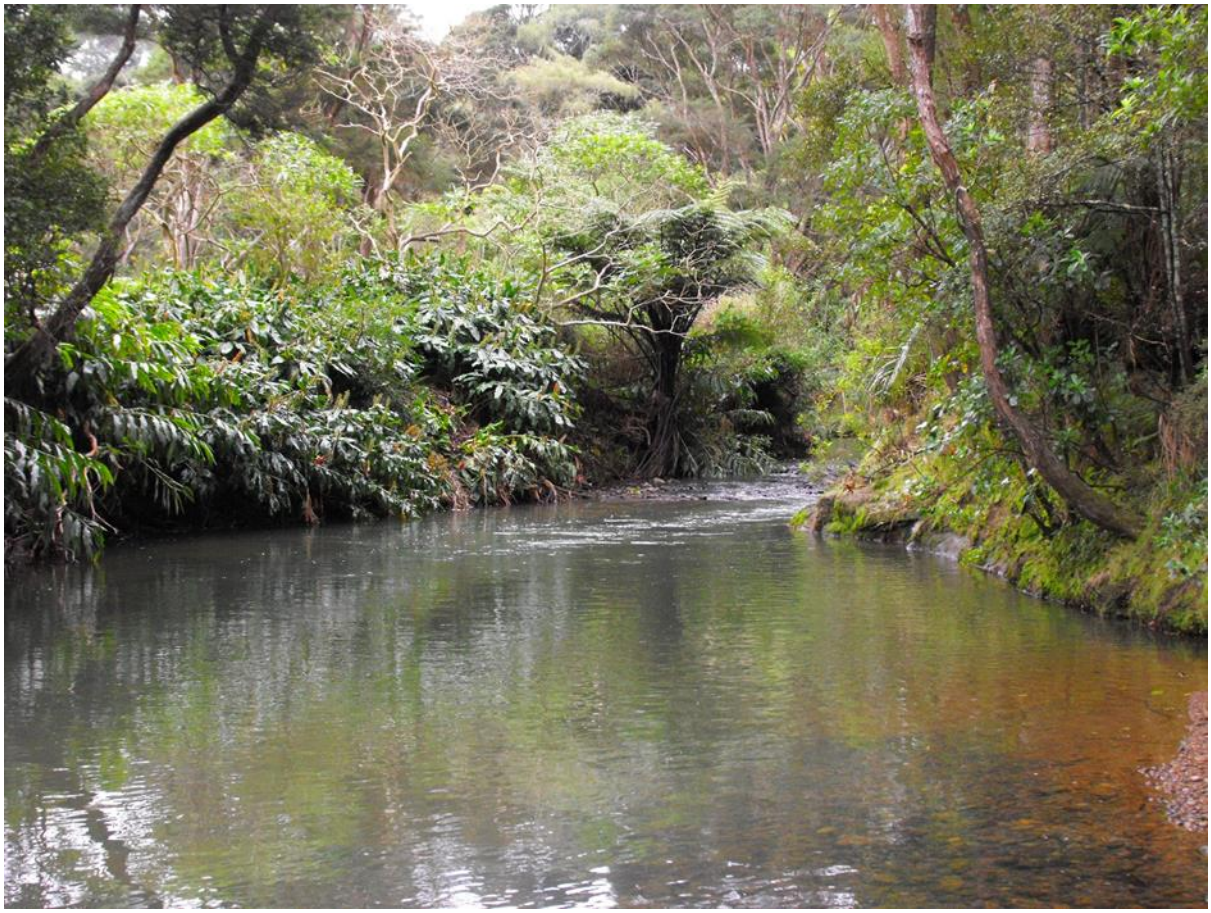
October 2018

Predicting stream macroinvertebrate reference conditions using a multivariate model – Phase 2



Predicting stream macroinvertebrate reference conditions using a multivariate model – Phase 2

Martin Neale

Prepared by Puhoi Stour Limited for the Ministry for the Environment under sub-contract to Cawthron Institute. [MfE Contract 18502]

October 2018

Executive summary

It is important in bioassessment systems that indices provide consistent meaning irrespective of geographic location or stream type. This issue is commonly addressed by the reference condition approach, whereby indices are benchmarked against appropriate reference conditions and the deviation from reference indicates the degree of impairment.

Such approaches have been used internationally for over 40 years but have not been widely adopted in New Zealand. There were some regional trials in the 1990s and 2000s, but large-scale applications have been lacking. However, in 2017 the Ministry for the Environment funded a proof-of-concept study as part of a wider project on macroinvertebrate indicators of ecosystem health. That work included the development of national scale predictive models based on biological and environmental classifications. The predictive performance of the biological model was similar to that observed for other multivariate models developed internationally and the environmental model performed equally as well.

The proof-of-concept study represented the first attempt to develop a national scale application of the reference condition in New Zealand, and whilst initial assessments of performance were promising, there remained several opportunities to improve performance. This work sought to pursue some of these opportunities to develop a model that could be used in an operational manner, including to expand the reference site training dataset, optimise model build, include additional metrics and test the model with independent data.

Unfortunately, much of the reference site data used in the proof-of-concept study was not available for this project and rather than building on an existing reference site dataset, the compilation of a new reference dataset from publicly available data was required. This newly compiled dataset was substantially smaller than the dataset used for the proof-of-concept study (359 sites compared with 538 sites) and was also less representative of the New Zealand river network.

The limitations of the dataset had a large effect on the performance of the models built using it. The biological model developed here had poorer predictive performance than the proof-of-concept model in a like-for-like comparison. However, when addional tests were carried out, involving exclusion of under-represented river types and using invertebrate data not used in the development of the model, the predictive performance of the biological model was similar to the proof-of-concept model.

The future use of a reference condition model in New Zealand is contingent on the availability of addional reference site data. A large and representative reference site dataset is foundational for a predictive model and it is important that natural variability is adequately captured to create a model that performs well across the heterogeneous environment of New Zealand. Consequently, the key recommendation from this work is to increase the size of the reference site dataset. This could include seeking agreement to use some of the privately collected data used in the proof-of-concept study, but will also likely involve the collection of new invertebrate community data from under-represented stream types. This data collection may be informed by the analysis of FENZ classifications undertaken here and may be done in conjunction with the further development of stressor specific metrics.

When a suitable reference site dataset is available, the approach described here to develop predictive models can be re-applied to produce a bioassessment tool that may be used in operational monitoring and reporting programmes for stream ecological health.

Table of contents

List of tables

List of figures

List of Appendices

1. Introduction

1.1 Reference condition

A major challenge for bioassessment is ensuring that any index provides consistent meaning in different environmental settings; that is, a given score from an index should indicate the same biological condition irrespective of geographic location or stream type (Mazor et al. 2016). This issue is commonly addressed in bioassessment through the use of the reference condition approach (Reynoldson et al. 1997), which attempts to distinguish natural variability from anthropogenic impacts by comparing the biological attributes from a test site with a group of similar sites that are in a minimally-disturbed reference condition.

The concept has become central to many bioassessment programmes, where the target is referred to as the 'expected' (E) value and the result from a test site is referred to as the 'observed' (O) value. The ratio (O/E) of these values is now commonly used as a basis for bioassessment, with values close to 1 indicating sites close to reference condition and values closer to 0 indicating impaired sites. Importantly, the concept of O/E can be applied to any measure or metric of the stream macroinvertebrate community.

There have been two broad approaches for predicting the reference condition (E) for use in bioassessment programmes, with the key difference being how the reference sites are used to support the prediction of the expected fauna (E). Predictions are either based on an 'environmental' classification of stream types or a model based on multivariate 'biological' data that uses environmental attributes to predict the macroinvertebrate community. The key differences between these two approaches are described in Clapcott et al (2017).

The biological approach has been used extensively in freshwater bioassessment programmes around the world; for example, the approach was first used in the U.K. (Moss et al., 1987; Wright et al., 1993), with subsequent applications in Australia (Parsons & Norris, 1996), North America (Reynoldson et al., 1995, Hawkins et al., 2000a; Mazor et al., 2016), Sweden (Johnson et al., 2003), Czech Republic (Kŏkes et al., 2006) and Spain (Poquet et al., 2009). Where the performance of the two approaches has been compared, the biological approach has generally provided greater predictive accuracy than the environmental approach (Reynoldson et al., 1997; Hawkins et al., 2000b; van Sickle et al., 2005; Davy-Bowker et al., 2006; Mazor et al., 2006; Neale & Rippey, 2008).

1.2 New Zealand application

The biological approach has been trialled previously at regional scales in New Zealand (e.g. Waikato (Coysh & Norris 1999; Death & Collier 2010) and Manawatu-Wanganui (Joy & Death 2003)), but until recently the approach had not been applied at a national scale.

The Ministry for the Environment funded a project on benthic macroinvertebrate indicators of ecosystem health (Contract 21630), which included a work package to explore the use of a multivariate approach to assessing ecosystem health. This involved the development of a proof-of-concept biological classification model at the national scale based on reference site dataset that was compiled from a range of public and private data sources. The resulting biological classification was used to predict the invertebrate reference condition for all segments in the training dataset (Clapcott et al., 2017).

The predictive accuracy of the biological classification approach was similar to that observed for other multivariate models developed overseas. The performance of the biological approach was compared

with that of a biologically-optimised environmental classification approach - the Freshwater Ecosystems of New Zealand (FENZ) classification of sites (Leathwick et al., 2010), which performed equally well.

