Options for river water quality management classifications for the Northland region

June 2017

Prepared By:

Ton Snelder – Land Water People Ltd

Tim Kerr – Aqualinc Research Ltd

For any information regarding this report please contact:

Ton Snelder

Phone: 027 575 8888 or 03 377 3755

Email: ton@lwp.nz

Land Water People Ltd

PO Box 70

Lyttelton 8092

New Zealand

LWP Client Report Number: LWP Client Report 2017-02

Report Date: 7 June 2017

LWP Project: LWP Client Report 2017-02

Quality Assurance Statement

Version	Reviewed by		
1	Caroline Fraser	20 May 2017	
Final	Caroline Fraser	7 June 2017	

Table of Contents

Executive Summary

This report follows previous work undertaken for Northland Regional Council (NRC) to define freshwater management units (FMUs). A key step in the definition of FMUs was to classify the region's rivers for water quality management. The purpose of the "water quality management classification" is to broadly discriminate variation in the characteristics of the water bodies that are relevant to management including their current water quality state, values, and capacity for resource use.

The original water quality management classification recommended by Snelder (2015) divided the region in to two classes, where river reaches were defined as hill or lowland, depending on the average upstream slope. The classification was mapped using a digital representation of the stream and rivers of the Northland region. This simple classification was recommended because variation in water quality in the region is complex, and attempts to discriminate finer scale patterns in the variation in general water quality in Northland were not successful.

NRC determined that there may be advantages for regional land and water management if the characteristics of water bodies that are relevant to management were more finely discriminated than the original, two-class, water quality management classification. This report details an investigation of alternative classifications that used the most up-to-date data and statistical modelling to derive and test alternative classifications.

Classifications were derived by first extracting site median values of several water quality variables that had been measured at monthly intervals for periods of between one and 19 years at 63 sites distributed across Northland. Statistical cluster analysis was used to group the sites into 2, 3 and 4 classes based on the water quality data. Three types of statistical model were then used to determine relationships between the cluster-based site classifications and a suite of environmental predictors that represent the characteristics of each site's upstream catchment. These relationships were used to make predictions of the water quality classes for all segments of a digital network representing the region's rivers and produce maps of the alternative classifications. The ability of all classifications to explain variation in each of the observed water quality variables was assessed using analysis of variance (ANOVA) and a multi-variate form of ANOVA (called ADONIS). The performance of the statistically derived classifications was compared to the original management classification recommended by Snelder (2015).

The ANOVA and ADONIS tests indicated that the classifications based on statistical clustering discriminated water quality variation significantly better than the original water quality management classification. The classes based on statistical clustering could be understood from analysis of the water quality data and described a gradient across the region from the highest to lowest water quality conditions. A key finding was that discrimination of water quality increased significantly when three classes were defined compared to two classes but that the increase in discrimination was less from three to four classes.

The relationships between the cluster-based site classifications and environmental predictors were used to describe the environmental characteristics of the classes. The models indicated that aspects of catchment geology and catchment topography are strongly associated with water quality variation in the Northland region.

The study results indicate that a water management classification that is defined using statistical modelling will discriminate regional water quality variation better than the original water quality management classification recommended by Snelder (2015). However, the

original classification is easily explained. By contrast, statistically-based classifications are not easily explained and do have subjective elements. This may make it difficult for the community to accept a statistically-based water quality management classification.

The information presented in this report and the associated supplementary data provide a basis for NRC to consider alternatives for the water quality management classification. An important consideration is whether the increase in the discrimination of variation in water quality variation over the original classification is sufficiently advantageous that it outweighs the increased complexity and difficulty in explaining the classification. The decision to adopt an alternative classification is a subjective (political) one. This report does not recommend adoption of any alternative but does provide the information to assist that decision.

The statistical approach to defining the classifications was limited by the number of sites with water quality data and the representativeness of these sites of regional water quality patterns. As a result, some environments that may be regarded as distinctive regionally will not be represented by the classifications. We therefore recommend that if any of the statistically defined classifications presented here were to be used, some refinement of the classification may be appropriate. The refinement could be carried out by incorporating 'local knowledge' which is not reflected in the site data and spatial modelling. Care would need to be taken to avoid generating many small-scale classes based on local knowledge as this would undermine the general objective of providing a succinct classification for management purposes.

1 Introduction

This report follows previous work undertaken for Northland Regional Council (NRC) to define freshwater management units (FMUs) (Snelder, 2015). FMUs were designed to provide a regional spatial framework for managing river water quality and quantity in a new regional water plan that will implement the National Policy Statement – Freshwater Management (NPSFM). A key step in the definition of FMUs was to classify the region's rivers for water quality management. The purpose of the "water quality management classification" was to broadly discriminate variation in the characteristics of the water bodies that are relevant to management including their values and capacity for resource use.

A key assumption underlying the water quality management classification recommended by Snelder (2015) was that objectives and policies would aim to maintain the current state of water quality. The maintain-current-state requirement effectively sets the capacity for use of water bodies in each class. This means that a water quality management classification needs to discriminate variation in existing water quality. To be effective, the classification should also discriminate variation in other values of water bodies, including instream values and the economic values of their upstream catchments. In addition, a water quality management classification is ideally based on "inherent" factors (i.e. factors that are natural and unaffected by resource use, such as climate, topography and geology) so that the discrimination of the characteristics of water bodies is independent of existing activities.

Snelder (2015) analysed Northland's 'general' river water quality (i.e. water quality as defined by a mix of physical, chemical, and biological parameters) and found broad differences were associated with variation in catchment topography. Steep hill catchments are associated with relatively higher overall water quality than lowland (low gradient) catchments. Catchment topography not only discriminates many water quality parameters but is also broadly associated with differences in the economic value of upstream catchments (e.g., lowland areas tend to be more intensively farmed and urbanised than hill country areas) and other relevant management considerations such as river size and hydrological regime, which affect contaminant dilution, transport and assimilation.

The original water quality management classification recommended by Snelder (2015) was therefore a simple one that comprised two classes: hills and lowlands. This classification was applied to a digital representation of the stream and rivers of the Northland region. The region's rivers were represented as individual segments, each of which was classified on the basis the topography of the upstream catchment. Individual segments were classified as hill class if the average slope of the upstream catchment was greater than 10 degrees and lowland if the average slope was less than 10 degrees (Figure 1).

Attempts by Snelder (2015) to discriminate finer scale patterns in the variation in general water quality in Northland were not successful. This is because variation in water quality in the region is complex. Individual water quality variables tend to vary independently (i.e. some variables have low correlation with other variables). In addition, there is variation in the strength of the relationships between the individual variables and catchment characteristics that are potentially used to define classes such as topography, geology, land cover, and climate.

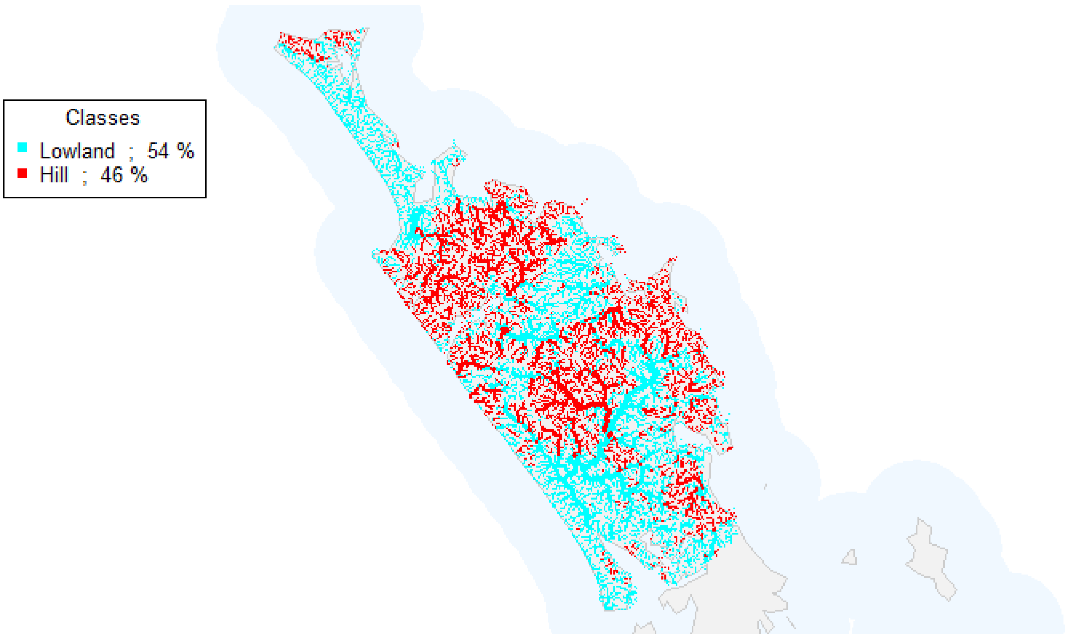

Figure 1. The original water quality classification recommended by Snelder (2015). The classification is based on the average slope of upstream catchments being greater ("Hill" class) or less than ("Lowland") 10°.

Subsequent to the development of the original FMUs, NRC developed a new regional water plan. During that process, NRC determined that there may be advantages for regional land and water management if the characteristics of water bodies that are relevant to management were more finely discriminated than the original water quality management classification recommended by Snelder (2015). NRC approached LWP Ltd and Aqualinc Research Ltd to study alternative classifications.

