_3-N and NNN based on the five-year flow adjusted trends were not significant.

27. Differences between the results for five-year trends for dissolved nitrogen, as described above, creates increases the uncertainty around the conclusion that there is no evidence of region-wide degradation in dissolved nitrogen.

CONCLUSION

28. The experts agree that the additional analyses have increased the robustness of the findings and it is reiterated that the experts agree that there is no evidence of *region-wide* degradation over the ten-year or five-year time periods tested by Dr Snelder.

Signed on 19 February 2018

Kate McArthur

Ton Snelder

Adam Canning

APPENDIX 1. FURTHER ANALYSIS OF REGIONAL TREND DATA

Adjustment of binomial test p-values for multiple comparisons

Snelder (2017a) performed multiple binomial tests (each for a different water quality variable). When multiple hypothesis tests are performed, the possibility of a type one error (i.e. falsely rejecting the null hypothesis) is larger than indicated by the nominated alpha value (in this case 0.05) because the tests are performed repeatedly. Arguably, Snelder (2017a) should have corrected for "false discovery". This was not done because the binomial test could have resulted in both regional degrading as well as regional improving trends, so the analysis was even-handed. However, the analysis found only improving trends were significant and the absence of adjusting for false discovery may have over-stated the number of significant improving trends.

In this analysis, the significance levels for the individual binomial tests made by Snelder (2017a) were adjusted to account for multiple comparisons using the false discovery rate (FDR) adjustment method (Benjamini and Hochberg, 1995). If the adjusted binomial test p-value was less than 0.05, the null hypothesis was rejected, i.e., it was concluded that there were more trends in the region than could be expected by chance and that there was an 'regional-trend'. The regional-trend direction was determined as positive if the proportion of positive trends was greater than 50%, and negative if the reverse were true. The magnitude of the regional trend is quantified by the median of all RSS values.

Results

Ten-Year trends

For the ten-year period, there were significant improving regional-trends in Clar, TP, NO3-N, NNN, and TN and Chla (Table 1). All other regional-trends were not significant. After adjustment, the individually significant improving trend in Chla did not reach significance at 0.05 level but was significant when significance was relaxed to 0.1. There were no significant degrading regional-trends.

Table 1. Ten-year regional trends based on raw site trends.

Variable	No. sites	No. decreasing	No. increasing	Binomial p-value	Adjusted p-value	Regional trend	Regional trend magnitude (%)
Clar	52	5	47	0	0.000	Improving	3.86
Turb	56	35	21	0.081	0.182	Not Significant	-1.5
DRP	35	20	14	0.311	0.431	Not Significant	-0.2
TP	45	36	10	0	0.000	Improving	-2.09
NO_3-N	50	35	14	0.003	0.011	Improving	-1.57
NNN	55	40	14	0	0.000	Improving	-1.45
TN	40	36	5	0	0.000	Improving	-2.12
TOC	51	33	19	0.092	0.184	Not Significant	-0.38
E. coli	55	29	26	0.788	0.834	Not Significant	0
Chla	47	32	15	0.019	0.057	Not	-7.18

Variable	No. sites	No. decreasing	No. increasing	Binomial p-value	Adjusted p-value	Regional trend	Regional trend magnitude (%)
						significant	
Fils-Max	45	28	17	0.135	0.203	Not Significant	0
Fils-Mean	45	19	26	0.371	0.477	Not Significant	0
Mats-Max	45	27	17	0.135	0.203	Not Significant	0
Mats-Mean	45	24	21	0.766	0.834	Not Significant	0
%EPT	54	24	30	0.497	0.596	Not Significant	0.17
%EPT_Taxa	54	20	33	0.134	0.203	Not Significant	0.58
MCI	54	26	27	1	1.000	Not Significant	0.03
QMCI	54	33	20	0.076	0.182	Not Significant	-0.16

Five year trends

For the five-year period, there were significant improving regional-trends in in Clar, Turb and TOC (Table 2). All other regional-trends were not significant. After adjustment, the individually significant improving trend in *E. coli* did not reach significance at 0.05 level but was significant if the required significance was relaxed to 0.1. There were no significant degrading regional-trends.