The proof-of-concept study represented the early development of a predictive model for assessing macroinvertebrate communities in New Zealand rivers and identified a series of potential options to improve the biological model for assessing ecosystem health. Improvements could be achieved by:

- Compiling a larger reference site training dataset,
- optimising model build, including testing clustering options and environmental predictors,
- include additional metrics (beyond number of taxa and MCI), and
- testing the model with independent data.

1.3 Scope

The outputs of the proof-of-concept study informed the objectives of this scope of work, which include:

- Compile a reference site dataset based only on public data and identify under-represented river types
- Construct a biological and environmental model with the new dataset and test options in the model build to optimise performance
- Test model performance within the training dataset and with independent data

Hence, in this report I describe the steps taken to meet the study objectives. The relative performance of the environmental and biological approaches is presented as a comparison with a null model (i.e. a model that makes no attempt to explain variability in communities (van Sickle et al. 2005)). Based on the analyses completed for this work package, recommendations are made on the suitability of the multivariate approach to complement existing stream health assessment.

2. Methods

2.1 Dataset

Much of the data that was used in the proof-of-concept study was from private research datasets and unfortunately was not available for this project. Therefore, a new dataset of reference sites from public data sources was compiled, which was primarily sourced from Regional Council monitoring programmes.

This compilation began with a dataset compiled by Martin Unwin (NIWA) for the MfE sediment project (Depree et al., 2017) and subsequently groomed by Annika Wagenoff (Cawthron) for the previous MfE macroinvertebrate project (Clapcott et al., 2017). This dataset included records of 178 taxa, for 15,508 samples collected between 1994 and 2016, from 1,770 sites. I extracted sites from this dataset that had a coverage of 90% or greater of native forest in the upstream catchment (based on the FENZ measure USNATIVE), which yielded an initial 220 reference sites. Where repeat samples were available for sites (i.e. multiple years), the most recent sample was used for model build and selected older samples were used to test model performance.

During a stakeholder workshop at the start of this work programme, discussion with Regional Council staff indicated that there was likely to be additional reference site data that had been collected after the data compilation carried out for the MfE sediment project. A subsequent data request to Regional

Council staff yielded a further 61 reference sites that met the 90% native forest criteria. This included 39 sites from Environment Southland, 16 sites from Hawke's Bay Regional Council, five sites from Northland Regional Council and a single site from Horizons Regional Council.

Finally, the public data from the proof-of-concept study dataset was added to the Regional Council data, which yielded a further 78 reference sites that met the 90% native forest criteria. This included 38 sites from Stewart Island (Chadderton, 1988), 33 sites from upland streams in the South Island (Richard Storey, unpublished data) and seven sites from Northland (Dave West, unpublished data).

Due to inconsistencies in the invertebrate data from the different sources, the taxonomic information was standardised prior to analysis. This involved transforming the occurrence data into presence-absence (the base data was a mixture of counts, relative abundance categories and presence-absence). The taxonomic information was generally provided, and used, at MCI-level (i.e. mainly genus). However, records of Chironomidae were aggregated to family level because of inconsistencies in resolution amongst datasets (i.e. genus, tribe and sub-family were all present in the datasets).

This process resulted in a dataset of 359 reference sites (Figure 1) for model development and the representativeness of these sites was assessed using the FENZ classification of river types.

2.2 Model development

Predictive models were constructed using the different site classification (biological and environmental) approaches and used to predict the expected fauna as follows.

2.2.1 Biological model

The predictions based on the biological classification of sites were made by constructing a River Invertebrate Prediction and Classification System (RIVPACS)-type model using the framework described by Clarke et al. (2003), although more recently developed statistical tools were utilised where appropriate. The model was constructed using R scripts provided by John van Sickle (US EPA).

First, biological classification of the reference site data was developed using agglomerative nesting (agnes) routine (Kaufman & Rousseeuw 1990) in the R package 'cluster' based on the Sorenson dissimilarity index. For the proof-of-concept study, the dendrogram from the clustering was used to visualise the classification and the select the number of end groups used in the model development (Clapcott et al., 2017). In this phase of model development, I used a more objective analysis to test the classification strength of different clustering options.

The classification strength of different clustering options was assessed by comparing the within-class biotic similarity relative to the between-class similarity (Van Sickle and Hughes, 2000; Hawkins and Norris, 2000). In a strong classification, similarities between sites that are in the same group should be substantially greater than similarities between sites that are in different groups. The ANOSIM (ANalysis Of SIMilarities) routine in the PRIMER v7 (Plymouth Routines In Multivariate Ecological Research (Clarke and Warwick, 1994)) software package was used to carry out this comparison of within- and between-group similarities (Clarke, 1993). The key output from this analysis is the *R* statistic, which ranges from -1 to 1. The *R* statistic equals 1 if all sites within classification groups are more similar to each other than any sites from different groups. Similarly, $R = 0$ if inter-site similarities between and within classification groups are on average the same. Although the *R* statistic is a useful comparative measure of the degree of separation of sites, a permutation statistical test was also carried out to determine whether the *R* statistic is significantly greater than zero (Clarke, 1993).

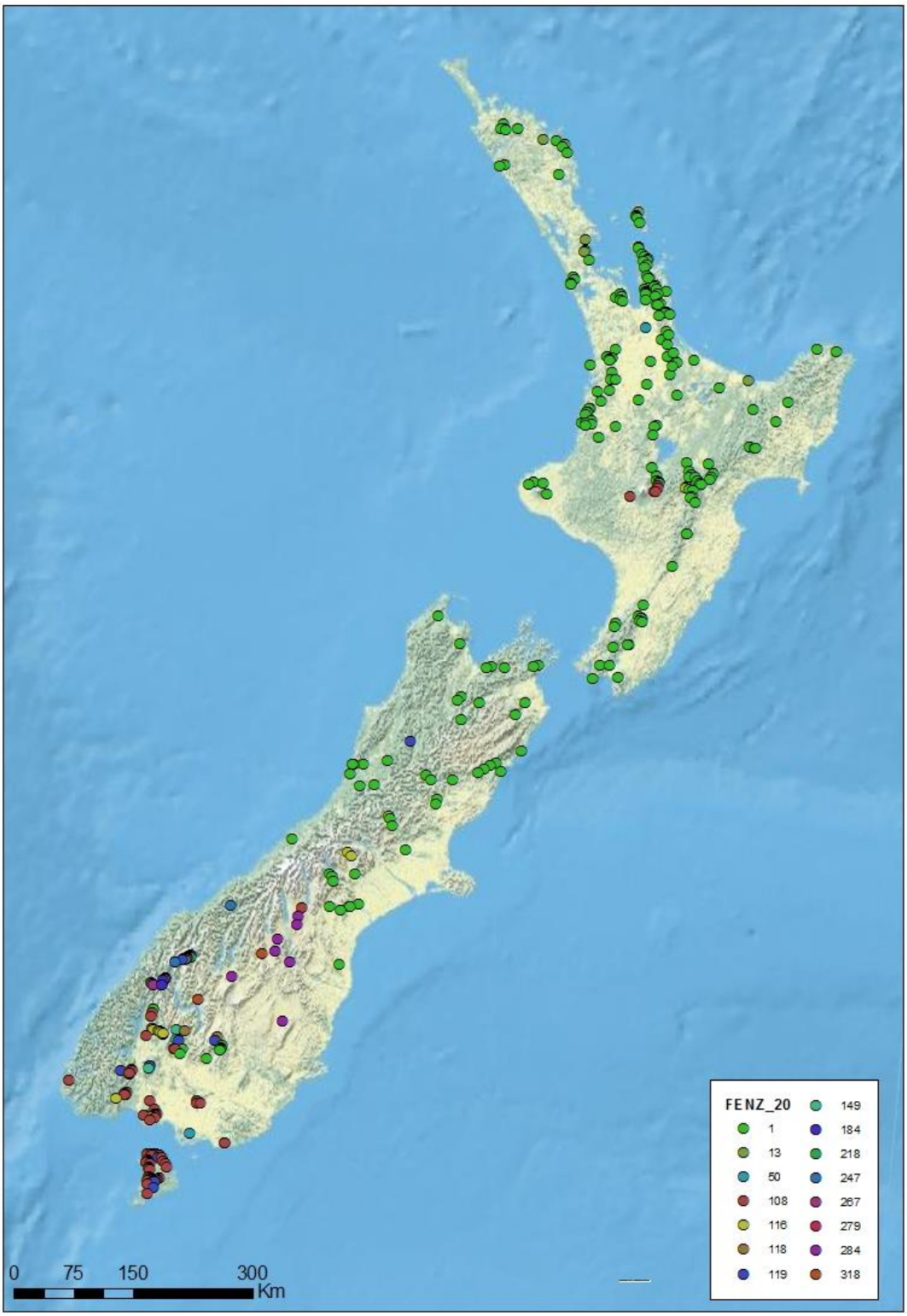


Figure 1: The location of the 359 reference sites used in this study labelled with FENZ level 1 (20-end groups) classification.

I used ANOSIM to test the strength of different biological classifications, with clustering options between 10 and 40-end groups, to inform the classification used in model development. This range of end groups was based on those used in similar models elsewhere (e.g. Wright, 2000; Mazor et al., 2016). Whilst the selection of the number of end groups remains subjective (Mazor et al, 2016), the approach taken here was to identify a clustering option with a high classification strength, while ensuring each end group contained at least five reference sites consistent with the recommendation by Wright et al. (1993). End groups with smaller than five sites should be avoided as they can lead to poor representation and modelling errors (Simpson & Norris, 2000).