This report details an investigation of alternative water quality management classifications that aim to improve the discrimination of the original classification recommended by Snelder (2015). The approach taken in the present study has made a key assumption that a water quality management classification needs to discriminate variation in existing water quality. As for the original study, it has also been assumed that it is useful to understand the relationship between different water quality states and catchment characteristics. This study has therefore used statistical techniques applied to river water quality measurements made across Northland to explore these relationships. The statistical analysis can be broadly understood as three key steps:

1. Statistical clustering was used to group sites into classes that have similar general water quality.

2. Statistical spatial modelling was used to associate the classes with characteristics of their upstream catchments (e.g. topography, geology etc.).

3. The statistical model was used to extrapolate the classes to all segments of the digital river network in Northland.

This report describes the methods used in these analyses, the ability of the classes to discriminate water quality, and the performance of the spatial models. The report describes the trade-off decision that would be needed to decide on an alternative water quality management classification and suggests the most suitable candidates from among the various options that were developed.

2 Data

As of the end of 2016, NRC had carried out monthly monitoring at 59 river sites for periods of between one and 19 years. This comprises 30 sites that have at least 10 years of monthly observations, and 29 recently established ‘priority catchment’ sites with up to three years of monthly observations. A variety of physical, chemical, and biological indicators of water quality are measured at these sites (Table 1). In addition, water quality and biological monitoring has been carried out by NIWA since 1989 at the 4 river sites in the Northland region as part of the National River Water Quality Network (NRWQN).

NRC provided all available water quality data for these sites. In addition to water quality data, NRC provided metadata for sites, including: site name, location, identifier, NZMS260 grid reference, and NZReach number (as defined in the River Environment Classification (REC) geodatabase).

Table 1. Summary of water quality used to define and test the classifications. The number of sites, the mean of the site median values and the range of the site median values.

Water quality variable	Units	Number of Sites	Mean of site medians	Min:Max of site medians
Ammoniacal Nitrogen	g/m^3	63	0.017	0.0025:0.1
Dissolved Reactive Phosphorus	g/m^3	63	0.021	0.005:0.18
Escherichia coli	MPN/100ml	63	329	63:1046
Nitrite/nitrate nitrogen	g/m^3	63	0.3	0.004:2.6
Total Nitrogen	g/m^3	63	0.54	0.05:2.8
Total Phosphorus	g/m^3	63	0.043	0.009:0.36
Turbidity	NTU	63	4.8	1:14
Ammoniacal Nitrogen*	g/m^3	60	0.0086	0.0022:0.039
Clarity	m	60	1.3	0.3:2.2
pH	g/m^3	63	7.3	6.6:7.9
Dissolved Oxygen	%	63	90	39:109
Dissolved Oxygen	g/m^3	63	9	3.8:11
Taxa Richness	NA	19	116	12:25
MCI	NA	19	37	68:141
%EPT Taxa	NA	19	20	0:60
Chlorophyll a (benthic)	mg/m^3	37	40	1.6:233

* pH adjusted

3 Methods

3.1 Selection of sites and preparation of data

For the analysis that follows, it was a requirement that all sites had a value for all the water quality variables. Therefore, the water quality variables were reduced to a core set for which a maximum number of sites had observations. The biological variables: Chlorophyll, MCI, Taxa Richness and % EPT Taxa, were excluded because these were measured at a relatively small subset of sites (Table 1). We also excluded the variables pH and dissolved oxygen. These variables are subject to diurnal variability and therefore the site median values are not necessarily a reliable characterisation of the central tendency for the site. This left a core set of variables (the first nine variables shown in Table 1) which were available at most of the 63 sites.

We calculated the median value of each of the nine core variables at each site. Where a site did not have data for a particular variable (i.e. for ammoniacal nitrogen and visual clarity at three sites) we substituted the mean value of the median values over the other sites with available observations. This treatment slightly reduces the reliability of the statistics associated with ammoniacal nitrogen and clarity but retained 63 sites for analysis. For all the analysis that follow we log (natural) transformed the median values of each variable so that they were approximately normally distributed.

3.2 Cluster-based classifications of sites

Cluster analysis or clustering was used to classify the sites so that sites in the same cluster (called a class) are more similar to each other than to sites in other classes. The similarity between sites is measured using the combined differences between all pairs of sites of each water quality variable. We used K-means clustering, which is a non-hierarchical clustering method (Crawley, 2002). The K-means algorithm clustered the 63 sites into a user-defined number of classes such that the sum of squares of differences from sites to the means of their assigned classes is minimized.

Clustering solutions are influenced by the input data. To allow us to investigate the effects of the input data on the cluster solutions, we made three separate input data sets based on different treatments of the original data and made clusters from each dataset. First, before clustering, we scaled all log transformed variables to have a mean of 0 and a standard deviation of 1. This ensured that all variables had similar numeric ranges so that they had equal weight in the analyses that follow. These data were called the “raw” dataset.

Second, if some of the input variables are correlated this will produce classifications that more strongly emphasise variation in those (correlated) variables at the expense of the other variables. Correlation among our water quality variables was very likely, especially between the nitrogen species because these variables are strongly related. To remove correlation in the input data we subjected the log transformed and scaled variables to rotated principle component analysis (PCA) (Crawley, 2002). We retained the scaled site scores on the first two components as an alternative input dataset. These data represent a simplified and uncorrelated (i.e. orthogonal) projection of the original data. The rotation maximises the correlation of the retained components with the original variables, enabling the components to be interpreted. This dataset was called the “PCARot” set.

The third treatment of the input variables was to deliberately weight a variable to enhance its discrimination by the classification. In consultation with NRC we decided to weight the cluster solution to enhance the discrimination of Clarity. We achieved this by first finding the rotated component of the PCARot that had the highest correlation with the original Clarity data. We then multiplied the site scores on this component by a factor of 3. This was called the "RotWt" set.

The 63 sites were clustered into two, three or four classes using each of the three input data sets. This resulted in nine different classifications of the sites. Every site was assigned a membership of a class for each of the nine classifications. The names and methods for the 9 cluster-based classifications are described in Table 2.

Table 2. Names and descriptions of the 9 cluster-based classifications of sites.

Name of classification	Input data set	Number of clusters
Raw2Clusters	Scaled logs of parameter medians	2
Raw3Clusters	Scaled logs of parameter medians	3
Raw4Clusters	Scaled logs of parameter medians	4
PCARot2Clusters	First 2 components of rotated PCA	2
PCARot3Clusters	First 2 components of rotated PCA	3
PCARot4Clusters	First 2 components of rotated PCA	4
RotWt2Clusters	Weighted rotated PCA	2
RotWt3Clusters	Weighted rotated PCA	3
RotWt4Clusters	Weighted rotated PCA	4

3.3 Spatial modelling of cluster-based classifications

The next step used spatial statistical modelling to determine the relationship between the cluster-based site classifications and a suite of environmental predictors that represent the characteristics of each site's upstream catchment. We used three types of statistical model (from relatively simple to complex) to represent these relationships:

1. Linear Discriminant Analysis (LDA). LDA is similar to regression in that it finds the best linear combination of the predictors that discriminates class membership (Crawley, 2002). LDA solutions can therefore be expressed as linear models (i.e. an equation).

2. Classification and Regression Trees (CART). A CART model is a tree-like structure that splits the sites into successively smaller groups for which the class membership of the group becomes more homogeneous (Breiman et al., 1984). The splits are defined by particular values of one of the predictors so that the tree can be understood as a set of decisions made using the predictors to "find" the cluster membership of each site.

3. Random Forest (RF) model (Breiman, 2001). RF models are a machine learning technique and are an ensemble of many CART models, all varying subtly from each other. RF models generally always predict the response more accurately than CART models but are "black boxes" in that they cannot be expressed as either equations (like LDA) or simple tree diagrams (like CART).

Our rationale for using these three modelling methods was to explore the impact of model complexity on the predictive performance of the modelling methods. In general, it is preferable that the basis for a water quality management classification can be easily understood and appreciated. It is also important, however, that the spatial modelling is accurate (i.e. that the relationships between water quality measured at a site and the characteristics of its upstream catchment is robust). There is a trade-off between model complexity and prediction performance that can be demonstrated by the three modelling methods.

The spatial framework for the spatial models was a digital representation of Northland's drainage network extracted from the New Zealand river environment classification (REC, version 2) (Snelder and Biggs, 2002). The first version of the REC was used as the spatial framework to define the original water quality management classification and FMUs (Snelder, 2015). The digital network was derived from a digital elevation model (DEM) with a spatial resolution of 50 m (Snelder and Biggs, 2002). Computer analysis of the DEM identified drainage paths, network segments and associated catchment boundaries. Version 2 of the REC improved some of the details of the representation of the region's stream and rivers by the digital network. The digital network represents Northland's rivers as 29,000 segments (delineated by upstream and downstream confluences) and their catchments, and is contained in a spatial database.