Table 2. Five-year regional trends based on raw site trends.

Variable	No. sites	No. decreasing	No. increasing	Binomial *p*-value	Adjusted p-value	Regional trend	Regional trend magnitude (%)
Clar	51	0	51	0	0.000	Improving	9.05
Turb	55	44	11	0	0.000	Improving	-5.08
DRP	37	20	17	0.743	0.836	Not Significant	0
TP	44	28	17	0.174	0.313	Not Significant	-1.51
NO_3-N	50	31	19	0.119	0.268	Not Significant	-0.33
NNN	54	36	19	0.04	0.132	Not Significant	-0.73
TN	42	28	14	0.044	0.132	Not Significant	-1.44
TOC	52	50	1	0	0.000	Improving	-7
E. coli	55	38	18	0.014	0.063	Not Significant	-5.32
Chla	45	25	20	0.551	0.661	Not	-7.96

Variable	No. sites	No. decreasing	No. increasing	Binomial *p*-value	Adjusted p-value	Regional trend	Regional trend magnitude (%)
						Significant	
Fils-Max	43	18	26	0.222	0.363	Not Significant	0.99
Fils-Mean	43	19	25	0.36	0.498	Not Significant	0
Mats-Max	41	23	18	0.533	0.661	Not Significant	0
Mats-Mean	41	16	26	0.117	0.268	Not Significant	0
%EPT	41	20	20	1	1.000	Not Significant	0
%EPT_Taxa	42	21	20	0.878	0.930	Not Significant	-0.21
MCI	42	25	17	0.28	0.420	Not Significant	-1.05
QMCI	42	26	16	0.164	0.313	Not Significant	-0.57

Conclusions

The adjustment for false discovery rate does not change the conclusions of Snelder (2017a). The analyses of regional trends indicate significant regional improvement in several water quality variables over both time periods. There were improving regional-trends in Clarity, and nutrients (including TP, and nitrogen species) over the ten-year period and clarity, turbidity and TOC over the five-year period. If alpha is relaxed to 0.1, there was an improving regional-trend in the periphyton measure Chla over the ten-year period and an improving regional trend in *E. coli* over the five-year period. There were no significant regional degrading trends even when the required p-value was relaxed to 0.1.

Analysis of potential site pseudo replication

The GWRC monitoring network has catchments with multiple sites (Figure 1). Table 1 shows the catchments with multiple sites. Sites located in the same catchment are influenced by the same conditions and a component of the water at downstream sites is measured at the upstream sites. This raises the possibility that the results of the regional study by Snelder 2017a) are unduly influenced by pseudo replication.