Second, the selected biological classification was used to construct a 500-tree random-forest model (Cutler et al. 2007) using the randomForest package in R (Liaw & Wiener 2002) to predict group membership for sites. Environmental predictor variables were assembled from FENZ and River Environment Classification ((REC) Snelder et al, 2004) databases (Table 1). Predictor variables were only considered for inclusion in the model if they were considered to have a causal relationship with the distribution of macroinvertebrates and were not likely to be substantially affected by human activity.

The relative importance of each predictor variable was assessed using two tests following Breiman (2003). First, changes in group prediction error are assessed when each variable is randomly permuted while all others remain unchanged (Mean Decrease Accuracy). Randomly permuting an important variable will have a greater effect on model accuracy. Second, in random forests the gini criterion is a measure of the accuracy of each node and is used to build the decision tree. The sum of decreases in the gini criterion for a given variable is summed and normalised to provide the decrease in the gini criterion for that variable in the classification (Mean Decrease Gini; greater decrease = greater importance of predictor variable).

Third, the probability of each taxa occurring at a site was determined based on the probability of group membership for the test site and the frequency of occurrence of each taxon within each group. The probability of occurrence for a taxon at a test site was calculated as the weighted average of the frequencies of occurrence of a taxon across groups in which frequencies are weighted by the probabilities of group membership.

Three metrics were calculated from the site-specific predictions of taxon occurrence (number of taxa, MCI and EPT richness). The number of taxa expected at a site was determined as the sum of all probabilities for that site. The expected EPT richness was determined in the same way, but only using the probabilities for EPT taxa. The expected MCI score for each site was calculated from the probability of each taxon's occurrence at a site using the method described by Clarke et al (1996).

2.2.2 Environmental model

The Freshwater Ecosystems of New Zealand (FENZ (Leathwick et al. 2010; 2011)) was used as the environmental classification on which to base a predictive model. The FENZ classification is based on a range of catchment-scale and segment-scale descriptors of the stream environment (e.g. geology, climate) and was biologically-optimised using macroinvertebrate and fish community data to ensure river types had biological relevance.

The FENZ Level 1 class (20 group) and Level 2 class (100 group) were used as environmental classification options and the classification strength of these two options was assessed using ANOSIM to allow comparison with the biological classification.

The expected macroinvertebrate metrics (e.g. number of taxa, EPT richness and MCI) for a test site were determined as the mean of the observed metric values for all reference sites in the FENZ group to which the test site is assigned.

2.2.3 Null model

In the null model the predicted metrics for sites are the average observed value for all reference sites (van Sickle et al. 2005). The null model makes no attempt to explain the variation in taxa occurrences among the reference sites and as a result provides an upper limit for standard deviation (SD) (O/E), a limit that would be achieved if a predictive model failed to account for any of the variation in macroinvertebrate communities.

Table 1: Environmental variables used as predictors in a random forest model (including data source).

Variable	Description
Elevation	Metres above sea level
SEGLOWFLOW	Mean annual 7-day low flow (m3/sec), derived from hydrological models, provided by Jochen Schmidt, NIWA (see http://wrenz.niwa.co.nz/webmodel/ for details).
SEGJANAIRT	Summer (January) air temperature (degrees C)—used in the absence of robust estimates of water temperature (FENZ)
SEGMINTNOR	Average minimum daily air temperature (degrees C) normalised with respect to SegJanAirT—negative values indicate strongly seasonal climates and positive values indicate weakly seasonal climates (FENZ)
SEGSLOPE	Segment slope (degrees), derived from GIS calculation using length and difference between upstream and downstream elevation for each segment (FENZ)
SEGSLOPESQ	Square-root transformed segment slope (slope +1) (FENZ)
DSDIST2COA	Distance to coast (km), from mid-point of each river segment (FENZ)
USDAYSRAIN	Days/year with rainfall greater than 25 mm in the upstream catchment (FENZ)
USAVGTNORM	Average air temperature (degrees C) in the upstream catchment, normalised with respect to SegJanAirT, with negative values indicating colder (higher elevation) headwaters than average, given the segment temperature, and positive values indicating warmer temperatures (FENZ)
USAVGSLOPE	Average slope in the upstream catchment (degrees), describes catchment-driven modification of flow variability (FENZ)
USHARDNESS	Average hardness (induration) of surface rocks using values derived from the underlying LENZ layers (FENZ)
SEGFLOW	Mean annual flow (m3/sec), derived from hydrological models (FENZ)
ORDER_	Strahler stream order (REC)
CATCHAREA	Catchment area upstream of each location (REC)
Easting	
Northing	

2.3 Model assessment

The approach described by van Sickle et al. (2005) was used to investigate the predictive performance of models based on the differing classification approaches.

In a useful model, the observed (O) fauna for reference sites should closely resemble the expected (E) fauna based on the model's predictions. Therefore, the O/E ratio for a metric should vary around unity (e.g. O/E = 1) and the standard deviation (SD) of the O/E ratio for a metric should be low (e.g. SD (O/E) < 0.2 (Davy-Bowker et al., 2006)) in a well-performing model. In this study, the assessment of predictive accuracy was based on the O/E values of three metrics; number of taxa, EPT richness and MCI.

The SD (O/E) of each of the environmental and biological classification approaches were compared directly to assess the accuracy of the faunal predictions. Additionally, the SD (O/E) of each model was compared with that of a null model, where for the environmental and biological models we calculated the percent reduction in SD (O/E) compared with the null model. The percent reduction provides a standardised measure of performance that can be used to compare the performance of models developed in different locations.

As with other studies, this analysis was carried out on the training dataset (i.e. those data used to develop the model). Initially this carried out using an 'internal' test of the model, whereby the ability of the models' to predict the data on which they were developed is assessed. This was done for all of the reference sites used in model development, and repeated excluding FENZ groups that were represented by less than five sites.

In addition, for those sites where repeat samples were available, this analysis was also carried out for samples collected in earlier years. For these sites, the samples collected in the three years prior to the sample used in model development were selected to test the performance of the models. This analysis is a more severe test of the models' performance as it does not use invertebrate data that was used in the model development.

3. Results

3.1 Dataset

The site selection process resulted in a dataset of 359 reference sites from public sources, with those samples containing records of 131 taxa at the taxonomic resolution described in the methods. Most taxa were infrequent in the dataset (74 taxa occurred at frequency of occurrence <10%) and only four taxa were present in the dataset at greater than 80% frequency of occurrence; Chironomidae (96.8%); *Deleatidium* (93.9%); Elmidae (83.8%) and *Hydrobiosis* (81.6%) (see Appendix 1 for full details).

The distribution of the 359 reference sites across FENZ classes was used to assess the representativeness of the new public dataset. The 359 reference sites contained representatives of 15 of the 20-group FENZ classification (Table 2). Whilst 15 groups were represented in the reference dataset, 81% of the reference stream length (288 sites) fell into two FENZ classes (Groups C and G). In addition, most of the sites in the North Island were in a single FENZ group (Groups C – see Figure 1).

Four of the groups with no reference sites were small, collectively representing less than 1% of the river network captured by FENZ (Groups F, T, M and R). However, Group J represents 4.2% (16,990 km) of the river network by length and was not represented in the reference site dataset. Furthermore,

Group A is a large group representing 21.5% (or 86,314 km) of the FENZ network that was under-represented in the references site dataset (7 reference sites, or 1.2% of the reference stream length).

A similar analysis using the FENZ 100 group classification showed that the reference site dataset contained representatives of 48 of the Level 2 groups (Appendix 2). This analysis provides greater granularity about under and over-represented river types in the reference site dataset and may be used to inform future invertebrate sample collections. For example, the under-represented Level 1 Group A is comprised of six Level 2 Groups, of which three groups are not represented in the reference site dataset. In addition, the greater granularity in the Level 2 classification revealed representativeness issues that were not apparent at the Level 1 classification. For example, Group H (7.2% of the river network) appears well represented in the reference dataset (4.9%) based on the Level 1 classification, however four of the six Level 2 groups that comprise Group H are not represented in the reference site dataset.

Table 2: The distribution of all streams in the FENZ classification and the reference site dataset by Level 1 FENZ group.

FENZ classification				Reference site dataset		
Group code	Group number	Length in group (km)	% of river network	Number of sites	Length of sites (km)	% of reference dataset
A	13	86314	21.5	7	5.2	1.2
B	50	1458	0.4	2	2.6	0.6
C	1	176549	43.9	214	253.3	58.0
D	284	18592	4.6	7	9.6	2.2
E	318	1124	0.3	2	1.9	0.4
F	67	1924	0.5	0	0.0	0.0
G	108	41692	10.4	74	100.3	23.0
H	116	28884	7.2	14	21.4	4.9
I	247	1600	0.4	2	1.0	0.2
J	130	16990	4.2	0	0.0	0.0
K	279	329	0.1	1	0.4	0.1
L	267	666	0.2	5	5.2	1.2
M	382	120	0.0	0	0.0	0.0
N	119	14126	3.5	9	9.1	2.1
O	184	4265	1.1	6	5.0	1.1
P	149	3372	0.8	7	10.6	2.4
Q	118	1410	0.4	4	6.2	1.4
R	374	29	0.0	0	0.0	0.0
S	218	1695	0.4	5	4.8	1.1
T	276	832	0.2	0	0.0	0.0

3.2 Biological classification

For the purposes of the biological model, the classification strength of 31 clustering options was tested using ANOSIM. Classification strength increased as the number of end groups increased, indicating greater biotic similarity amongst groups with finer divisions of the clustering dendrogram (Figure 2). The permutation test indicated that the ANOSIM *R* statistic was significantly greater than zero for all clustering options, with the *R* statistic of the random permutations never greater than the observed *R*. However, the relationship between classification strength and number of clustering groups was broadly asymptotic, with classification strength increasing rapidly between five and 22-end groups, with marginal increases in classification strength above 22-end groups.

