Memo

From	**Paul Franklin, Rick Stoffels, Doug Booker, Joanne Clapcott**
To	Ministry for the Environment
Date	13 March 2020
Subject	Contract 23184: Task 3 - Response to peer reviews of Appendix J of Franklin et al. (2019)

1 Background

Franklin et al. (2019) applied a range of different methods to characterise stressor-response relationships of macroinvertebrates and fish to turbidity, visual clarity and deposited fine sediment. These methods included: boosted regression trees (BRT); quantile regression; generalised linear modelling; extirpation analyses; and a community deviation method. Appendix J of Franklin et al. (2019) describes the community deviation method for characterising fish-sediment ESV and macroinvertebrate-sediment ESV responses. The community deviation method was developed specifically for this project. It was designed to characterise changes in community composition resulting from declines in sediment ESV state from predicted reference conditions. The community deviation method had the advantage of characterising impact across species/taxa from within a community, and accounting for the overall probability of occurrence of each species/taxa, but had not been previously published. In December 2019 the Ministry for the Environment (MfE) commissioned two peer reviews of Appendix J of Franklin et al. (2019).

MfE have requested a response to this peer review. MfE requested that "if the changes are limited to minor editorial edits and/or minor production of statistics, the Supplier will make these changes". MfE also requested a memo responding to comments and outlining additional work needed (if any).

2 Responses to Reviewer 1 (Gerry Closs)

Reviewer 1 stated that the analysis was "complex and rigorous" and "clearly represents a methodologically cutting-edge approach to defining landscape-specific thresholds for detecting community-level responses to fine sediment inputs in streams".

The reviewer referred to comments on p56 and p57 that stated that deposited sediment cover data paired with macroinvertebrate and used in the BRT analysis were not evenly distributed across the full gradient of deposited cover because more data were available from cleaner sites. The reviewer asked: "does this suggest that rapid and substantial community responses to increased sediment occur at relatively low levels of sediment input, with relatively limited changes as further inputs occur?". We clarify that these statements related to potential bias in the distribution of observed deposited fine sediment. The statements do not relate to community responses. However, various methods we applied did suggest that responses to increased fine sediment cover occur at relatively low levels of sediment input. We agree with the reviewer's assertion that if changes in community are occurring across a relatively narrow band of sediment input, then deriving several thresholds within that narrow range is difficult. This is related to a further point raised by Reviewer 1 regarding concerns over the practicalities of accurately distinguishing between attribute bands with relatively narrow ranges between their thresholds. We agree that accurately detecting differences between narrow bands may be hampered by uncertainties associated with field observations. However, we point out that the thresholds are defined by median values derived from two years of monthly data, and that field observations of median values should be considerably more certain than one-off observations. We believe that the question around likelihood that uncertainty in field observations would incorrectly place a site in the wrong band remains unresolved. A variety of factors might contribute to observation uncertainty for visual methods (deposited fine sediment or visual clarity) including inter-operator variability, lack of consistency in observation location, and variable light conditions.

Reviewer 1 has raised a more general concern related to poor understanding of the uncertainty associated with the proposed thresholds; this is partly related to a lack of information on uncertainty in the applied models. Reviewer 1 shared the authors' particular concern regarding compatibility between observed deposited fine sediment data contained within the New Zealand Freshwater Fish Database (NZFFD) and the data likely to be derived from the proposed method for observing deposited fine sediment (SAM2). This is a specific source of uncertainty within the analysis. Additional observations of deposited fine sediment using the instream visual method (SAM2) have provided more data since completion of the Franklin et al. (2019) report. In response to reviewer comments, we compared SAM2- and NZFFD-derived estimates of reference condition. The comparison indicated a bias between the two datasets, with the NZFFD containing systematically higher deposited fine sediment values relative to SAM2 observations, regardless of whether observations were from reference sites. The memo "Task 4 – Check of deposited fine sediment reference states against all available SoE data" provides further details of this analysis.

We agree that there are several further sources of uncertainty associated with the various analyses discussed in Franklin et al. (2019). Mathematical uncertainty associated with the proposed threshold bands is determined by a complex interplay between various input datasets (including their pre-processing, quality control, spatial resolution, temporal resolution, spatial coverage across NZ, coverage across each ESV), model conceptualisation, data post-processing, and output visualisation. Each of these aspects will be associated with epistemic uncertainties. Epistemic uncertainty stems from things that we could know across all of time and space in principle, but we can only represent approximately within available data and models, which in practice are coarse representations of reality.