The digital river network has been combined with spatial data layers describing the climate, topography, geology, and land cover of New Zealand to produce a large number of catchment characteristics for each network segment (Leathwick et al., 2011). We extracted from these data several predictors that described the character of the catchments upstream of the monitoring sites and used them as inputs to the statistical models (

Name	Description	Units	Land cover predictor?
us_catarea	Area of the upstream catchment	m^2	NO
us_rain	Mean annual rainfall of the upstream catchment	mm	NO
us_pet	Mean annual potential evapotranspiration of the upstream catchment	mm/yr	NO
QMean	Mean annual stream flow	m^3/s	NO
seg_elev	Mean elevation of the river segment	M ASL	NO
us_lakePerc	Percentage of landcover in lakes in the upstream catchment	%	NO
us_elev	Mean elevation of the upstream catchment	M ASL	NO
us_slope	Mean slope of the upstream catchment	degrees	NO
us_tmin	Catchment average mean June air temperature	degrees C x 10	NO
us_twarm	Catchment average mean January air temperature	degrees C x 10	NO
us_rnvar	Catchment average coefficient of variation of annual rainfall	mm/yr	NO
us_rd10	Catchment average frequency of rainfall > 10mm	days/mo	NO
us_rd20	Catchment average frequency of rainfall > 20mm	days/mo	NO
us_rd100	Catchment average frequency of rainfall > 100mm	days/mo	NO
*us_hard	Catchment average induration or hardness value	Ordinal	NO
*us_phos	Catchment average phosphorous	Ordinal	NO
*us_psize	Catchment average particle size	Ordinal	NO
*usCalcium	Catchment average calcium	Ordinal	NO
us_IntensiveAg	Percentage landcover in intensive agriculture based on LCDBV3 in the upstream catchment	%	Yes

us_PastoralLight	Percentage landcover in light pasture based on LCDBV3 in the upstream catchment	%	Yes
us_NativeForest	Percentage landcover in native forest based on LCDBV3 in the upstream catchment	%	Yes
us_Urban	Percentage urban landcover based on LCDBV3 in the upstream catchment	%	Yes
us_Scrub	Percentage landcover in scrub based on LCDBV3 in the upstream catchment	%	Yes
us_Wetland	Percentage landcover in wetland based on LCDBV3 in the upstream catchment	%	Yes
us_ExoticForest	Percentage landcover in exotic forest based on LCDBV3 in the upstream catchment	%	Yes
us_Bare	Percentage landcover in bare land based on LCDBV3 in the upstream catchment	%	Yes

Table 3).

Some of the input predictors represent catchment land cover which reflects human activities (i.e. land use) whereas other predictors represent "inherent" catchment characteristics such as climate, topography and geology. Ideally a water quality management classification would be based only on "inherent" catchment characteristics and would be independent of human activities. However, we recognised that disentangling these two aspects is difficult due to the strong association between "inherent" catchment characteristics and human activities. In addition, excluding land use from statistical models of catchment water quality may reduce model performance significantly due to the strong relationships between the two (Larned et al., 2016). We therefore produced two sets of spatial models. The first set used all available predictors, including predictors that represent land cover (

Name	Description	Units	Land cover predictor?
us_catarea	Area of the upstream catchment	m^2	NO
us_rain	Mean annual rainfall of the upstream catchment	mm	NO
us_pet	Mean annual potential evapotranspiration of the upstream catchment	mm/yr	NO
QMean	Mean annual stream flow	m^3/s	NO
seg_elev	Mean elevation of the river segment	M ASL	NO
us_lakePerc	Percentage of landcover in lakes in the upstream catchment	%	NO
us_elev	Mean elevation of the upstream catchment	M ASL	NO
us_slope	Mean slope of the upstream catchment	degrees	NO
us_tmin	Catchment average mean June air temperature	degrees C x 10	NO
us_twarm	Catchment average mean January air temperature	degrees C x 10	NO
us_rnvar	Catchment average coefficient of variation of annual rainfall	mm/yr	NO
us_rd10	Catchment average frequency of rainfall > 10mm	days/mo	NO
us_rd20	Catchment average frequency of rainfall > 20mm	days/mo	NO
us_rd100	Catchment average frequency of rainfall > 100mm	days/mo	NO
*us_hard	Catchment average induration or hardness value	Ordinal	NO
*us_phos	Catchment average phosphorous	Ordinal	NO

*us_psize	Catchment average particle size	Ordinal	NO
*usCalcium	Catchment average calcium	Ordinal	NO
us_IntensiveAg	Percentage landcover in intensive agriculture based on LCDBV3 in the upstream catchment	%	Yes
us_PastoralLight	Percentage landcover in light pasture based on LCDBV3 in the upstream catchment	%	Yes
us_NativeForest	Percentage landcover in native forest based on LCDBV3 in the upstream catchment	%	Yes
us_Urban	Percentage urban landcover based on LCDBV3 in the upstream catchment	%	Yes
us_Scrub	Percentage landcover in scrub based on LCDBV3 in the upstream catchment	%	Yes
us_Wetland	Percentage landcover in wetland based on LCDBV3 in the upstream catchment	%	Yes
us_ExoticForest	Percentage landcover in exotic forest based on LCDBV3 in the upstream catchment	%	Yes
us_Bare	Percentage landcover in bare land based on LCDBV3 in the upstream catchment	%	Yes

Table 3). The second set of models were fitted using a subset of predictors that did not include land cover (i.e. the second set of models used only “inherent” catchment characteristics as predictors). Our rationale for defining two sets of spatial models was to explore the impact of choice of predictors on the performance of the models.

In total, six different model-predictor set combinations (three model methods and two sets of input predictors) were applied to estimate the nine different cluster sets (i.e. 54 models). The fitted models were used to predict the class membership for all 29,000 segments of the digital river network representing the Northland region. The predictions were used to produce maps of the region for all 54 models, with all river segments colour coded to reflect their assigned classes.

Table 3. Model predictors used by the spatial models.

Name	Description	Units	Land cover predictor?
us_catarea	Area of the upstream catchment	m^2	NO
us_rain	Mean annual rainfall of the upstream catchment	mm	NO
us_pet	Mean annual potential evapotranspiration of the upstream catchment	mm/yr	NO
QMean	Mean annual stream flow	m^3/s	NO
seg_elev	Mean elevation of the river segment	M ASL	NO
us_lakePerc	Percentage of landcover in lakes in the upstream catchment	%	NO
us_elev	Mean elevation of the upstream catchment	M ASL	NO
us_slope	Mean slope of the upstream catchment	degrees	NO
us_tmin	Catchment average mean June air temperature	degrees C x 10	NO
us_twarm	Catchment average mean January air temperature	degrees C x 10	NO
us_rnvar	Catchment average coefficient of variation of annual rainfall	mm/yr	NO
us_rd10	Catchment average frequency of rainfall > 10mm	days/mo	NO
us_rd20	Catchment average frequency of rainfall > 20mm	days/mo	NO
us_rd100	Catchment average frequency of rainfall > 100mm	days/mo	NO
*us_hard	Catchment average induration or hardness value	Ordinal	NO
*us_phos	Catchment average phosphorous	Ordinal	NO
*us_psize	Catchment average particle size	Ordinal	NO
*usCalcium	Catchment average calcium	Ordinal	NO
us_IntensiveAg	Percentage landcover in intensive agriculture based on LCDBV3 in the upstream catchment	%	Yes
us_PastoralLight	Percentage landcover in light pasture based on LCDBV3 in the upstream catchment	%	Yes
us_NativeForest	Percentage landcover in native forest based on LCDBV3 in the upstream catchment	%	Yes
us_Urban	Percentage urban landcover based on LCDBV3 in the upstream catchment	%	Yes
us_Scrub	Percentage landcover in scrub based on LCDBV3 in the upstream catchment	%	Yes
us_Wetland	Percentage landcover in wetland based on LCDBV3 in the upstream catchment	%	Yes
us_ExoticForest	Percentage landcover in exotic forest based on LCDBV3 in the upstream catchment	%	Yes
us_Bare	Percentage landcover in bare land based on LCDBV3 in the upstream catchment	%	Yes

*The variables usHardness and usParticleSize describe the physical character of the catchment regolith and usPhosphorus and usCalcium characterise its fertility based on values assigned to LRI top-rock categories by (Leathwick et al., 2003).

3.4 A priori classifications

We defined six additional classifications that were based on combinations of REC categories and the slope classification used by the original classification recommended by Snelder (2015). We refer to these as *a priori* classifications because they were not informed by the water quality data in the way that the cluster based classifications were. Rather we used prior knowledge of the likely drivers of river water quality variation across Northland to formulate alternative classifications. The classifications used the original slope classification and existing REC-based categories to discriminate differences in catchment topography and geology.

These *a priori* classifications have the advantage that their definition is less complicated than the statistically derived classifications. A description of the six *a priori* classifications are provided in Table 4.

Table 4. The a priori classifications based on REC attribute combinations

Name	Description
SlopeClassification	This classification represents the original classification recommended by Snelder (2015). REC version 1 upstream slope attribute divided into two classes with stream segments above 10º in the "1" class, and those below in class "0".
GeologyClassification	REC version 1 Geology level divided into two classes, with soft sedimentary, volcanic acidic and volcanic basic in class "1" and all other geology categories in class "0" This distinction recognises that class 1 geology is expected to be naturally more enriched with nutrients than class 2 (Snelder and Biggs, 2002).
TopoClassification	REC version 1 Source of Flow level divided into two with lake and lowland sourced rivers in class "1", and hill sourced rivers in class "0".
SlopeGeologyClassification	A four class classification created by combining the above slope and geology classifications.
SlopeTopoClassification	A four class classification created by combining the above slope and source of flow classifications.
GeolTopoClassification	A four class classification created by combining the above geology and source of flow classifications

3.5 Assessment of cluster-based classification performance

The ability of all classifications to explain variation in each of the observed water variables was assessed using analysis of variance (ANOVA). The ANOVA r^2 value was used as a measure of discrimination of water quality variation. This resulted in nine ANOVA tests for each classification (one for each of the water quality variables represented in the raw dataset).