In addition, when the number of end groups increased above 22, several of the clustering options had end groups with less than five reference sites, therefore the 22-end group clustering option was selected for further model development (indicated by the yellow marker in Figure 2).

The classification strength of the 22-end group clustering option was 0.758 (ANOSIM *R*) and the number of reference sites in each of the end groups ranged from 5 to 37, with a mean group size of 16.3 (Table 3). The environmental characteristics for each end group showed evidence of inter-group variability, indicating the potential to predict group membership based on these parameters (Table 3). The geographic distribution of the 22 groups is shown in Figure 3.

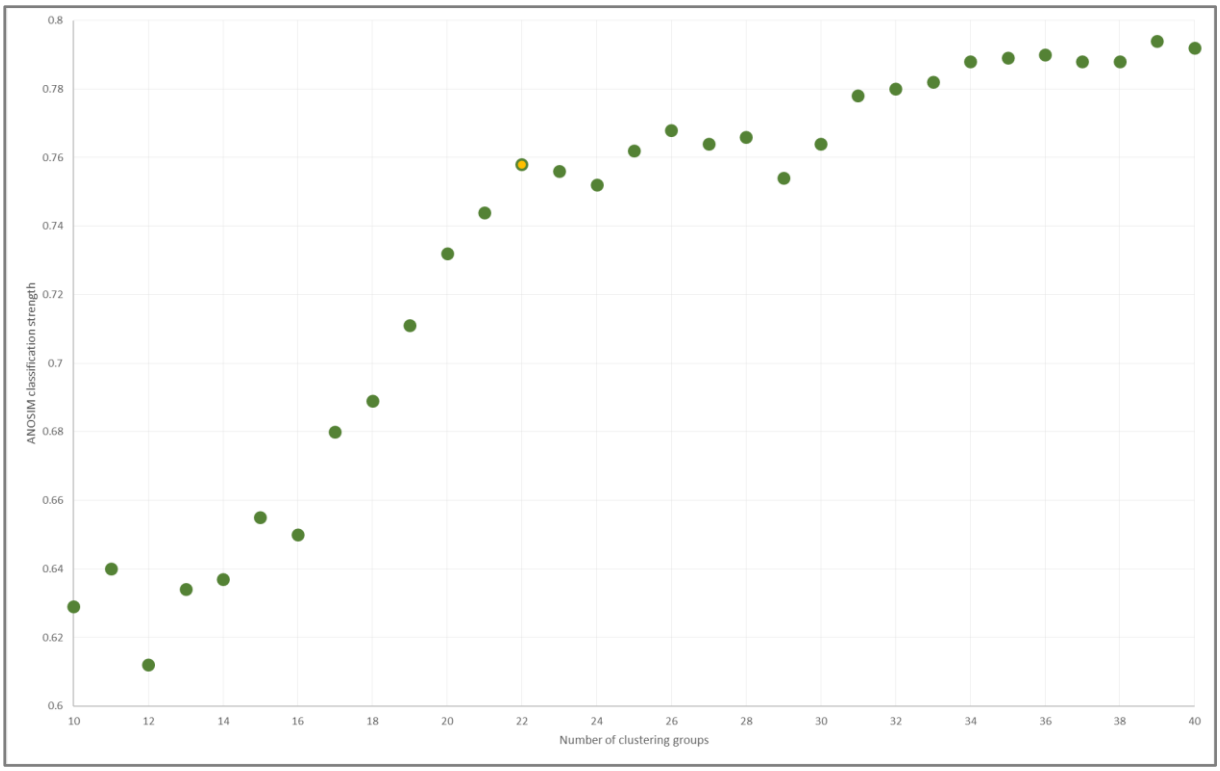


Figure 2: Classification strength (indicated by ANOSIM R) for cluster options with 5 to 40-end groups.

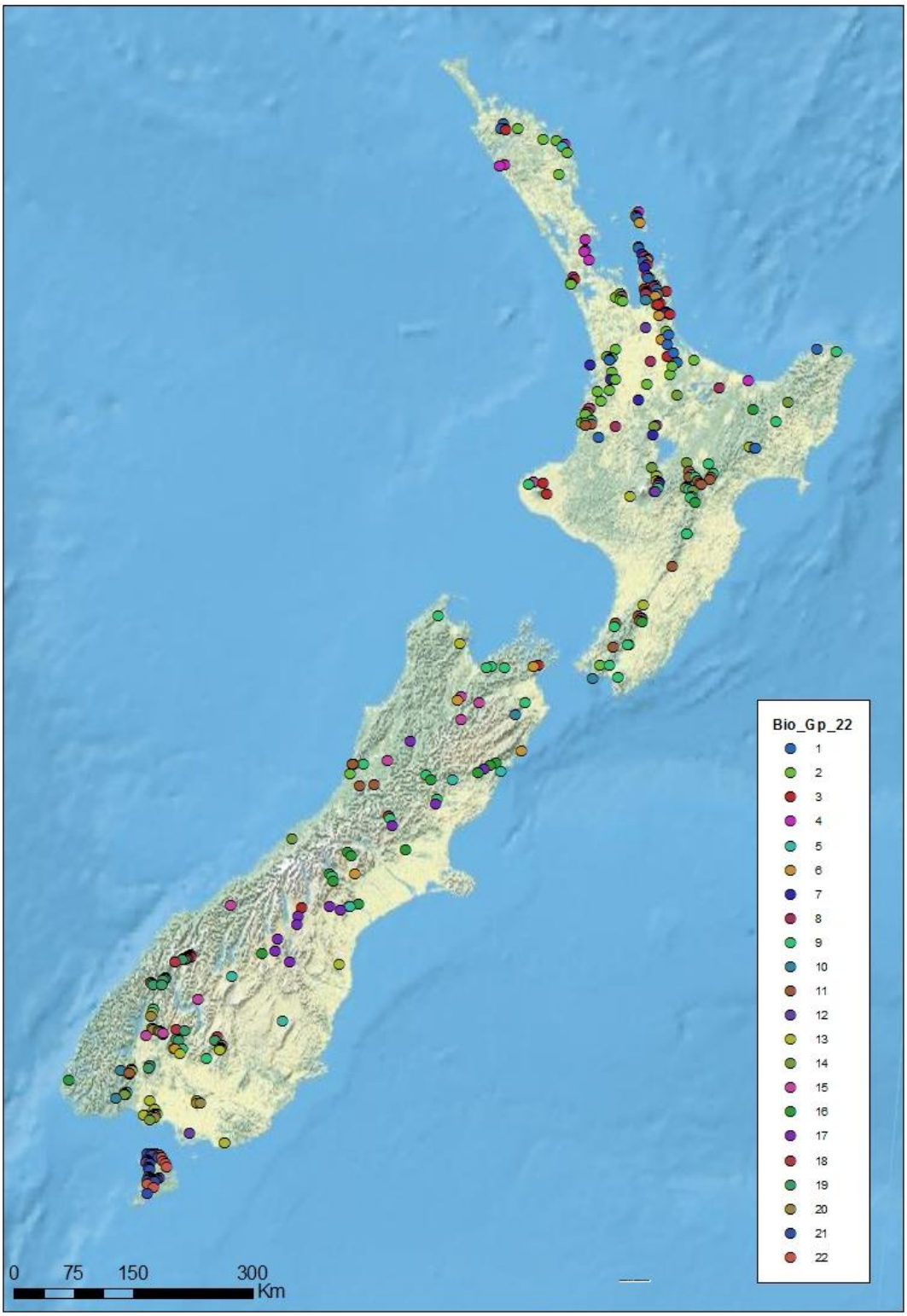


Figure 3: The location of the 359 reference sites used in this study labelled with the 22-end group biological classification.

Table 3: The group size and mean environmental conditions for the 22-end group biological classification. See Table 3 for explanation of environmental variables.