We suggest that some sources of epistemic uncertainty could be quantified through further analysis of existing data (e.g. bootstrap resampling to quantify uncertainty in derivation of sediment classes or estimated reference conditions), but these analyses would have to assume that the existing data represented a sample of sites that was large enough to include sufficient replication (including replication of rare classes), and which represents a gradient of ESV conditions within each class in an unbiased manner. Evidence shown in Appendix D of Franklin et al. (2019) indicates that these assumptions are not valid for existing data sets, especially for turbidity (Figure D-13) and visual clarity (Figure D-17).

We also suggest that it is not possible to quantify some further sources of epistemic uncertainty, such as operator errors within existing data, which are extremely difficult to identify.

We further suggest that some sources of epistemic uncertainty could only be quantified following extensive investment. For example, inclusion of rare climate-topography-geology (CTG) classes in the data-driven approach to deriving sediment classes (and therefore estimated reference conditions) would only be possible following collection of data from several sites belonging to these classes.

We suggest that effective deployment of the proposed attributes and threshold bands will also be subject to considerable management decision uncertainties. These are likely to include decisions around level of investment in monitoring, which thresholds to target with management actions, and which management actions to apply. We point out that analysis of the importance of epistemic uncertainty versus decision uncertainty was not included within the scope of work addressed by Franklin et al. (2019).

3 Responses to Reviewer 2 (Angus McIntosh)

3.1 General response to Reviewer 2's concern over the regionalised approach

Reviewer 2's primary criticism of Franklin et al. (2019) concerned the regionalisation of fine sediment attribute thresholds. Both the Ministry and Franklin et al. recognised the need for a regionalised approach based on the following axioms:

1. Even in the absence of anthropogenic catchment modification, the natural supply and retention regimes of New Zealand's streams will vary regionally, and be dependent on the geological, topographical and climatological context of the region. The climatological, topographical and geological setting of a stream influences supply to and retention of sediment in that stream.
2. Aquatic invertebrates and vertebrates vary in their tolerance of deposited and suspended sediment, and so we can expect natural variation in the composition of aquatic biological communities among regions, even in the absence of any anthropogenic catchment modification.

Although Reviewer 2 agrees that (a) ecological communities are a good indicator of the impacts of fine sediment; (b) that fine sediment varies across the landscape; and (c) that aquatic species vary in their response, he doesn't agree that limits should be adjusted regionally to reflect that variation. It was not clear from the written review what Reviewer 2 would recommend as an alternative, but during the STAG meeting he recommended a non-regionalised approach, involving the development of a single set of attribute band thresholds that would be applied uniformly to all streams throughout New Zealand.

Reviewer 2 also argues that a regionalised fine sediment NOF will be difficult for managers to implement and difficult for the public to understand.

When estimating fine sediment attribute thresholds, we based our approach on the following:

A. Use of well established ecological principles/axioms concerning how the attributes to be managed affect ecological responses at the national scale.
B. Use of approaches that have been taken to similar policy problems in other countries undergoing water policy reform.
C. Careful consideration of the risks that may come with alternative approaches.

With respect to (A) the axioms mentioned earlier are very well supported by the ecological and geomorphological literature. Indeed, Reviewer 2 agreed with those axioms.

With respect to (B) a regionalised approach to the setting of water quality policy has been recommended in the scientific literature in several national contexts: North America (Dodds and Oakes 2004, Dodds et al. 2006, Herlihy and Sifneos 2008, Hawkins et al. 2010), China (Huo et al. 2012) and New Zealand (McDowell et al. 2013). Examples of studies where non-spatial approaches have been used to set water quality policy exist in the scientific literature, but they usually concern individual river catchments or local governmental policy problems, not national-scale problems.

With respect to (C) there are risks associated with both regionalised and non-regionalised approaches. Taking a spatially-uniform, non-regionalised approach creates the risk of being unrealistically 'over-protective' of instream values in some regions and 'under-protective' of instream values in other regions.