In addition, we applied a multi-variate form of ANOVA (ADONIS) to the whole water quality dataset in a single analysis. ADONIS uses a dissimilarity matrix as the input, which we calculated by applying a Euclidean distance measure to the raw water quality data (Anderson, 2001). The ADONIS r^2 was used as a measure of discrimination of multivariate water quality variation.

3.6 Evaluating spatial models of cluster-based classifications

3.6.1 Spatial model performance

The performance of the spatial modelling of the cluster-based classifications was evaluated by the fitted misclassification rates of the respective LDA, CART and RF models. The misclassification rate is the proportion of sites that are incorrectly allocated to a class by the model. The fitted misclassification rate describes how well the model predicted the class of sites in the fitting dataset. As for all statistical models, the fitted performance is higher than the expected performance of the model to predict the class of sites that are not part of fitting dataset. We note that a more objective test of performance would be to use a cross validation procedure to assess the performance when individual sites are held out of the fitting data. This analysis was beyond the scope of this study and we assumed that relative difference in performance of the methods was indicated by the fitted misclassification rates.

3.6.2 Relationships with predictors

In addition to the performance of the different spatial models, the model structures, and the relative importance of the model predictors can provide insight into the environmental characteristics that discriminate the clusters. We examined these relationships in three ways. First, we assessed the average value of each predictor in each class. This analysis indicates how the classes differ environmentally from each other and where the classes sit on the environmental gradient represented by the predictors. Second, for e RF models we examined the importance measure for all predictors included in the model. RF model importance quantifies the contribution of each predictor to the model prediction accuracy (Cutler et al., 2007). Third, we used partial dependence plots (PDPs) to describe the fitted predictor-response relationships (Cutler et al., 2007). A PDP shows the marginal contribution of a predictor to a response variable and can be interpreted as an approximation of the fitted predictor-response relationships.

4 Results

4.1 Water quality classification

4.1.1 Characterisation of regional water quality variation

There was considerable variation in the site median values of the water quality variables (Table 1). This variation was summarised by a PCA performed on the site median log transformed and scaled water quality variables. A biplot of the rotated PCA indicates that different groups of variables were correlated (Figure 2). The variables clarity, dissolved reactive phosphorus and turbidity were correlated (clarity inversely to dissolved reactive phosphorus and turbidity) and were most correlated with the first principle component. The nitrogen species and *E. coli* were most correlated with the second axis and each other. This indicates that the two groups of variables (i.e. clarity, dissolved reactive phosphorus and Turbidity versus the nitrogen species and *E. coli*) were relatively independent of each other (Figure 2).

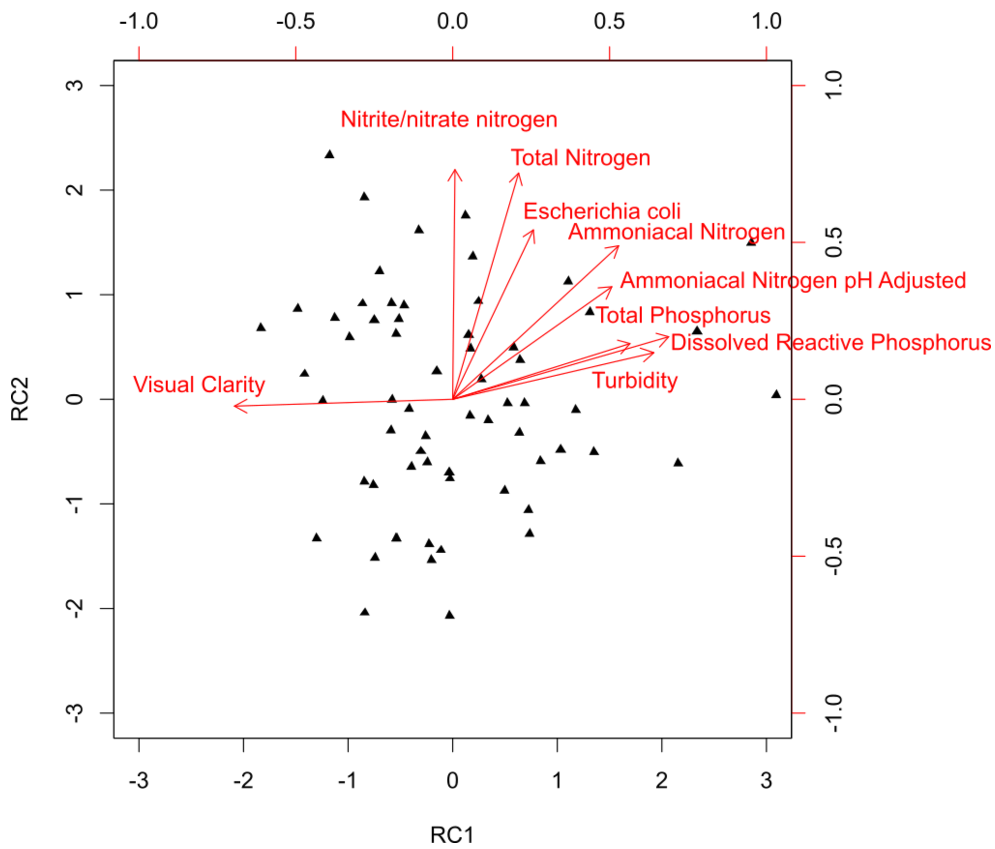

Figure 2. Biplot of the rotated PCA performed on the log transformed and scaled water quality variables.

4.1.2 Explanation of variation in regional water quality by cluster models

The ability of all site classifications to explain variation in each of the observed water variables is described by ANOVA r^2 and ADONIS r^2 values shown in Table 5. For the three sets of cluster-based classifications there was a general and expected pattern of increasing r^2 values with increasing number of classes. Another clear pattern was the superior performance of the cluster-based classifications compared to the REC-based classifications.

The r^2 values for the raw two class cluster-based classifications were all superior to the original water quality management classification based on two slope categories. This confirms the assumption that a cluster-based classification is likely to better discriminate water quality than the original water quality management classification.

The rotated PCA version of the cluster-based classification (PCARot) had lower r^2 values for most water quality variables than the cluster-based classification based on the raw data (Raw). This was true for all levels of detail (2, 3 and 4 clusters). The PCARot cluster-based classification had higher r^2 values for some variables (e.g. Clarity at the 4-cluster level, Nitrate-nitrogen across all levels). This indicates that there was a small effect of correlation on the relative weighting of variables associated with the clustering based on the Raw data.

The weighted rotated PCA version of the cluster-based classification (RotWt) had higher r^2 values for Clarity, Turbidity and Total Phosphorus than the classification based on the raw data (Raw). The increased discrimination of these three variables as a result of weighting the Clarity gradient arise because these variables are correlated (Figure 2). All other variables had lower r^2 values than the cluster-based classification based on the raw data (Raw). This is an expected outcome that demonstrates that enhancing the discrimination of a specific water quality gradient comes at the expense of the discrimination of the variables that are not correlated with that gradient.

Table 5. Explanation of the water quality observations by the site classifications.The table shows r^2 values for ANOVA tests performed on the individual water quality variables and r^2 values for the ADONIS test. The r^2 values are coloured (relatively) from highest (red) to lowest (green).

	Cluster-based classifications									*a priori* classifications					
Water quality variable	Raw2Clusters	Raw3Clusters	Raw4Clusters	PCARot2Clusters	PCARot3Clusters	PCARot4Clusters	RotWt2Clusters	RotWt3Clusters	RotWt4Clusters	SlopeClassification	GeologyClassification	TopoClassification	SlopeGeologyClassification	SlopeTopoClassification	GeolTopoClassification
Ammoniacal nitrogen	0.46	0.61	0.72	0.43	0.48	0.6	0.24	0.39	0.48	0.18	-0.02	0.01	0.16	0.17	-0.01
Dissolved reactive phosphorus	0.26	0.4	0.52	0.15	0.26	0.39	0.26	0.47	0.44	0.04	-0.02	0.02	0	0.04	0
Escherichia coli	0.25	0.3	0.35	0.26	0.21	0.27	-0.01	0.03	0.1	0.03	-0.01	0.09	0.06	0.1	0.09
Nitrite/nitrate nitrogen	0.37	0.42	0.49	0.53	0.62	0.65	-0.02	0	0.18	0.33	0.07	0.02	0.52	0.32	0.08
Total nitrogen	0.5	0.52	0.66	0.55	0.52	0.62	0.04	0.09	0.27	0.28	0	0	0.35	0.27	-0.01
Total phosphorus	0.34	0.58	0.6	0.2	0.46	0.56	0.45	0.65	0.63	0.06	-0.01	0.03	0.04	0.08	0.02
Turbidity	0.26	0.38	0.44	0.13	0.45	0.57	0.47	0.48	0.62	0.05	-0.01	0.01	0.03	0.04	-0.01
Ammoniacal nitrogen	0.32	0.53	0.58	0.24	0.34	0.45	0.21	0.43	0.42	0.05	-0.01	-0.01	0.04	0.04	-0.01
Visual clarity	0.16	0.44	0.4	0.04	0.38	0.53	0.47	0.46	0.7	0.02	-0.02	0.02	-0.01	0.03	0
Mean of r^2 values	0.33	0.46	0.53	0.28	0.41	0.52	0.24	0.33	0.43	0.11	0	0.02	0.13	0.12	0.02
ADONIS	0.37	0.53	0.6	0.33	0.47	0.57	0.28	0.4	0.5	0.15	0.01	0.04	0.19	0.17	0.05

4.1.3 Water quality characteristics of the classifications

The water quality characteristics of each class for each of the cluster-based classifications are summarised by the mean of the site medians for each water quality variable. Table 6 shows the water quality characteristics for two classifications: Raw3Clusters and Raw4Cluster (the results for all classifications have been provided as supplementary data). A map of the 63 water quality monitoring sites coded according to their membership of the Raw3Clusters classification is shown in Figure 3. Table 6 indicates that class 2 of the Raw3Clusters classification comprised rivers and streams that are characterised by relatively poor water quality (i.e. compared to the region). Class 2 had the highest mean median site concentrations of all water quality variables, except for Nitrite/nitrate-nitrogen, which was highest for class 1. Class 3 had the highest water quality overall with lower concentrations of all chemical variables and *E. coli*. For Raw4Clusters classification, the overall ranking of the classes from highest to lowest water quality is 2,1,4,3. The ranking is generally the same for the individual variables, except for DRP and TP where classes 2 and 1 switch, and Nitrite/nitrate nitrogen where classes 3 and 4 switch.