Group number	Number of sites	Elevation	SEG LOW FLOW	SEG JAN AIRT	SEG MINT NOR	SEG SLOPE	SEG SLOPE SQ	DS DIST 2COA	US DAYS RAIN	US AVGT NORM	US AVG SLOPE	US HARD NESS	SEG FLOW	ORDER	CATCH AREA	Easting	Northing
1	17	153	0.09	18.2	1.20	2.4	1.7	17.3	18.0	-0.56	17.2	3.6	0.5	2.8	12828283	2736247	6452299
2	37	175	0.03	17.9	1.17	2.9	1.9	51.3	16.6	-0.10	17.2	3.5	0.2	2.0	4139828	2675377	6393381
3	19	182	0.26	17.6	0.87	1.5	1.5	34.0	23.3	-1.08	18.5	3.5	1.1	3.2	23365250	2667117	6329833
4	11	74	0.01	18.7	1.73	1.6	1.6	4.6	11.8	0.08	12.8	3.4	0.1	1.8	4503250	2669227	6530870
5	11	333	0.59	15.7	-0.58	1.7	1.5	93.2	10.5	-1.74	21.4	3.8	2.2	3.8	67398532	2422179	5838472
6	15	192	0.03	17.3	0.92	2.6	1.8	42.8	14.7	-0.45	19.5	3.8	0.2	2.4	6150221	2598480	6197030
7	11	440	0.04	16.4	0.47	5.5	2.5	142.2	15.1	-0.16	17.9	3.6	0.1	1.9	3679011	2635695	6202217
8	9	141	0.05	18.1	0.73	3.0	1.9	69.8	16.1	-0.31	16.6	3.5	0.2	2.3	5466580	2739157	6402631
9	31	356	1.75	16.0	0.19	1.2	1.5	95.4	20.4	-1.48	23.4	3.5	6.9	3.8	106026956	2591485	6016735
10	11	589	0.93	13.9	0.97	3.9	1.9	166.6	32.1	-1.49	22.7	4.0	3.9	3.5	122119497	2379418	5846334
11	24	333	0.53	15.9	0.71	1.2	1.4	99.8	25.5	-0.99	24.3	3.5	2.5	3.4	31807293	2640985	6092972
12	5	254	0.07	14.1	1.95	1.6	1.5	96.0	12.5	0.95	6.6	3.1	0.2	2.6	7210397	2374448	5748767
13	24	317	0.76	14.8	0.50	1.8	1.6	138.8	13.6	-1.08	16.0	3.5	3.0	3.5	64408443	2364179	5751741
14	11	572	0.04	14.9	0.55	5.3	2.3	165.1	18.5	0.57	13.6	2.8	0.2	1.9	3807088	2626807	6070142
15	11	384	10.90	15.4	-0.13	1.1	1.4	130.2	26.8	-2.34	24.8	3.6	35.6	4.7	430637717	2415463	5887610
16	15	443	6.11	15.2	-0.59	2.4	1.7	87.6	18.0	-2.25	25.1	3.6	19.6	4.0	376904384	2463262	5833674
17	11	515	0.43	15.1	-2.11	1.5	1.5	109.1	8.2	-1.36	18.2	3.5	2.1	3.9	141495755	2360957	5725170
18	11	1259	0.79	10.1	1.24	14.2	3.5	279.7	55.2	-0.53	27.8	4.2	3.1	2.1	25507445	2151746	5601131
19	20	1106	0.11	10.4	1.33	12.4	3.4	183.4	35.7	0.62	29.7	4.1	0.3	1.7	3869577	2133612	5550857
20	20	368	0.01	13.5	0.72	3.9	2.1	109.8	6.9	0.45	13.8	3.4	0.1	2.3	4573665	2129964	5485708
21	19	167	0.04	12.3	3.28	5.4	2.4	8.3	10.2	1.10	15.2	4.5	0.3	1.6	7968363	2112226	5359916
22	16	160	0.02	12.3	3.30	6.4	2.6	4.6	9.0	0.83	16.6	4.5	0.1	1.8	3788494	2119181	5360650

3.3 Predicting taxa occurrence with a biological model

A random forest model was developed to predict group membership using the 16 environmental predictors selected as described in the methods (Table 1). The relative importance of each predictor was assessed by two indicators of predictor importance that are calculated in different ways. However, both indicators rank the same four most important variables (Figure 4). Furthermore, despite a substantially different reference site dataset, the relative importance of environmental predictors in the random forest model developed here were similar to those for the proof-of-concept model. For example, the four most important environmental predictors were identical (northing, easting, SEGJANAIRT and SEGMINTNOR).

The error rate for group prediction was 57.4% for the model developed here. This compares poorly with the proof-of-concept model error rate of 48% (Clapcott et al., 2017). Whilst it is not critical for model performance that a site is predicted to the correct group, it is important to minimise this error rate. This is because the model produces a probability of each site belonging to all groups and the invertebrate community predictions are based on all of these probabilities.

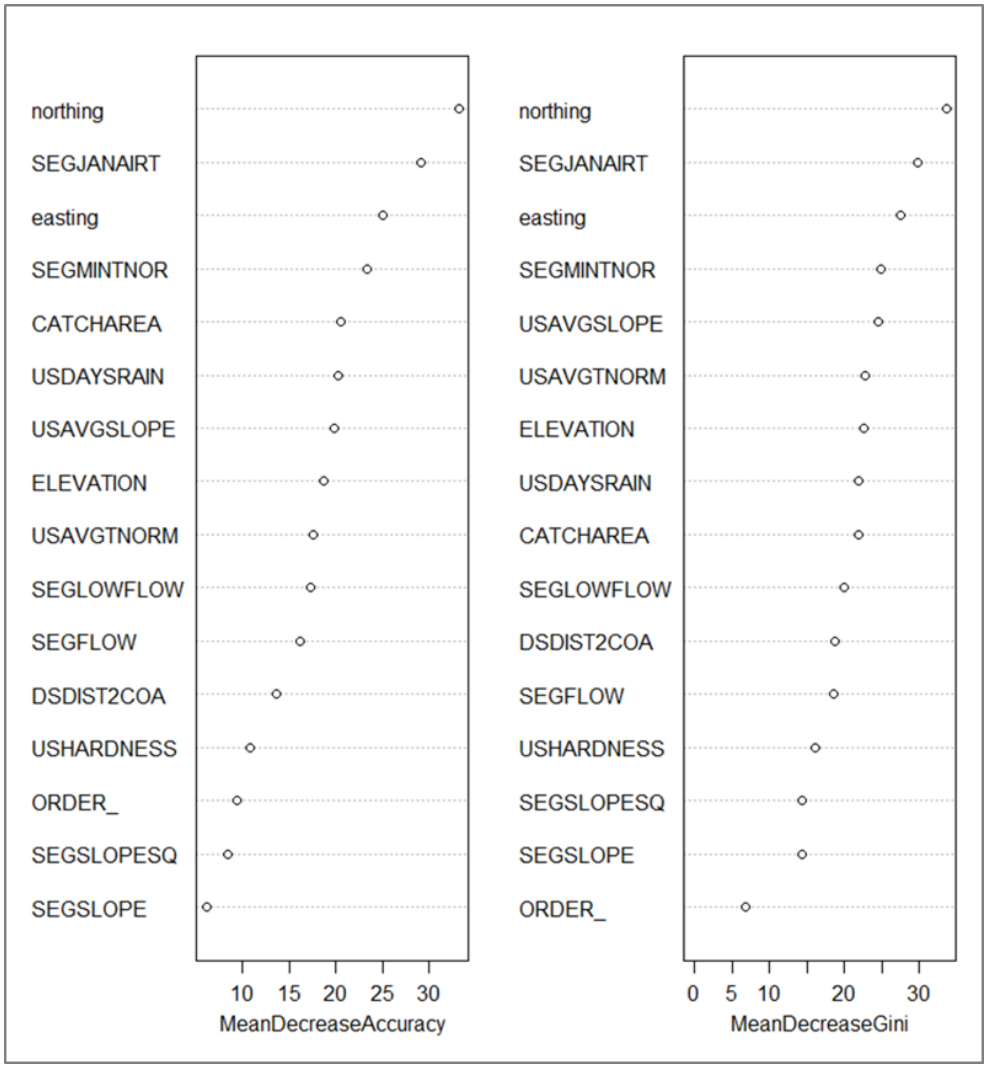


Figure 4: Relative importance of environmental predictors for the random forest model based on the 22-group biological classification. Higher Gini index = higher importance for that predictor and a greater contribution to model accuracy.

3.4 Environmental model

The two FENZ classification levels used in the representativeness assessment were used to form the basis of an environmental model.

The FENZ Level 1 classification had 15 of the 20 groups represented in the reference site data set and the number of reference sites per group ranged from 1 to 214. The mean group size was 23.9, but note that two groups were very large compared with the others (Table 2; Figure 1).

The FENZ Level 2 classification had 48 of the 100 groups represented in the reference site data set and the number of reference sites per group ranged from 1 to 49 (Appendix 2; Figure 5). The mean group size was 7.5. For comparison, the level 2 classification was used in the proof-of-concept study, where 50 groups were represented, and group membership ranged from 1 to 84, with a mean of 11 reference sites per group (Clapcott et al., 2017).

The classification strength of the two FENZ classifications assessed using ANOSIM were 0.414 (FENZ 100 group) and 0.367 (FENZ 20 group), which indicates that the FENZ classifications are lower in strength than the selected biological classification (0.758). The higher the classification strength, the greater the within group biological similarity compared with between group biological similarity, which indicates a better classification.

3.5 Predictive performance

3.5.1 Internal test

The use of an environmental or biological classification resulted in a reduction of the SD (O/E) for all three test metrics, and hence an increase in predictive performance when compared with the null model (Table 4). In all of the models the SD (O/E) was lower for MCI (or similar indices) than for metrics that use number of taxa (Number of taxa and EPT richness), as has been found in similar evaluative studies elsewhere (Moss et al. 1999; van Sickle et al. 2005; Davy-Bowker et al. 2006; Neale & Rippey 2008; Clapcott et al. 2017).

For the complete dataset, the environmental model based on the FENZ 100 group classification produced the lowest SD (O/E) for all three test metrics, resulting in a reduction of 18% for Number of taxa, 23% for EPT richness and 26% for MCI (Table 4).