- In regions that naturally have higher sediment supply and retention, a single, nation-wide threshold may not ever be achievable.
- In other regions that naturally have very low sediment supply and retention, a single threshold may be under-protective, giving license to land use degradation.

Consequently, a non-regionalised approach creates the potential for greatly eroding stakeholder confidence in both policy and policy-makers; attribute thresholds need to be realistic and achievable.

When taking a regionalised approach, one could argue that the difficulties associated with implementing and interpreting a NOF with, say, 12 different classes of attribute thresholds nationwide may make the NOF too unruly to implement 'on the ground'. This issue is for regional councils to comment on and we would be happy to hear their feedback. Another potential risk associated with taking a regionalised approach may arise from the introduction of "biased" thresholds derived from analyses that are not supported by sufficient data. The more we decompose New Zealand's river network into different classes, the less data we have for, say, estimation of reference states within individual classes. We were aware of this risk prior to taking a regionalised approach, and we were very careful to check data quality and quantity in each class prior to estimation of attribute thresholds.

The hierarchical classification provided by Franklin et al. (2019) was specifically designed to allow assessment of the trade-offs that occur as spatial complexity is increased. When fewer classes are defined, more data are available for each class, allowing more precise estimation of the reference condition for the overall class. In this situation there is greater spread of sites around the estimated reference condition because there is more natural variation within the class, and therefore a greater chance of the reference condition being inappropriate for some sites within the class. When more classes are defined, less data are available for each class, and therefore a less precise (more uncertain) estimate of the reference condition for the overall class is obtained. In this situation there is less spread of sites around the estimated reference condition because there is less natural variation within the class, and therefore a reduced chance of the reference condition being inappropriate for sites within the class exists.

On balance, therefore, our approach is consistent with ecological axioms and approaches recommended in the scientific literature. Although we acknowledge that our approach comes with risks, those risks are certainly no greater than the risks that come with a non-regionalised approach. In fact, a relatively recent critique of processes associated with setting water quality policy highlights the problems created by non-regionalised approaches and calls for future approaches to take into account natural variability (Page et al. 2012). We cannot, therefore, concur with Reviewer 2's comments that we used a "flawed approach" and that we should instead adopt a non-regionalised approach.

3.2 Response to Reviewer 2's specific comments on the regionalised approach

> *"However, just because a flat region of the country tends to accumulate more fine sediment, the impact on those ecological values (and life supporting capacity) isn't less! It means a flat area will be more prone to degradation, and it certainly doesn't mean the limits set there should be discounted."*

Franklin et al. have not suggested that lowland streams are more resistant to fine sediment. Franklin et al. stated that certain lowland streams will have naturally had higher sediment supply and retention characteristics and, consequently, would naturally support different aquatic communities to, say, headwater streams (Franklin et al. (2019), Section 2.2.1, p 26). Following the establishment of class-specific reference/baseline states, we then developed class-specific ecological response models that account for differing sensitivities of the community to degradation across classes.

We agree that attribute thresholds in lowlands streams should not be "discounted", which is why we did not discount them, as was made clear in Franklin et al. (2019).

> *"If I'm interpreting the classification and clustering of landscape-dependent fine sediment classes correctly (Appendix D), it is based on data which includes those from already impacted locations. If the classification is to accurately define classes, then it would need to be based on data from non-degraded sites. Thus, I worry that the classification is not a fair representation of the fine sediment state associated with locations because I can't see an attempt to deal with the already degraded state of many of the location used."*

It is true that when developing our sediment state classifications (SSCs) we used all available data on attribute states, including data from streams already impacted by anthropogenic land use. A risk in taking this approach is that we might derive an SSC that characterises gradients in anthropogenic disturbance, rather than the natural sediment supply and retention regimes of the streams as desired. We managed this risk in two ways:

First, before allowing sediment data to influence our SSC we first classified streams according to their climatological, topographical and geological contexts, as determined using the River Environment Classification (REC). Once streams were assigned to a climate-topography-geology (CTG) class, we then collated all available attribute data within those CTG classes and used hierarchical clustering to determine further possible levels of spatial aggregation/lumping. By using CTG classes as a basis for our hierarchical SSC we ensured that the SSC reflected conceptual differences in sediment supply and retention, and not just spatial patterns in land use.