A complete version of this table for all cluster classifications is provided in the supplementary file called "AllWQVariableMedians.csv".

Table 6. Water quality variable medians for each cluster in the Raw3Clusters and Raw4Clusters classifications. See Table 1 for details of the water quality variables including measurement units. Values are coloured (relatively) from highest water quality (green) to poorest water quality (red).

Classification:	***Raw3Clusters***			***Raw4Clusters***			
Cluster class:	***1***	***2***	***3***	***1***	***2***	***3***	***4***
Number of sites:	***33***	***6***	***24***	***21***	***8***	***6***	***28***
Water quality variable							
Ammoniacal nitrogen	*0.015*	*0.044*	*0.0068*	*0.01*	*0.005*	*0.044*	*0.017*
Dissolved reactive phosphorus	*0.016*	*0.054*	*0.0082*	*0.0075*	*0.018*	*0.054*	*0.016*
Escherichia coli	*330*	*600*	*200*	*240*	*170*	*600*	*340*
Nitrite/nitrate nitrogen	*0.34*	*0.3*	*0.038*	*0.11*	*0.014*	*0.3*	*0.36*
Total nitrogen	*0.55*	*0.91*	*0.22*	*0.32*	*0.15*	*0.91*	*0.57*
Total phosphorus	*0.033*	*0.12*	*0.021*	*0.021*	*0.025*	*0.12*	*0.039*
Turbidity	*4.9*	*9*	*3.2*	*3.7*	*2.1*	*9*	*4.9*
Ammoniacal nitrogen	*0.0073*	*0.018*	*0.0051*	*0.0055*	*0.0042*	*0.018*	*0.0084*
Clarity	*1.2*	*0.71*	*1.6*	*1.4*	*1.8*	*0.71*	*1.2*

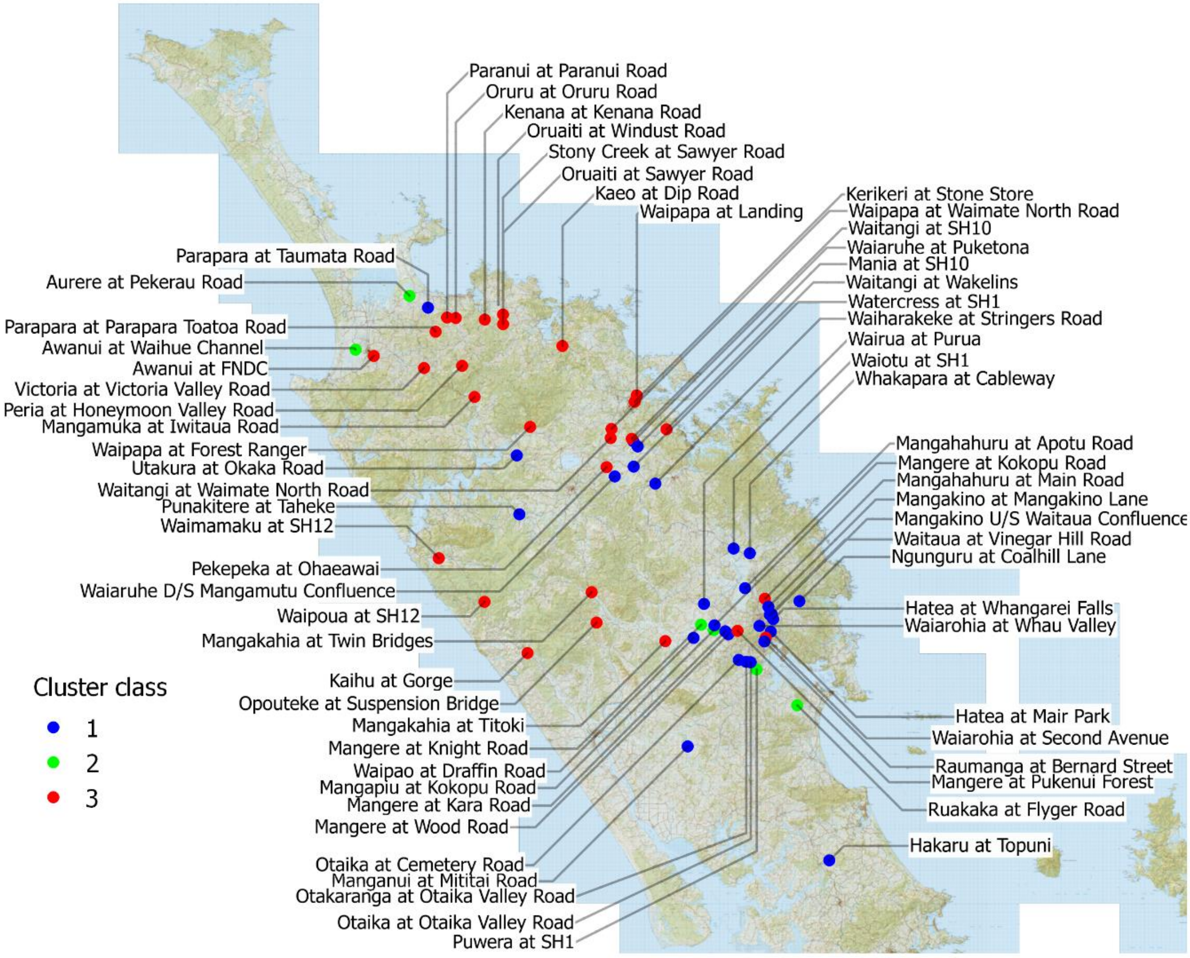

Figure 3. Map of the 63 water quality monitoring sites coded according to their membership of the Raw3Clusters classification.

4.2 Spatial models

4.2.1 Model performance

The performance of the spatial models is quantified by the fitted misclassification rates shown in Table *7*. The random forest model returned the lowest misclassification rates (Table *7*). The performance (i.e. misclassification rate) for the RF models were not affected by the exclusion of the land cover (LC) predictors. The performance of the LDA and CART models was decreased by a small margin for most of the classifications.

The CART models failed to discriminate some classes (i.e. sites in the non-discriminated class were assigned to alternative (incorrect) classes by the CART model). For example, class 2 of the Raw3Clusters classification was not discriminated by the CART model. This occurred regardless of whether the predictors included the land cover predictors. This results in some classes not being represented on the maps that are based on predictions from these models. See the supplementary output “ClassificationPlots.PDF” and Figure 5.

Table 7. Fitted misclassification rates (% of sites) of the spatial models of the cluster-based classifications.

Model and predictors	Cluster-based classification								
	Raw2Clusters	Raw3Clusters	Raw4Clusters	PCARot2Clusters	PCARot3Clusters	PCARot4Clusters	RotWt2Clusters	RotWt3Clusters	RotWt4Clusters
RF-All	0	0	0	0	0	0	0	0	0
RF-noLC	0	0	0	0	0	0	0	0	0
LDA-All	6.4	3.17	4.8	6.4	12.7	6.4	7.9	6.4	9.5
LDA-noLC	6.4	6.4	6.4	7.9	20.6	11.1	14.3	14.3	17.5
***CART-All**	12.7	17.5	55.6	9.5	19.1	19.1	12.7	15.9	31.8
***CART-noLC**	15.9	17.5	30.2	11.1	23.8	22.2	12.7	20.6	30.2

* Note that the CART models failed to discriminate some classes.

4.2.2 Relationships with predictors

Table 8 shows the 10 most important predictors for the RF models that used only "inherent" catchment characteristics as predictors (i.e. no land cover predictors). The most important predictors used by the RF models (over all models) were us_hard and us_slope. This indicates aspects of catchment geology and catchment topography are strongly associated with water quality variation in the Northland region.

When land cover was used as a predictor the variables us_IntensiveAg also had high importance. However, model performance did not significantly decrease when the land cover predictors were excluded from the models (Table *7*). This indicates that variation in the proportion of catchment occupied by agricultural land cover (i.e. us_IntensiveAg) is strongly associated with the non-land cover variables (e.g. mean catchment slope, us_slope). This means the inclusion of us_IntensiveAg in the model is effectively redundant and that the proportion of catchment in agricultural land is strongly associated with the topography of catchments in Northland.