Water quality monitoring sites

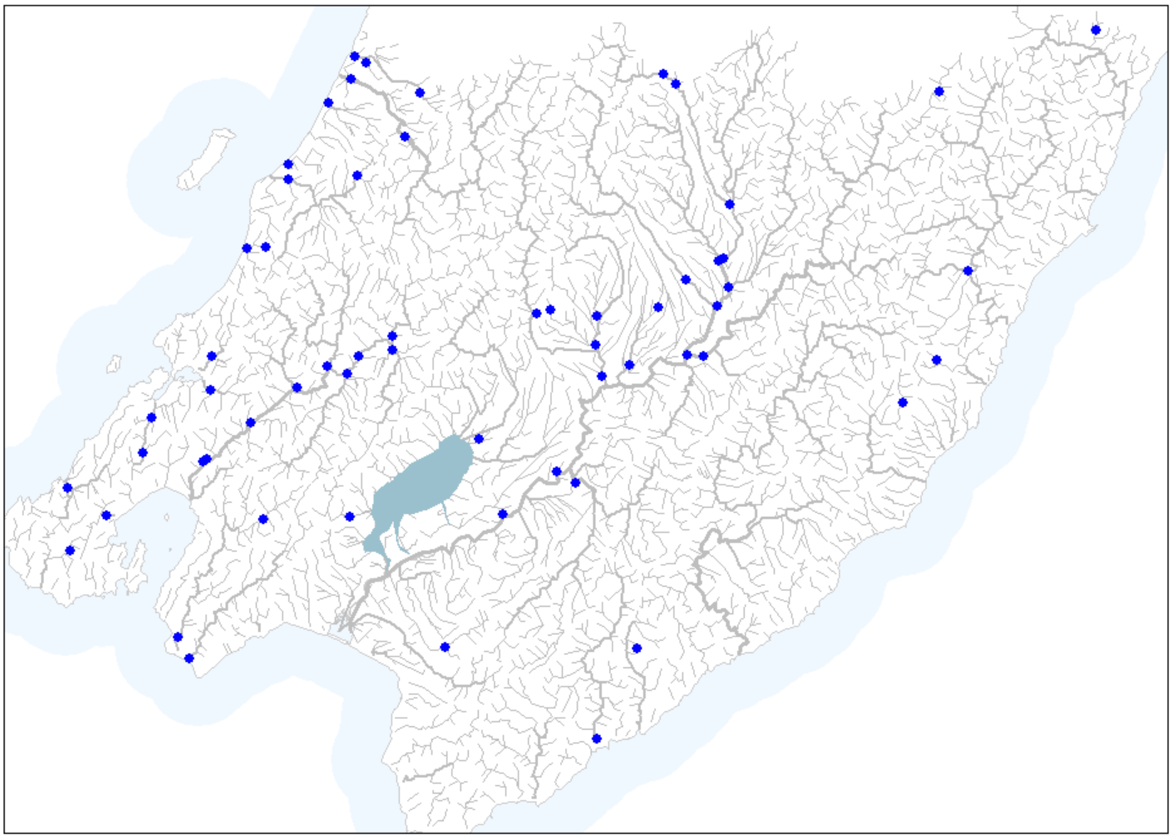

Figure 1. Water quality monitoring sites used by Snelder (2017a) showing river network of order 4 or greater. Note that some catchments have more than one site.

Table 3. Catchments in the water quality monitoring network with replicate sites (showing the number of replicate sites)

Catchment	Number of sites
Hutt	4
Otaki	2
Parkvale	2
Porirua	2
Ruamahanga	7
Taueru	2
Waikanae	2
Wainuiomata	2
Waiohine	2
Waitohu	2
Whareroa	2

A sensitivity analysis was carried out to assess the degree to which results may have been influenced by pseudo replication. The analysis was performed on the 10-year raw dataset. The analysis took multiple realisations of the data such that only one site was allowed to represent catchments with more than one site. In each realisation, the site representing each catchment was drawn at random from the data (Table 3). This was repeated for 100 realisations. The results were then summarised by evaluating the proportion of sites that were (1) at least as likely as not to be improving and (2) at least likely to be improving and comparing these results to those obtained for the full dataset.

Results

A summary over all variables of the results of the sensitivity analysis are shown in Table 4. The results by variable are shown in Table 5. The results indicate that very similar results are obtained for the sensitivity analysis as for the full dataset. This result is unsurprising because of the dominance of improving trend direction and the absence of relationships between trend direction and catchment characteristics reported by Snelder (2017b).

Table 4. Results of sensitivity of regional analysis to possible catchment-based pseudo replication. The values show the proportion of sites (over all water quality variables) that have improving trends at two levels of confidence for the full dataset and for the simulation. The results for the sensitivity analysis are the mean and standard deviation (in parentheses) for 100 realisations.