Second, once the SSC for each attribute was developed, we then used models to (a) estimate reference state and (b) estimate band thresholds. These models rely on having a good spread of data across the three key anthropogenic stressor variables (% upstream heavy pasture, forestry and urbanisation) within each class. We observed surprisingly good spread of these stressor variables within each class, giving us confidence that our SSC was not describing spatial patterns in anthropogenic disturbance, but spatial variation in natural sediment supply and retention characteristics.

> *"The absence of actual reference sites for both fish and macroinvertebrate data for a number of spatial classes considerably reduces the rigour because reference state has had to be 'estimated'. I also found it hard to distinguish the rationale for what was regarded as a reference site or to determine what the effect of 'reference site estimation' had on the relationships. It would be useful to see some comparison of sites that were estimated versus those not, and a more comprehensive justification of what constitutes a reference site."*

Our method for estimating reference state is based on a well-documented scientific method (Dodds and Oakes 2004). In Appendix D we compare reference states estimated using the model-based approach with observed data from reference sites.

> *"There is an assumption (appendix D) that the REC framework represents the underlying processes that govern fine sediment distribution and this has not been adequately justified. I was expecting an analysis of actual fine sediment measures – did I miss it?"*

In Appendix D we have explained why the REC climate, topography and geology classes should coarsely represent the supply and retention regimes of streams. As explained above, once we used REC CTG classes as a basis for our SSC we then brought data into the analysis, to further aggregate spatial classes. In Appendix D we present violin plots showing the distribution of observed attribute states within CTG classes. The violin plots show that the CTG classes do indeed characterise fine sediment states of streams in intuitive ways. For example, 'warm-wet' areas are characterised by higher concentrations of deposited fine sediment than 'cool-wet' areas (Figure D-2 in Franklin et al. 2019).

3.3 General response to Reviewer 2's concern over the community deviation method

Reviewer 2's primary concern with the community deviation method is that it relies on "layer upon layer of prediction" and, therefore, it is "hard to evaluate what level of confidence we could have". We note that Reviewer 1 commented that the approach "clearly represents a methodologically cutting-edge approach to defining landscape-specific thresholds for detecting community-level responses to fine sediment inputs in streams", but also noted the risks related to uncertainty in the modelling approach. We have discussed aspects of uncertainty in the modelling approach in our response to Reviewer 1 and consider it equally applicable here. All analytical approaches used by Franklin et al. (2019), including the community deviation method, quantile regression, BRT and extirpation analyses involved some degree of prediction and modelling of ecological responses. We followed the US EPA (2016) weight of evidence assessment framework to evaluate the relative strengths and weaknesses of all the modelling approaches, which included consideration of uncertainty and representativeness of the different methods. We accept that there are uncertainties in the method, some of which are unquantifiable, but this is true of all the methods to some extent. We used a data-driven approach that is transparent and reproducible. We note that Reviewer 2 did not suggest an alternative analytical approach that met the requirements and guiding principles set out in section 3 of Franklin et al. (2019).

Reviewer 2 also suggested that statistical measures of uncertainty would be useful. We point out that the community deviation method essentially comprised two components: 1) a reference condition estimate, and 2) a set of probability of occurrence estimates for each of several species/taxa. With regard to reference condition estimates, uncertainties in estimated reference conditions are depicted in Figure D-8 for deposited fine sediment, Figure D-12 for turbidity, Figure D-16 for visual clarity. AIC, Akaike weight of model, and log-likelihood statistics for Model 4 fitted to data at Levels 1 through to 4 are shown in Table D-5 for deposited fine sediment, AIC Table D-7 for turbidity, and Table D-9 for visual clarity. With regard to probability of occurrence estimates, AUC and AIC of two alternate models of each invert taxa are shown in Figure J-7 and Figure J-8. AUC and AIC of two alternate models of each fish species are shown in Figure J-9 and Figure J-10.

Reviewer 2 suggested a need for verification using independent data. We agree with this suggestion. Please see accompanying memo (Task 4), related to assessment of deposited fine sediment threshold bands in comparison with independent SoE data. Analysis for Task 4 has provided strong evidence for re-evaluating the deposited fine sediment thresholds downwards. We also point out that Reviewer 1 stated that "the authors have accessed and made appropriate use of available relevant datasets".