Table 8. Ranked importance of the top ten predictors within the Random Forest models. The results are for models that used only "inherent" catchment characteristics as predictors (i.e. no land cover predictors). The numbers show the importance rank within each model. The order of the predictors is based on the mean importance rank over all models. See

Table 3 for an explanation of each of the predictor variables. Colours range from green

Name	**Description**	**Units**	**Land cover predictor?**
us_catarea	Area of the ups*tream catchment*	m^2	NO
us_rain	Mean annual rainfall of the upstream catchment	mm	NO
us_pet	Mean annual potential evapotranspiration of the upstream catchment	mm/yr	NO
QMean	Mean annual stream flow	m^3/s	NO
seg_elev	Mean elevation of the river segment	M *ASL*	NO
us_lakePerc	Percentage of landcover in lakes in the upstream catchment	%	NO
us_elev	Mean elevation of the upstream catchment	M ASL	NO
us_slope	Mean slope of the upstream catchment	degrees	NO
us_tmin	Catchment average mean June air temperature	d*egrees C x 10*	NO
us_twarm	Catchment average mean January air temperature	degrees C x 10	NO
us_rnvar	Catchment average coefficient of variation of annual rainfall	mm/yr	NO
us_rd10	Catchment average frequency of rainfall > 10mm	days/mo	NO
us_rd20	Catchm*ent average frequency of rainfall > 20mm*	days/mo	NO
us_rd100	Catchment average frequency of rainfall > 100mm	days/mo	NO
**us_hard*	Catchment average induration or hardness value	Ordinal	NO
**us_phos*	Catchment average phosphorous	Ordinal	NO
**us_psize*	Catc*hment average particle size*	Ordinal	NO
**usCalcium*	Catchment average calcium	Ordinal	NO
us_IntensiveAg	Percentage landcover in intensive agriculture based on LCDBV3 in the upstream catchment	%	Yes
us_PastoralLight	Percentage landcover in light pasture bas*ed on LCDBV3 in* the upstream catchment	%	Yes
us_NativeForest	Percentage landcover in native forest based on LCDBV3 in the upstream catchment	%	Yes
us_Urban	Percentage urban landcover based on LCDBV3 in the upstream catchment	%	Yes
us_Scrub	Percentage l*andcover in scrub based on LCDBV3 in the* upstream catchment	%	Yes
us_Wetland	Percentage landcover in wetland based on LCDBV3 in the upstream catchment	%	Yes
us_ExoticForest	Percentage landcover in exotic forest based on LCDBV3 in the upstream catchment	%	Yes
us_Bare	Percentage landcover in bare land based on LCDBV3 in the upstream catchment	%	Yes

(rank=1) to red (rank=17)

Predictor variable	Cluster-based classification								
	Raw2Clusters	Raw3Clusters	Raw4Clusters	PCARot2Clusters	PCARot3Clusters	PCARot4Clusters	RotWt2Clusters	RotWt3Clusters	RotWt4Clusters
us_hard	2	1	1	9	1	4	1	1	2
us_slope	5	3	2	1	2	5	8	10	3
us_elev	4	2	6	2	4	2	9	8	7
us_rd10	6	11	5	4	5	1	3	2	12
us_phos	11	8	10	15	3	6	2	5	1
us_rain	1	7	3	3	9	7	11	4	17
us_rd20	3	6	4	11	14	3	12	3	9
us_twarm	8	5	9	7	12	10	7	11	4
us_rnvar	9	9	8	5	6	9	15	6	6
us_psize	10	4	11	14	8	8	5	9	5

Figure 4 shows the PDP for the random forest model of the Raw3Clusters classification. The PDP indicates that the marginal probability that a site belongs to class 2 of the classification decreases with increasing value of us_hard, us_slope and us_psize. The PDP indicates that the probability that a site belongs to class 2 of the classification increases with increasing value of us_twarm. By contrast, the PDP indicates that class 3 has the opposite relationships with the predictors to class 2 and class 1 has relationships that are intermediate to those of classes 2 and 3 (Figure 4).

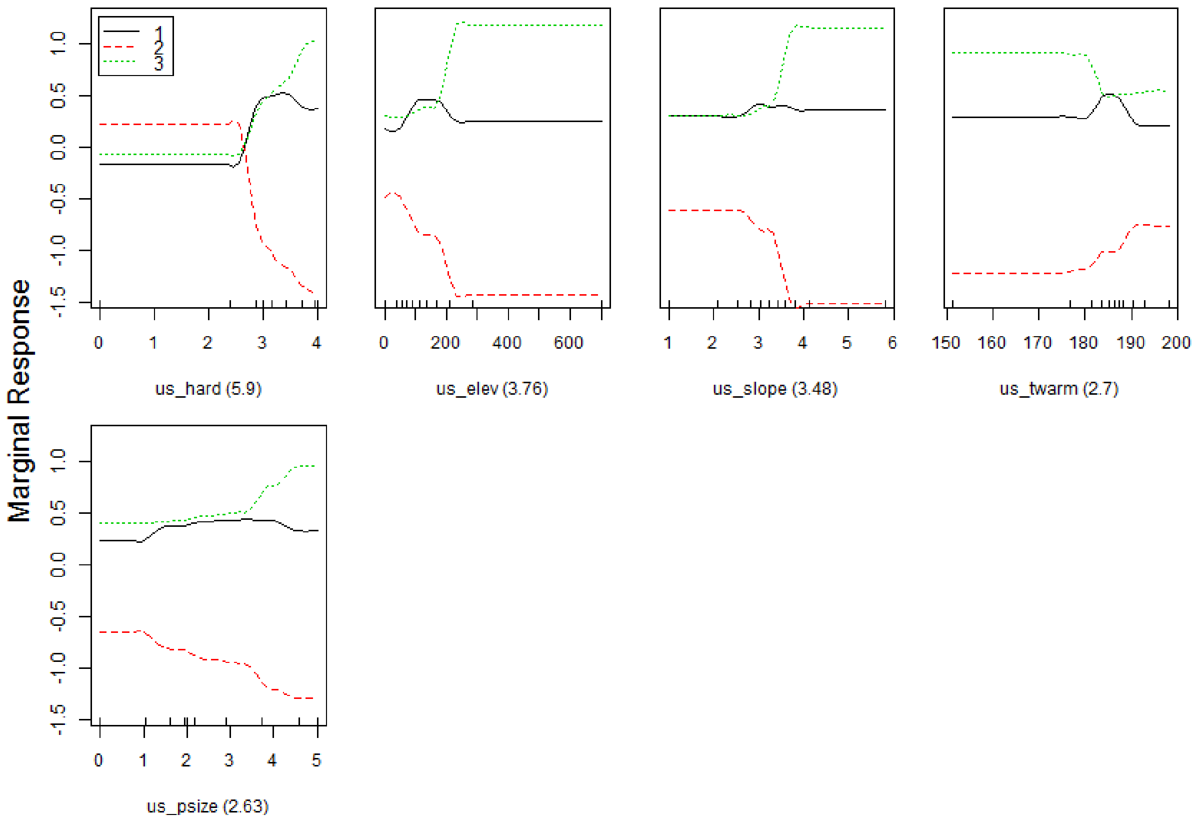

Figure 4. Partial plots of the top five predictors for the random forest model of the Raw3Clusters classification. See

Name	Description	Units	Land cover predictor?
us_catarea	Area of the up*stream catchment*	m^2	NO
us_rain	Mean annual rainfall of the upstream catchment	mm	NO
us_pet	Mean annual potential evapotranspiration of the upstream catchment	mm/yr	NO
QMean	Mean annual stream flow	m^3/s	NO
seg_elev	Mean elevation of the river segment	M *ASL*	NO
us_lakePerc	Percentage of landcover in lakes in the upstream catchment	%	NO
us_elev	Mean elevation of the upstream catchment	M ASL	NO
us_slope	Mean slope of the upstream catchment	degrees	NO
us_tmin	Catchment average mean June air temperature	*degrees C x 10*	NO
us_twarm	Catchment average mean January air temperature	degrees C x 10	NO
us_rnvar	Catchment average coefficient of variation of annual rainfall	mm/yr	NO
us_rd10	Catchment average frequency of rainfall > 10mm	days/mo	NO
us_rd20	Catchm*ent average frequency of rainfall > 20mm*	days/mo	NO
us_rd100	Catchment average frequency of rainfall > 100mm	days/mo	NO

**us_hard*	Catchment average induration or hardness value	Ordinal	NO
**us_phos*	Catchment average phosphorous	Ordinal	NO
**us_psize*	Catch*ment average particle size*	Ordinal	NO
**usCalcium*	Catchment average calcium	Ordinal	NO
us_IntensiveAg	Percentage landcover in intensive agriculture based on LCDBV3 in the upstream catchment	%	Yes
us_PastoralLight	Percentage landcover in light pasture ba*sed on LCDBV3 in* the upstream catchment	%	Yes
us_NativeForest	Percentage landcover in native forest based on LCDBV3 in the upstream catchment	%	Yes
us_Urban	Percentage urban landcover based on LCDBV3 in the upstream catchment	%	Yes
us_Scrub	Percentage l*andcover in scrub based on LCDBV3 in the* upstream catchment	%	Yes
us_Wetland	Percentage landcover in wetland based on LCDBV3 in the upstream catchment	%	Yes
us_ExoticForest	Percentage landcover in exotic forest based on LCDBV3 in the upstream catchment	%	Yes
us_Bare	Percentage landcover in bare land based on LCDBV3 in the upstream catchment	%	Yes

Table 3 for an explanation of each of the predictor variables. The numbers in brackets are the importance values of the predictors.

The environmental characteristics of the cluster-based classifications can also be evaluated by the mean value of the predictor variables for all segments that the spatial model has assigned to a particular class. Table 9 displays the mean values of the top eight predictors for the RF models of the Raw3Clusters and Raw4Clusters classifications (the results for all classifications have been provided as supplementary data). Table 9 shows for the Raw3Clusters classification that the classes are ranked (from highest to lowest) 2,1,3 in terms of mean catchment geological hardness, slope and rainfall (indicated by the predictors us_hard, us_slope, us_rain, us_rd20 and us_rd10). This ranking is 3, 4, 1 and 2 for the Raw4Clusters classification. It is noted that the numbers used to identify the class names are arbitrarily defined by the K-means clustering algorithm.