	Summary for selected level of confidence	
	At least as likely as not	At least likely
Full dataset	68.4	51.2
Sensitivity simulation	68.7 (1.2)	52.4 (1.4)

Table 5. Results of sensitivity of regional analysis to possible catchment-based pseudo replication by variable. The values show the proportion of sites (by water quality variable) that have improving trends at two levels of confidence for the full dataset and for the simulation. The results for the sensitivity analysis are the mean and standard deviation (in parentheses) for 100 realisations.

npID	Sensitivity simulation		Full dataset	
	At least as likely as not	At least likely	At least as likely as not	At least likely
Clar	91.9 (2.7)	89.1 (2.9)	90	88
Turb	67.1 (3.1)	56.8 (3.1)	66	55
DRP	62 (4)	46.6 (3.8)	66	49
TP	85.4 (2.1)	72.6 (3.2)	87	71
NO3-N	70.4 (2.7)	67.8 (2.5)	72	70
NNN	70.7 (3)	68.1 (2.9)	75	71
TN	93.8 (2.1)	86.4 (3.6)	92	82
TOC	73.5 (2.1)	41.2 (1.9)	75	45
E. coli	53 (2.4)	41 (3.1)	58	47
Chla	72.2 (1.9)	60.4 (2.2)	68	55
Fils-Max	73.9 (1.7)	42.9 (2.3)	76	40
Fils-Mean	52.8 (3.3)	29.5 (2.1)	53	29
Mats-Max	80.9 (2.4)	47.3 (2.9)	80	44
Mats-Mean	74.8 (2.1)	31.6 (2.5)	73	31
%EPT	61.4 (3.3)	46.5 (2.1)	57	41
%EPT_Taxa	65 (3.2)	53.8 (3.5)	63	50
MCI	53.9 (3.4)	40.6 (2.6)	52	35
QMCI	43.7 (2.6)	26.5 (1.9)	39	22

Conclusions

Other approaches could be taken to assessing pseudo replication. The approach taken by this is stringent in that only one site from any catchment is used in each realisation of the data. In reality, sites within a catchment can be entirely independent (e.g., all on tributaries such that none are upstream of one another). In addition, even sites on main stems can often be regarded as independent if they are sufficiently separated within the catchment network.

This analysis indicates that controlling for pseudo replication does not change the conclusions of Snelder (2017a).

References

Benjamini, Y. and Y. Hochberg, 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological):289–300.

Snelder, T., 2017a. Analysis of regional-scale river water quality trends in the Wellington region; Period 2007 to 2016. LWP Client Report: 2017-07, Christchurch.

Snelder, T., 2017b. Analysis of Water Quality Trends for Rivers and Lakes of the Wellington Region. LWP Client Report: 2017-01, Christchurch.

APPENDIX 2. EXCEPTIONS TO THE REGIONAL TREND IN WATER QUALITY

Three exceptions to the general pattern of improving water quality are noteworthy. First, one periphyton measure (Fils-Max) and two invertebrate measures (MCI and QMCI) have degraded at more sites than they have improved over the five-year time-period although there are not significant regional trends for these variables.

Second, there are some sites for which several water quality variables have degrading trends. These sites are inconsistent with the general regional pattern of water quality improvement and may reflect the effects of specific activities in the upstream catchments of the sites.

The make-up of sites with degrading trends for several variables differs between time periods but some sites are common to both time periods. For example, sites with more than five trends that were at least likely to be degrading in both time periods included Mangaone Stream at Sims Road Bridge, Otaki River at Pukehinau, Pakuratahi River 50m Below Farm Creek and Taueru River at Castlehill.

Third, five-year flow adjusted trends suggest that the dissolved nitrogen variables NO_3-N and NNN were at least as likely as not to be degrading at 61% and 54% of sites. However, the raw trends for these variables were at least as likely as not to be improving at 66% and 69% of sites. In addition, the regional-scale trends for NO_3-N and NNN based on the five-year flow adjusted trends were not significant. There is therefore some evidence for regional degradation of water quality with respect to dissolved nitrogen, however this is less compelling than the evidence for regional improvement in the majority of variables.