3.4 Response to Reviewer 2’s specific comments on the community deviation method

> *“The Community Deviation method had to ‘fill in’ missing data in some (many?) cases using adjacent sites”*

We understand that Reviewer 2 is largely referring to the section ‘Matching fish data with observed ESV data’ in Appendix J. In this section we describe a process we undertook to evaluate whether it was possible to match observed fish data from the New Zealand Freshwater Fish Database (NZFFD) to observed sediment ESV data (deposited and suspended) in space and time. In no cases were data infilled from adjacent sites and used in the community deviation method. For deposited sediment, the community deviation method applied to fish used paired observations of fine sediment cover and fish recorded in the NZFFD and collected at the same time and place. The community deviation method applied to macroinvertebrates used paired observations of fine sediment cover and macroinvertebrates from a combined research and regional council State of the Environment (SoE) data set. For suspended sediment, no paired observations of fish and suspended sediment were available, so modelled site median estimates of turbidity and visual clarity were paired with fish observations. Paired observations of suspended sediment ESVs and macroinvertebrates were available from the SoE and National Water Quality Monitoring Network sites and were used for the community deviation analyses.

> *“One community assembly process was considered (competition/predation with trout) when creating models of community structure, but there are many other community assembly processes which are not included…”*

We accept that multiple community assembly processes may affect community structure. Some of these processes act primarily across space (i.e. whether trout are present or not), and others act primarily across time (e.g. flood disturbance). Our analyses were limited to use of existing data and to a consideration of spatial variations in community structure at a regional to national scale, which is consistent with assessment of attributes against a site summary statistics such as a median.

> *“The fish species modelled to produce the fish limits do not include a non-migratory galaxiid…They are also likely to be particularly vulnerable to fine sediment accumulations…”*

As stated in the proposed attribute tables in Franklin et al. (2019) “Bottom-line thresholds are anticipated to provide a sufficient level of protection at an overall fish community level (i.e., will cause <20% decrease in the *fish community deviation metric*). Bottom-line thresholds may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations for certain biota”. While we acknowledge that non-migratory galaxiids are an important and threatened group of fishes in New Zealand that may be susceptible to fine sediment accumulation, the majority of these species are very geographically limited and not representative of fish communities in New Zealand as a whole. This makes them unsuitable for specific inclusion in a model to derive national scale attribute limits. It is our view that such species should be subject to specific management plans and will also be provided some degree of protection by the NPS-FM requirement to avoid deterioration in state.

3.5 Supporting literature

Dodds, W. K., E. Carney, and R. T. Angelo. 2006. Determining ecoregional reference conditions for nutrients, Secchi depth and chlorophyll a in Kansas lakes and reservoirs. Lake and Reservoir Management **22**:151-159.

Dodds, W. K., and R. M. Oakes. 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnology and Oceanography: Methods **2**:333-341.

Franklin, P., R. Stoffels, J. Clapcott, D. Booker, A. Wagenhoff, and C. Hickey. 2019. Deriving potential fine sediment attribute thresholds for the National Objectives Framework. NIWA Report to the Ministry for Environment

Hawkins, C. P., J. R. Olson, and R. A. Hill. 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society **29**:312-343.

Herlihy, A. T., and J. C. Sifneos. 2008. Developing nutrient criteria and classification schemes for wadeable streams in the conterminous US. Journal of the North American Benthological Society **27**:932-948.

Huo, S. L., F. Y. Zan, Q. Chen, B. D. Xi, J. Su, D. F. Ji, and Q. G. Xu. 2012. Determining reference conditions for nutrients, chlorophyll a and Secchi depth in Yungui Plateau ecoregion lakes, China. Water and Environment Journal **26**:324-334.

McDowell, R. W., T. H. Snelder, N. Cox, D. J. Booker, and R. J. Wilcock. 2013. Establishment of reference or baseline conditions of chemical indicators in New Zealand streams and rivers relative to present conditions. Marine and Freshwater Research **64**:387-400.

Page, T., A. L. Heathwaite, B. Moss, C. Reynolds, K. J. Beven, L. Pope, and R. Willows. 2012. Managing the impacts of nutrient enrichment on river systems: dealing with complex uncertainties in risk analyses. Freshwater Biology **57**:108-123.

US EPA (2016) Weight of evidence in ecological assessment, EPA/100/R-16/001.