The environmental characterisation of the classifications provided by PDPs and tabulations of mean values of predictors can be used to describe the classes in narrative terms. For example, Figure 4 and Table 9 indicates that class 2 of the Raw3Clusters classification comprised rivers and streams with catchments that are characterised by relatively (i.e. compared to the region) soft geology, low slopes and low rainfall.

A complete version of this table for all cluster-based classifications and models with all predictors and just with fixed predictors, is provided in the supplementary file called 'AllPredictorMeans.csv'.

Table 9. Mean values of the Raw3Clusters and Raw4Clusters classifications for the top eight predictors for the Random Forest models. Cells are coloured from high (Green) to low (red).

Classification:	***Raw3Clusters***			***Raw4Clusters***			
Cluster class	***1***	***2***	***3***	***1***	***2***	***3***	***4***

Number of sites	*33*	*6*	*24*	*21*	*8*	*6*	*28*
Predictor (3 class rank,4 class rank,units)							
us_hard (1,1,ordinal)	*3.1*	*2*	*3.6*	*3.5*	*3.8*	*2*	*3.2*
us_elev (2,6,M ASL)	*100*	*51*	*201*	*196*	*284*	*50*	*99*
us_slope (3,2,degrees)	*3.1*	*2.2*	*3.6*	*3.3*	*4.2*	*2.2*	*3.1*
us_psize (4,11,ordinal)	*2.2*	*1.7*	*3.6*	*3.2*	*4.2*	*1.7*	*2.4*
us_twarm (5,9,ºC x 10)	*187*	*190*	*183*	*183*	*178*	*190*	*187*
us_rd20 (6,4,days/mo)	*1.6*	*1.4*	*1.8*	*1.9*	*1.7*	*1.4*	*1.5*
us_rain (7,3,mm)	*1499*	*1414*	*1643*	*1734*	*1621*	*1407*	*1477*
us_phos (8,10,ordinal)	*1.4*	*1.3*	*2.2*	*2*	*2.9*	*1.3*	*1.5*

4.2.3 Mapped classifications

Maps of example classifications are shown in Figure 5. The maps show the predictions for the Raw3Clusters classification made using the three modelling methods: LDA, CART and RF. The different modelling methods resulted in differences in the predictions and therefore the maps shown in Figure 5. Figure 5 shows that the mapped patterns are reasonably similar for RF and LDA but that the CART model failed to discriminate class 2. Figure 5 also indicates that the predictions and therefore the maps were very similar, but not identical, for RF models using all predictors or excluding the land cover predictors.

Maps of examples of the *a priori* classifications are shown in Figure 6. The classifications describe spatial patterns that are quite different to the cluster-based classifications. Maps of all classifications have been provided in the supplementary file called ‘ClassificationPlots.pdf’.

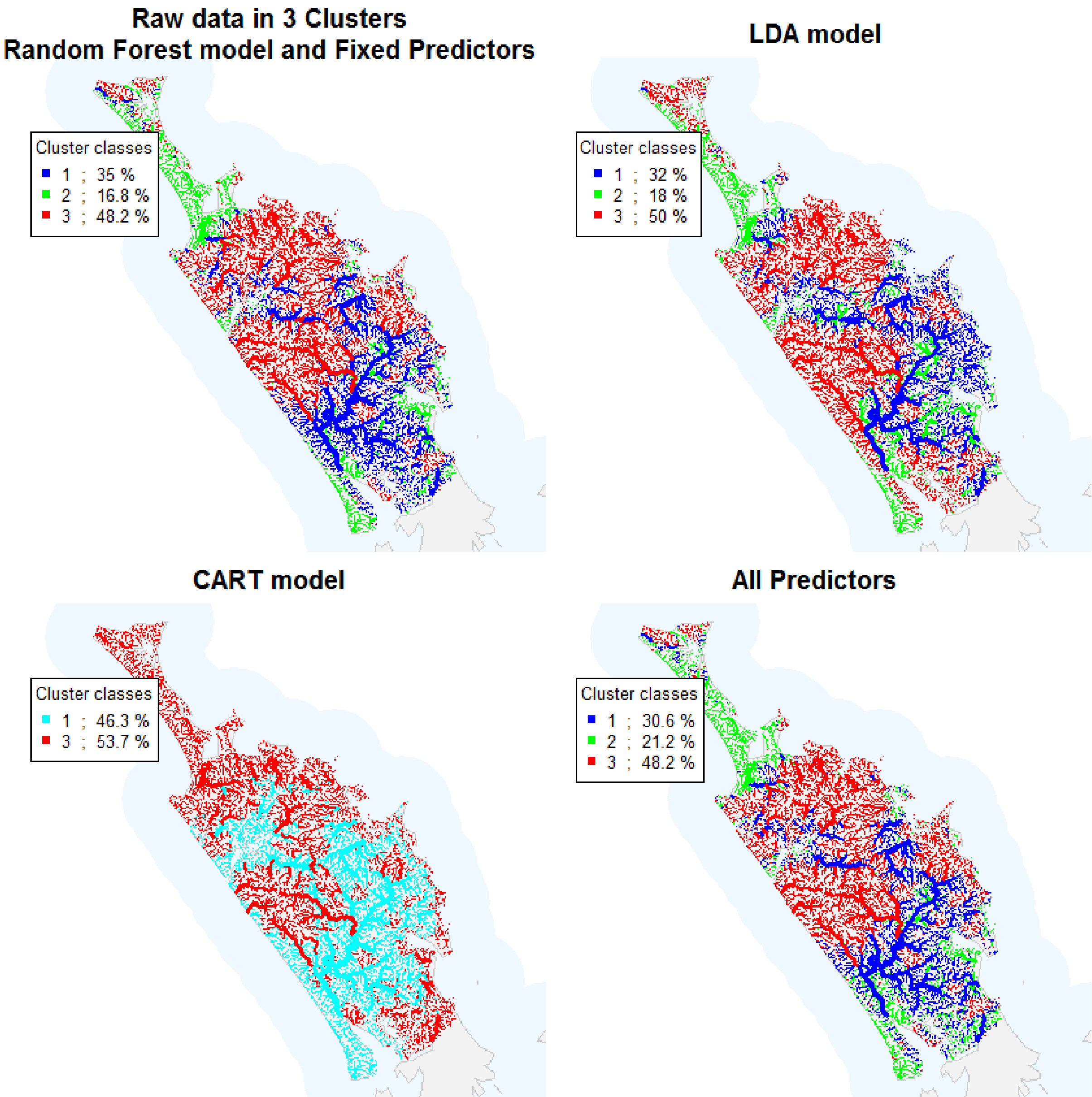

Figure 5. Maps of the Raw3Clusters classification made using predictions derived from the three modelling methods; RF, LDA, CART. The top left plot represents predictions made using the RF model that excluded the land cover predictors. The bottom right plot represents predictions made using the RF model included all predictors. Note that CART failed to discriminate class "2". See related comment associated with Table 7.

Figure 6. Maps of the a priori classifications. The Slope classification is the original water quality management classification recommended by Snelder (2015). The other classifications are described in Table 4.

5 Discussion

5.1 Discrimination of water quality variation by the classifications

The *a priori* (REC-based) classifications performed poorly compared to the cluster-based classifications. This is an expected result because the water quality data was used to define the cluster-based classification whereas the *a priori* classifications are independent of the data. This result indicates that a water management classification that is defined using clustering can be expected to discriminate regional water quality variation better than an *a* *priori* classification. None of the alternative *a priori* classifications performed significantly better than the original water quality management classification (i.e. two slope classes) recommended by Snelder (2015). It is therefore concluded that the *a priori* classifications are not good candidates for an alternative water quality management classification.

The *a priori* (REC-based) classifications are relatively easily explained compared to the cluster-based classifications. Therefore, the trade-off for the increased discrimination that is achieved by the cluster-based classifications is the increased complexity. Clustering is not easily understood by non-experts. In addition, as demonstrated by this study, the exact solution and its ability to discriminate variation in water quality depends on subjective decisions associated with the analysis. For example, this study demonstrated that inclusion of correlated variables or weighting of certain water quality gradients affects the clustering solution. These points are not easily explained and may make acceptance of a cluster-based water quality management classification difficult. However, it is important to acknowledge that two national scale environmental classifications that were derived using complex multivariate statistical methods are widely accepted and used: Land Environments of New Zealand (LENZ; Leathwick et al., 2003) and Freshwater Environments of New Zealand (FWENZ; Leathwick et al., 2011).

The study demonstrated that a significant increase in explanation of water quality variation can be achieved by moving from the simple two class slope-based classification recommended by Snelder (2015) to a cluster-based classification (Table 5). The study showed that the three-class cluster-based classification increased the explanation of water quality variation result compared to two classes (Table 5). In addition, the increase in discrimination (i.e. r^2) associated with increasing from 3 to 4 classes was small compared to the increase associated with increasing from 2 to 3 classes.

Use of principle component analysis did little to improve the discrimination of water quality compared to the cluster-based classification based on the raw data. The additional complexity introduced by adding PCA to the classification process would increase the difficulty of explaining and defending the classification.