The environmental model developed here based on FENZ 100 group classification provided greater predictive performance than the equivalent proof-of-concept model (12% reduction in Number of taxa; 22% reduction in MCI) (Clapcott et al., 2017). In contrast, the biological model developed here had lower predictive performance than the equivalent proof-of-concept model (14% reduction in Number of taxa; 19% reduction in MCI).

When the predictive performance analysis was undertaken after excluding FENZ groups with less than five references sites, the differential in performance between the environmental and biological model was negligible for FENZ 100 group classification (Table 5) and the biological model performed better compared with the FENZ 20 group classification (Table 6).

The exclusion of the groups with less than five reference sites reduced the performance of the environmental models, but also improved the performance of the biological model. Indeed, the exclusion of these sites resulted in the biological model's predictive performance moving closer to that of the proof-of- concept biological model.

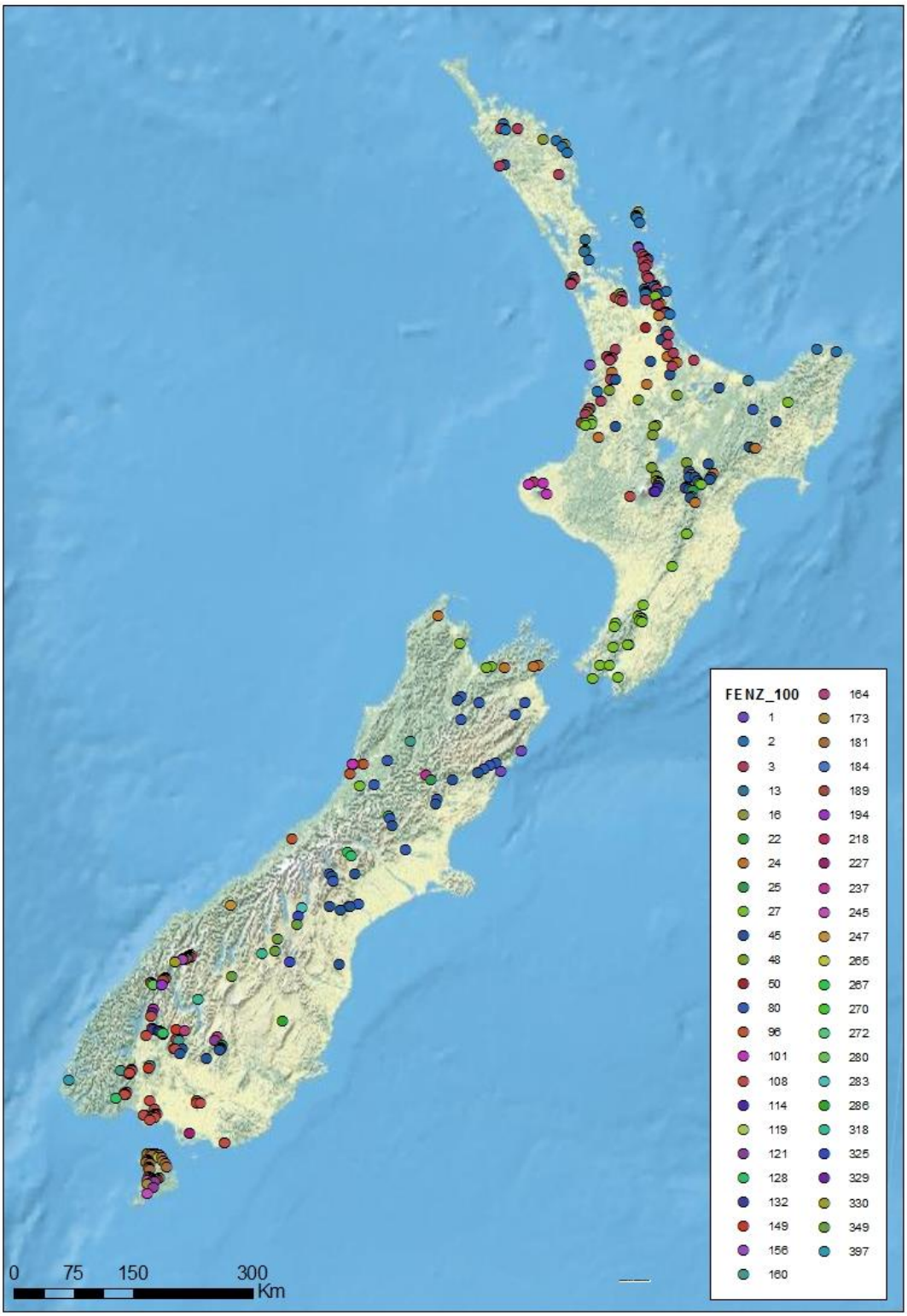


Figure 5: The location of the 359 reference sites used in this study labelled with FENZ level 2 (100 group) classification.

Table 4: Predictive performance of the models indicated by standard deviation (SD) of O/E for the three test metrics. Percent reduction compared with the null model is shown in parentheses. A lower SD (O/E) and higher % reduction indicates improving model performance.

Model	SD (O/E) Number of taxa	SD (O/E) EPT richness	SD (O/E) MCI
Null	0.287	0.368	0.116
Biological	0.271 (6%)	0.330 (10%)	0.098 (16%)
Environmental (FENZ 20)	0.256 (11%)	0.313 (15%)	0.094 (19%)
Environmental (FENZ 100)	0.236 (18%)	0.285 (23%)	0.086 (26%)

Table 5: Predictive performance of the environmental model based on the FENZ 100 group classification and biological models indicated by standard deviation (SD) of O/E for the three test metrics excluding FENZ groups with less than five reference sites (293 sites in 16 groups represented). Percent reduction compared with the null model is shown in parentheses. A lower SD (O/E) and higher % reduction indicates improving model performance.

Model	SD (O/E) Number of taxa	SD (O/E) EPT richness	SD (O/E) MCI
Null	0.281	0.360	0.109
Biological	0.250 (11%)	0.302 (16%)	0.088 (19%)
Environmental (FENZ 100)	0.248 (12%)	0.298 (17%)	0.091 (17%)

Table 6: Predictive performance of the environmental model based on the FENZ 20 group classification and biological models indicated by standard deviation (SD) of O/E for the three test metrics excluding FENZ groups with less than five reference sites (348 sites in 10 groups represented). Percent reduction compared with the null model is shown in parentheses. A lower SD (O/E) and higher % reduction indicates improving model performance.

Model	SD (O/E) Number of taxa	SD (O/E) EPT richness	SD (O/E) MCI
Null	0.284	0.362	0.112
Biological	0.254 (11%)	0.308 (15%)	0.090 (20%)
Environmental (FENZ 20)	0.274 (4%)	0.334 (8%)	0.099 (12%)

3.5.2 Independent test

For the independent test, there were 144 sites that had three-years of samples collected prior to the data used in model development (432 samples in total). For this assessment of model performance, the biological model resulted in the lowest SD (O/E) for all three metrics (Table 7).

Table 7: Predictive performance of the models indicated by standard deviation (SD) of O/E for the three test metrics on an independent set of invertebrate data (144 sites). Percent reduction compared with the null model is shown in parentheses. A lower SD (O/E) and higher % reduction indicates improving model performance.

Model	SD (O/E) Number of taxa	SD (O/E) EPT richness	SD (O/E) MCI
Null	0.283	0.372	0.120
Biological	0.252 (11%)	0.311 (16%)	0.097 (19%)
Environmental (FENZ 20)	0.273 (4%)	0.341 (8%)	0.110 (8%)
Environmental (FENZ 100)	0.257 (9%)	0.328 (12%)	0.100 (17%)

4. Discussion

The development of multivariate model for the prediction of stream invertebrate communities was hampered by the availability of reference site data. The proof-of-concept study utilised a dataset of 538 reference sites, which unfortunately were not all available for this work. The new dataset of publicly available data compiled for this study consisted of 359 reference sites, which was also less representative of the global population of river length in New Zealand.

The smaller and less representative dataset is likely to be the primary reason why the predictive performance of the biological model was worse than the proof-of-concept model in a like-for-like comparison. Somewhat counterintuitively, the limitations of the dataset also likely explain the improved performance of the environmental model based on the FENZ Level 2 classification. This classification included 16 groups that were represented by a single site and given the predictions from an environmental model are based on the mean metric scores for the sites in a FENZ group, this results in a perfect match of observed (O) and expected (E) metrics (and hence reduces the SD (O/E).

The differential between the model performance of the biological and environmental classifications disappeared when the groups with less than five sites were excluded from this analysis. Furthermore, when the performance of the models was assessed using invertebrate data not used in model development, the biological model produced the greatest reduction in SD (O/E). Collectively, these analyses indicate that the biological model is performing reasonably well for those river types that are well represented in the dataset. The reduction in SD (O/E) achieved when underrepresented river types were excluded, or the independent test, were at the lower end of the magnitude reported in similar studies of biological models that are used in operational monitoring and reporting programmes (for example, 12% to 34% in European rivers (Davy-Bowker et al. 2006); 14% to 53% in US rivers (van Sickle et al. 2005).