The weighting of the first rotated principal component achieved the desired effect of improving the discrimination of the water clarity (and correlated) variables. However, in so doing, it reduced explanation of variation of those variables more closely associated with the second component. In addition, the additional complexity and subjectivity associated with weighting, would increase the difficulty of explaining and defending the methodology.

The Raw3Cluster classification defined using the raw (log transformed scaled site median water quality variables) is possibly the best candidate for an alternative water quality management classification. This classification is the simplest of the cluster-based classifications (i.e. fewest analytical steps). The classification's discrimination of variation of all water quality variables was superior to the original water quality classification recommended by Snelder (2015) (Table 5). The choice of three classes is subjective. It could be justified on the basis that there is a large increase in discrimination of water quality going from two to three classes but that the increase from three to four is less pronounced.

5.2 Spatial modelling of classifications

Statistical modelling methods were used to spatially interpolate the cluster-based classification to all network segments in the Northland river network. The study found that RF modelling produced the most accurate method for the spatial interpolation. RF models are more complicated and difficult to understand than the alternative LDA and CART models. However, the results of the RF modelling can be described by maps of the classifications and tables describing the environmental characteristics of classes (as can the results of LDA and CART modelling). It is suggested that RF modelling is a good candidate for defining an alternative water quality management classification.

The statistical models were not strongly improved by including land cover predictors and no improvement was achieved for the RF models (Table 7). This result suggests land use patterns in Northland are strongly determined by "inherent" catchment characteristics such as climate, topography, and geology. The cluster-based classifications can be accurately spatially interpolated using the RF models based only on "inherent" catchment characteristics. This means the assignment of any stream or river segment is 'independent' of land use in its catchment and that the classification would not favour or disfavour locations because of current land use.

Our statistical approach to defining the classifications was limited by the number of sites with water quality data and the representativeness of these sites of regional water quality patterns. As a result, some environments that may be regarded as distinctive regionally will not be represented by the classifications. These environments will be included in classes that are "close" with respect to the predictor variables that were included in the spatial models but this inclusion might not be considered appropriate. An example of this is Class 2 of the Raw3Clusters classification (green on Figure 7). Class 2 results from a cluster that includes six water quality monitoring sites that had the poorest overall water quality (Table 6). These sites were associated with low values of the us_slope, us_elev, us_hard and us_psize (i.e. lowland catchments with softer and finer grained regolith (Table 9). Much of the Aupouri and Pouto Peninsulas are classified as Class 2 because these areas are characterised by lowland catchments with softer and finer grained regolith. However, these specific areas are characterised by low intensity agriculture and plantation forestry and have few permanently flowing rivers and streams. Their inclusion in Class 2 of the Raw3Clusters classification is associated with the absence of monitoring sites and is probably inappropriate. This issue cannot be over-come statistically or objectively due to the lack of site data. We suggest therefore that if any of the statistically defined classifications presented here were to be used, some refinement of the classification may be appropriate. The refinement could be carried out by incorporating 'local knowledge' which is not reflected in the site data and spatial modelling.

6 Conclusions

The information presented in this report and the associated supplementary data provide a basis for NRC to consider alternatives for the water quality management classification. Aspects that should be considered include the discrimination of variation in water quality variation (see Table 5) and the performance of the spatial model (see Table 7). An important consideration is whether the increase in the discrimination of variation in water quality variation over the original classification is sufficiently advantageous that it outweighs the increased complexity and difficulty in explaining the classification. The decision to adopt an alternative classification is a subjective (political) one.

This report does not recommend adoption of any alternative but does provide the information to assist that decision. The combination of analyses and tables presented in this report allow all the derived classifications, including their explanation of water quality variation, the performance of the spatial model and the water quality and environmental character of the classes to be understood in narrative terms. To demonstrate this, consider the Raw3Clusters classification that is spatially interpolated using an RF model (Figure 7). This is a good candidate for an alternative water quality management classification given its good explanation of water quality variation (Table 5), good performance of the spatial model (Table *7*) and independence from land cover.

The analyses and tables presented in this report enable the characteristic of the 3 classes of the Raw3Clusters to be defined. Table 6 indicates that Class 3 (red on Figure 7) had the highest overall water quality. Table 9 indicates that Class 3 comprises rivers and streams with catchments that are characterised by relatively (i.e. compared to the region) hard geology, steep slopes and high rainfall.

Table 6 indicates that Class 1 (blue on Figure 7) is characterised by most water quality variables being between Class 2 and Class 3. Table 9 indicates that rivers and streams in Class 1 are similar to Class 2 but are generally higher and steeper (i.e., higher values of us_slope and us_elev; Table 9) and differences in geology (i.e., Class 1 has higher values of us_psize and us_hard; Table 9).

Table 6 indicates that Class 2 (green on Figure 7) is characterised by high concentrations of *E. coli*, total nitrogen and ammoniacal nitrogen. Table 9 indicates that rivers and streams in Class 2 are characterised by relatively (i.e. compared to the region) soft geology (i.e. low values of us_hard; Table 9), low slopes and elevation, warm temperatures, and low rainfall. The catchments of rivers in this class tend to be associated with the most intensive agriculture. The rivers in this class are associated with poor water quality that, at least partly, reflects that resource use.

It is noted that the Aupouri and Pouto Peninsulas are classified as Class 2 but that this results from the absence of water quality monitoring sites in these areas and may not be appropriate. We recommend therefore that if any of the statistically defined classifications presented here were to be used, some refinement of the classification may be appropriate. The refinement could be carried out by incorporating 'local knowledge' which is not reflected in the site data and spatial modelling. Care would need to be taken to avoid generating many small-scale classes based on local knowledge as this would undermine the general objective of providing a succinct classification for management purposes.

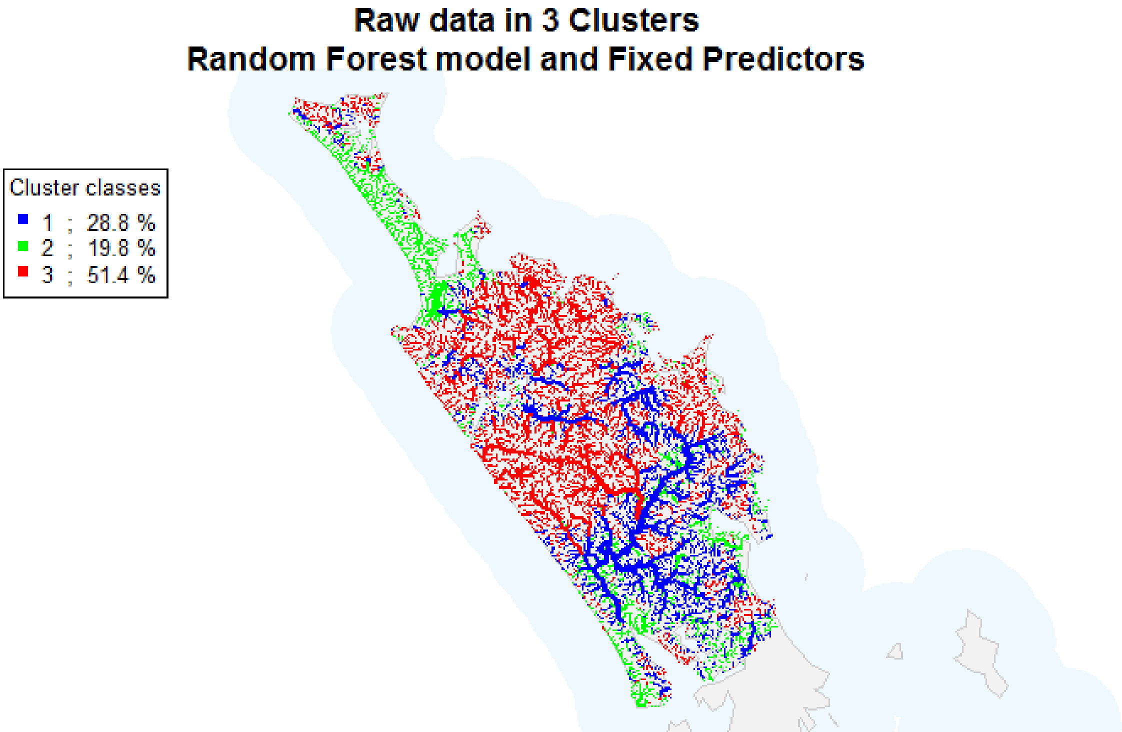

Figure 7. Map of the Raw3Clusters classification spatially interpolated using a Random Forest Model and fixed (i.e. non-land cover) predictors. This classification is a good candidate for an alternative water quality management classification.

7 References

Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral ecology 26, 32–46.

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32.

Breiman, L., Friedman, J.H., Olshen, R., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Belmont, California.

Crawley, M.J., 2002. Statistical computing: an introduction to data analysis using S-Plus. John Wiley & Sons Inc.

Cutler, D.R., Edwards, J.T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792.

Larned, S.T., Snelder, T.H., Unwin, M., McBride, G.B., 2016. Water quality in New Zealand rivers: current state and trends. NZJMFR.

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.

Leathwick, J.R., Overton, J.M., McLeod, M., 2003. An environmental domain analysis of New Zealand, and its application to biodiversity conservation. Conservation Biology 17, 1612–1623.

Snelder, T., 2015. Defining Freshwater Management Units for Northland: A Recommended Approach (LWP Client Report). LWP Ltd, Christchurch, New Zealand.

Snelder, T.H., Biggs, B.J.F., 2002. Multi-scale river environment classification for water resources management. Journal of the American Water Resources Association 38, 1225–1240.