Based on this work, the continued development of a multivariate model for predicting invertebrate reference conditions is contingent on an expanded and more representative reference site dataset. This is likely to require the collection of new stream invertebrate data from under-represented river types and the FENZ analysis presented here could be used to inform that data collection. Both FENZ classification levels indicated river types that were poorly represented in the reference site dataset. This data collection may be done in conjunction with future work on stressor specific metrics recommended by Wagenhoff et al (2018). There are also opportunities to explore the prediction of reference condition for the stressor specific metrics developed by Wagenhoff et al, as 46 of the 49 indicator taxa for stressors are predicted by the biological model developed here. However, the presence-absence nature of the biological model limits the calculation of the two abundance based stressor specific metrics (see Table 13 in Wagenhoff et al, 2018).

In addition, it would obviously be sensible to seek permission to use some of the private data used in the proof-of-concept study that was unavailable for this project. It is critically important to have a large and representative reference site dataset that captures natural variability to create an assessment tool that performs well across heterogeneous environmental settings (Cao et al., 2007; Mazor et al., 2016). Once that reference dataset is available for New Zealand, the model architecture developed in this project can be applied to that dataset with the aim of producing an bioassessment tool that may be used in operational environmental monitoring and assessment programmes.

5. Acknowledgements

I gratefully acknowledge those that have generated the data used in this work, including all the regional councils, Richard Storey, Dave West and Lyndsay Chadderton, and those that compiled and groomed the dataset (Martin Unwin and Annika Wagenhoff).

Thanks also to Russell Death, Evan Harrison and Joanne Clapcott for technical discussions and review comments on draft versions of this report. Thanks to the Ministry for the Environment for funding.

6. References

Breiman L. 2003. Manual on setting up, using, and understanding random forests v4.0. Available at https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf

Cao Y, Hawkins CP, Olson J, Kosterman MA. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. Journal of the North American Benthological Society 26:566–585.

Chadderton WL. 1988 Faunal and chemical characteristics of some Stewart Island streams. New Zealand Natural Sciences, 15, 43-50.

Clapcott J, Wagenhoff A, Neale M, Death R, Storey R, Smith B, Harding J, Matthaei C, Collier K, Quinn J, Young R. 2017. Exploring macroinvertebrate metrics for the National Policy Statement for Freshwater Management. Prepared for Ministry for the Environment. Cawthron Report No. 3073.

Clarke KR. 1993. Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology 18: 117–143.

Clarke KR, Warwick RM. 1994. Change in marine communities: an approach to statistical analysis and interpretation. Plymouth Marine Laboratory: Plymouth.

Clarke RT, Furse MT, Wright JF, Moss D. 1996. Derivation of a biological quality index for river sites: Comparison of the observed with the expected fauna. Journal of Applied Statistics 23: 311-332.

Clarke RT, Wright JF, Furse MT. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. Ecological Modelling 160: 219-233.

Coysh J, Norris R. 1999. Waikato Region AUSRIVAS / RIVPACS trial. Report prepared for the Ministry for the Environment. SMF Contract No. 5091.

Cutler DR, Edwards TC Jr., Beard KH, Cutler A, Hess KT. 2007. Random forests for classification in ecology. Ecology 88: 2783-2792.

Davy-Bowker J, Clarke RT, Johnson RK, Kokeš J, Murphy JF, Zahrádková S. 2006. A comparison of the European Water Framework Directive physical typology and RIVPACS-type models as alternative methods of establishing reference conditions for benthic macroinvertebrates. Hydrobiologia 566: 91–105.

Death RG, Collier KJ. 2010. Measuring stream macroinvertebrate responses to gradients of vegetation cover: when is enough enough? Freshwater Biology 55: 1447-1464.

Depree C, Clapcott J, Booker D, Franklin P, Hickey C, Matheson F, Shelley J, Unwin M, Goodwin E, Mackman J, Rabel H, Wagenhoff A. 2017. Development of ecosystem health thresholds for suspended

and deposited sediment in New Zealand rivers and streams. Prepared for Ministry for the Environment. 163 p.

Hawkins CP, Norris RH. 2000. Performance of different landscape classifications for aquatic bioassessments: introduction to the series. Journal of the North American Benthological Society 19: 367–369.

Hawkins CP, Norris RH, Hogue JN, Feminella JW. 2000a. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10: 1456–1477.

Hawkins CP, Norris RH, Gerritsen J, Hughes RM, Jackson SK, Johnson RK, Stevenson RJ. 2000b. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society 19: 541–556.

Joy MK, Death RG. 2003. Biological assessment of rivers in the Manawatu-Wanganui region of New Zealand using a predictive macroinvertebrate model. New Zealand Journal of Marine and Freshwater Research 37: 367-379.

Kaufman L, Rousseeuw P. 1990. Finding groups in data: An introduction to cluster analysis. Wiley, New York.

Kokeš J, Zahrádková S, Němejcová D, Hodovský J, Jarkovský J, Soldán T. 2006. The PERLA system in the Czech Republic: a multivariate approach to assess ecological status of running waters. Hydrobiologia 566: 343–354.

Johnson RK. 2003. Development of a prediction system for lake stony-bottom littoral macroinvertebrate communities. Archiv für Hydrobiologie 158: 517–540.

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

Liaw A, Wiener M. 2002. Classification and regression by randomForest. R News 2(3): 18-22.

Mazor RD, Reynoldson TB, Rosenberg DM, Resh VH. 2006. Effects of biotic assemblage, classification, and assessment method on bioassessment performance. Canadian Journal of Fisheries and Aquatic Sciecnes 63: 394-411.

Mazor RD, Rehn AC, Ode PR, Engeln M, Schiff KC, Stein ED, Gillett DJ, Herbst DB, Hawkins CP. 2016. Bioassessment in complex environments: designing an index for consistent meaning in different settings. Freshwater Science 35: 249-271.

Moss D, Furse MT, Wright JF, Armitage PD. 1987. The prediction of the macroinvertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17: 41–52.

Neale MW, Rippey B. 2008. A comparison of environmental and biological site classifications for the prediction of macroinvertebrate communities of lakes in Northern Ireland. Aquatic Conservation: Marine and Freshwater Ecosystems 18: 729-741.

Parsons M & Norris RH. 1996. The effect of habitat-specific sampling on biological assessment of water quality using a predictive model. Freshwater Biology 36: 419–434.

Poquet JM, Alba-Tercedor J, Puntí T, Sánchez-Montoya MM, Robles S, Álvarez M, Zamora-Muñoz C, Sáinz-Cantero CE, Vidal-Abarca MR, Suárez ML, Toro M, Pujante AM, Rieradevall M, Prat N. 2009. The

MEDiterranean Prediction And Classification System (MEDPACS): an implementation of the RIVPACS/AUSRIVAS predictive approach for assessing Mediterranean aquatic macroinvertebrate communities. Hydrobiologia 623: 153–171.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reynoldson TB, Bailey RC, Day KE, Norris RH. 1995. Biological guidelines for freshwater sediment based on Benthic Assessment of SedimenT (the BEAST) using a multivariate approach for predicting biological state. Australian Journal of Ecology 20: 198–219.

Reynoldson TB, Norris RH, Resh VH, Day KE, Rosenberg DM. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society 16: 833–852.

Simpson JC, Norris RH. 2000. Biological assessment or river quality: development of AUSRVAS models and outputs. In Assessing the Biological Quality of Freshwaters. RIVPACS and Other Techniques, Wright JF, Sutcliffe DW, Furse MT (eds). Freshwater Biological Association: Windermere; 1–23.

Snelder T, Biggs B, Weatherhead M. 2004. New Zealand River Environment Classification User Guide. Ministry for the Environment by the National Institute of Water and Atmospheric Research. 145 p.

Van Sickle J, Hughes RM. 2000. Classification strengths of ecoregions, catchments and geographic clusters for aquatic vertebrates in Oregon. Journal of the North American Benthological Society 19: 370–384.

Van Sickle J, Hawkins CP, Larsen DP, Herlihy AH. 2005. A null model for the expected macroinvertebrate assemblage in streams. Journal of the North American Benthological Society 24: 178–191.

Wagenhoff A, Goodwin E, Atalah J, Smith B, Harding J. 2018. Development and validation of stressor-specific macroinvertebrate metrics. Prepared for Ministry for the Environment. Cawthron Report No. 3197. 83 p. plus appendices.

Wright JF, Furse MT, Armitage PD. 1993. RIVPACS - a technique for evaluating the biological quality of rivers in the U.K. European Water Pollution Control 3: 15–25.

Wright JF. 2000. An introduction to RIVPACS. In Assessing the Biological Quality of Freshwaters. RIVPACS and Other Techniques, Wright JF, Sutcliffe DW, Furse MT (eds). Freshwater Biological Association: Windermere; 1–23.

7. Appendices

7.1 Appendix 1

Frequency of occurrence of the 131 taxa at the 359 reference sites used in this study.

Taxon	Number of records	Frequency (% of sites)	Taxon	Number of records	Frequency (% of sites)
Chironomidae	348	96.9	Paracalliope	22	6.1
Deleatidium	337	93.9	Collembola	21	5.8
Elmidae	301	83.8	Rallidens	21	5.8
Hydrobiosis	293	81.6	Taraperla	21	5.8
Psilochorema	240	66.9	Cristaperla	19	5.3
Zelandoperla	238	66.3	Microvelia	19	5.3
Oligochaeta	232	64.6	Rakiura	18	5.0
Austrosimulium	229	63.8	Psychodidae	16	4.5
Potamopyrgus	228	63.5	Tabanidae	16	4.5
Coloburiscus	224	62.4	Atalophlebioides	15	4.2
Hydraenidae	216	60.2	Isopoda	15	4.2
Aoteapsyche	207	57.7	Nemertea	15	4.2
Archichauliodes	201	56.0	Paratya	15	4.2
Austroperla	198	55.2	Staphylinidae	15	4.2
Olinga	194	54.0	Paralimnophila	14	3.9
Nesameletus	192	53.5	Sphaeriidae	13	3.6
Zelandobius	178	49.6	Ferrissia	12	3.3
Aphrophila	173	48.2	Molophilus	12	3.3
Eriopterini	171	47.6	Nematomorpha	12	3.3
Stenoperla	169	47.1	Oniscigaster	11	3.1
Austroclima	167	46.5	Copepoda	10	2.8
Acarina	162	45.1	Nothodixa	9	2.5
Costachorema	156	43.5	Xanthocnemis	8	2.2
Empididae	154	42.9	Alloecentrella	7	1.9
Pycnocentrodes	149	41.5	Ephydridae	7	1.9
Helicopsyche	127	35.4	Mauiulus	7	1.9
Oxyethira	126	35.1	Paroxyethira	7	1.9
Platyhelminthes	112	31.2	Physa	7	1.9
Zephlebia	109	30.4	Antipodochlora	6	1.7
Megaleptoperla	108	30.1	Berosus	6	1.7
Pycnocentria	108	30.1	Cladocera	6	1.7
Ceratopogonidae	103	28.7	Stratiomyidae	5	1.4
Hydrobiosella	98	27.3	Anisoptera	4	1.1

Taxon	Number of records	Frequency (% of sites)	Taxon	Number of records	Frequency (% of sites)
Beraeoptera	96	26.7	Lymnaeidae	4	1.1
Neozephlebia	84	23.4	Tepakia	4	1.1
Hudsonema	75	20.9	Zelandotipula	4	1.1
Orthopsyche	73	20.3	Gyraulus	3	0.8
Muscidae	73	20.3	Liodessus	3	0.8
Nematoda	69	19.2	Zelandoptila	3	0.8
Amphipoda	67	18.7	Zemelanopsis	3	0.8
Scirtidae	65	18.1	Antiporus	2	0.6
Zelolessica	65	18.1	Dolichopodidae	2	0.6
Neurochorema	58	16.2	Echyridella	2	0.6
Acroperla	56	15.6	Ecnomidae	2	0.6
Ptilodactylidae	56	15.6	Kempynus	2	0.6
Ostracoda	54	15.0	Oecetis	2	0.6
Triplectides	54	15.0	Peritheates	2	0.6
Ameletopsis	50	13.9	Pycnocentrella	2	0.6
Plectrocnemia	50	13.9	Sigara	2	0.6
Hydrochorema	46	12.8	Siphlaenigma	2	0.6
Acanthophlebia	43	12.0	Spaniocercoides	2	0.6
Hydrophilidae	43	12.0	Anisops	1	0.3
Ichthybotus	42	11.7	Arachnocolus	1	0.3
Hexatomini	41	11.4	Culicidae	1	0.3
Paraleptamphopus	41	11.4	Enochrus	1	0.3
Neocurupira	40	11.1	Hirudinea	1	0.3
Paradixa	40	11.1	Hydra	1	0.3
Latia	39	10.9	Hygraula	1	0.3
Limonia	35	9.7	Isothraulus	1	0.3
Mischoderus	34	9.5	Nesoperla	1	0.3
Spaniocerca	34	9.5	Procordulia	1	0.3
Oeconesidae	32	8.9	Saldidae	1	0.3
Philorheithrus	28	7.8	Scatella	1	0.3
Paranephrops	26	7.2	Sciomyzidae	1	0.3
Polyplectropus	23	6.4	Triplectidina	1	0.3
Confluens	22	6.1			

7.2 Appendix 2

The distribution of all streams in the FENZ classification and the reference site dataset by Level 2 FENZ group.

FENZ river network				Reference site dataset		
Group code	Group number	Length in group (km)	% of river network	Number of sites	Length of sites	% of reference dataset
A1	13	19890	4.95	3	1.3	0.31
A2	22	6356	1.58	1	1.0	0.23
A3	16	3821	0.95	3	1.2	0.28
A4	32	37453	9.32	0	0.0	0.00
A5	109	16488	4.10	0	0.0	0.00
A6	179	2306	0.57	0	0.0	0.00
B1	50	918	0.23	1	2.3	0.52
B2	227	357	0.09	1	0.4	0.08
B3	256	184	0.05	0	0.0	0.00
C1	1	4140	1.03	4	5.5	1.27
C10	25	11197	2.79	4	3.4	0.79
C11	57	1899	0.47	0	0.0	0.00
C12	80	6564	1.63	17	25.3	5.83
C13	237	1582	0.39	3	3.2	0.74
C2	96	7709	1.92	4	7.6	1.76
C3	101	4412	1.10	5	13.7	3.16
C4	2	7967	1.98	24	30.4	7.02
C5	3	21963	5.46	49	52.9	12.21
C6	24	17254	4.29	17	17.5	4.03
C7	27	16763	4.17	34	34.6	7.99
C8	45	56412	14.03	40	38.9	8.98
C9	48	18688	4.65	13	20.2	4.66
D1	284	2534	0.63	0	0.0	0.00
D2	286	3503	0.87	1	0.7	0.16
D3	325	6953	1.73	2	2.0	0.46
D4	349	5602	1.39	4	4.0	0.92
E1	318	788	0.20	2	2.0	0.46
E2	381	214	0.05	0	0.0	0.00
E3	367	122	0.03	0	0.0	0.00
F1	67	374	0.09	0	0.0	0.00
F2	86	1166	0.29	0	0.0	0.00
F3	319	384	0.10	0	0.0	0.00

FENZ river network				Reference site dataset		
G1	108	24842	6.18	31	47.6	10.98
G10	397	6	0.00	1	0.6	0.13
G2	114	3643	0.91	6	8.7	2.00
G3	283	4550	1.13	1	1.6	0.37
G4	222	41	0.01	0	0.0	0.00
G5	209	931	0.23	0	0.0	0.00
G6	334	1377	0.34	0	0.0	0.00
G7	173	2687	0.67	4	4.9	1.12
G8	181	2101	0.52	21	22.7	5.23
G9	245	1514	0.38	10	14.3	3.30
H1	116	7857	1.95	0	0.0	0.00
H2	324	1522	0.38	0	0.0	0.00
H3	126	4072	1.01	0	0.0	0.00
H4	171	2221	0.55	0	0.0	0.00
H5	128	2752	0.68	5	12.0	2.78
H6	132	10461	2.60	9	9.4	2.17
I1	247	1450	0.36	1	0.7	0.16
I2	395	59	0.01	0	0.0	0.00
I3	330	91	0.02	1	1.0	0.23
J1	130	2431	0.60	0	0.0	0.00
J10	309	941	0.23	0	0.0	0.00
J11	392	327	0.08	0	0.0	0.00
J12	174	1084	0.27	0	0.0	0.00
J13	396	118	0.03	0	0.0	0.00
J2	250	1157	0.29	0	0.0	0.00
J3	306	686	0.17	0	0.0	0.00
J4	308	276	0.07	0	0.0	0.00
J5	387	815	0.20	0	0.0	0.00
J6	170	1893	0.47	0	0.0	0.00
J7	249	1827	0.45	0	0.0	0.00
J8	246	3579	0.89	0	0.0	0.00
J9	235	1855	0.46	0	0.0	0.00
K1	279	70	0.02	0	0.0	0.00
K2	293	110	0.03	0	0.0	0.00
K3	329	60	0.01	1	1.0	0.23
K4	328	88	0.02	0	0.0	0.00
L1	267	215	0.05	1	1.4	0.31
L2	270	94	0.02	1	1.4	0.32
L3	272	66	0.02	1	1.0	0.23

FENZ river network				Reference site dataset		
L4	280	291	0.07	2	1.5	0.35
M1	382	58	0.01	0	0.0	0.00
M2	383	62	0.02	0	0.0	0.00
N1	119	2217	0.55	1	1.5	0.35
N2	121	3592	0.89	3	3.0	0.69
N3	151	2311	0.57	0	0.0	0.00
N4	160	2481	0.62	4	2.7	0.62
N5	156	2819	0.70	1	1.9	0.45
N6	240	706	0.18	0	0.0	0.00
O1	184	813	0.20	2	1.1	0.26
O2	194	954	0.24	1	1.0	0.23
O3	191	403	0.10	0	0.0	0.00
O4	388	886	0.22	0	0.0	0.00
O5	189	843	0.21	1	0.9	0.21
O6	265	203	0.05	2	1.9	0.44
O7	287	15	0.00	0	0.0	0.00
O8	300	104	0.03	0	0.0	0.00
O9	394	45	0.01	0	0.0	0.00
P1	149	1963	0.49	7	10.6	2.45
P2	197	464	0.12	0	0.0	0.00
P3	317	102	0.03	0	0.0	0.00
P4	188	832	0.21	0	0.0	0.00
P5	398	12	0.00	0	0.0	0.00
Q1	118	502	0.12	0	0.0	0.00
Q2	164	908	0.23	4	6.2	1.42
R1	374	9	0.00	0	0.0	0.00
R2	379	20	0.01	0	0.0	0.00
S1	218	1695	0.42	5	4.8	1.10
T1	276	832	0.21	0	0.0	0.00