Earth Sciences New Zealand

Comparison of the eFIBI using 6-replicate eDNA vs n-replicate eDNA samples

Prepared for Dairy NZ

June 2026

Prepared by:
Shad Mahlum
Cindy Baker

For any information regarding this report please contact:
Shad Mahlum
Quantitative Freshwater Ecologist
Freshwater Ecology
+64 7 321 5641
shad.mahlum@earthsciences.nz

New Zealand Institute for Earth Science Limited
PO Box 11115
Hamilton 3251

Phone +64 7 856 7026

Client Report No: 2026173HN
Report date: June 2026
Project No: DNZ26201

Revision	Description	Date
Version 0.1	Draft in preparation for client input	13/05/2026
Version 1.0	Final version sent to client	22/06/2026

Quality Assurance Statement		
	Reviewed by:	Mike Hickford
	Formatting checked by:	Melissa Greentree
	Approved for release by:	Michael Bruce

Contents

Tables

Figures

Executive summary

Freshwater fish populations are under pressure from human-driven stressors such as habitat modification, water abstraction, pollution, and climate change. Effective management depends on timely, reliable monitoring tools. Environmental DNA (eDNA) has emerged as a powerful method for detecting aquatic species, but its value for decision-making hinges on whether sampling approaches are robust and practical.

This report evaluates how reducing eDNA sampling effort, from the standard six-replicates to as few as one, affects ecological assessments. Specifically, it examines impacts on species richness, eDNA-based Fish Index of Biotic Integrity (eFIBI) scores (an index of fish community condition) and resulting management classifications.

Overall, increasing the number of replicates improves the accuracy of species detection and ecological condition estimates. While reduced sampling (two- to five-replicates) produces lower species richness and eFIBI scores, it generally maintains strong agreement with standard six-replicate classifications. As a result, management decisions based on these reduced designs are often comparable. In contrast, single-replicate sampling consistently underestimated biodiversity and ecosystem condition, and is not suitable for reliable monitoring.

The findings support a tiered approach to sampling design:

1. Six-replicate eDNA samples are necessary when monitoring is mandated by the NPSFM and consents or any central and local government context.
2. Single-replicate sampling is not recommended if wanting to calculate fish community metrics (e.g., species richness or indexes of biological integrity).
3. Two- to three-replicates may be sufficient for broad-scale assessments of ecosystem condition.
4. Full replication (six samples) should be retained when monitoring targets rare, threatened, or invasive species.
5. Replication should be increased in larger or more complex stream systems.
6. Results near eFIBI classification thresholds should be treated with caution and may require additional replication.

Decision-support tools developed in this study help identify when reduced sampling is appropriate, based on stream characteristics and uncertainty thresholds. These tools enable practitioners to balance cost-efficiency with confidence, ensuring that eDNA sampling strategies remain appropriate for a range of management contexts.

Future opportunities

Environmental DNA is rapidly becoming an important tool for assessing freshwater biological communities. Compared with conventional survey techniques, eDNA often provides higher detection sensitivity for aquatic species, particularly for rare, cryptic, or low-abundance taxa. In New Zealand, however, the application of eDNA to understand species distributions, habitat relationships, ecological condition, and responses to anthropogenic stressors at landscape scales is still developing. While eDNA provides substantial opportunities to improve biodiversity assessment and ecological monitoring, it is most effective when used alongside complementary ecological indicators and conventional survey approaches that provide information on abundance, growth, recruitment, condition, and other sublethal biological responses.

With respect to fish populations, research and method development are still required before eDNA can be routinely implemented as an integrated monitoring and management tool (Figure 1-1). Four interrelated areas of development are central to this goal. First, the biases and drivers of eDNA detection need to be better understood so that sampling protocols can be tailored to different monitoring contexts, as addressed in this report. Second, species distribution models built from eDNA data, combined with catchment connectivity information, can establish expected fish communities against which observed communities are compared. Third, linking these comparisons to anthropogenic stressors such as land-use intensity, water quality, and habitat modification will allow ecological thresholds to be identified. Finally, this stressor-response understanding can be used to prioritise restoration actions and evaluate their effectiveness over time. Together, these developments would provide a framework for using eDNA not just to describe fish communities, but to diagnose why they differ from expected condition and to identify the management actions most likely to improve them.

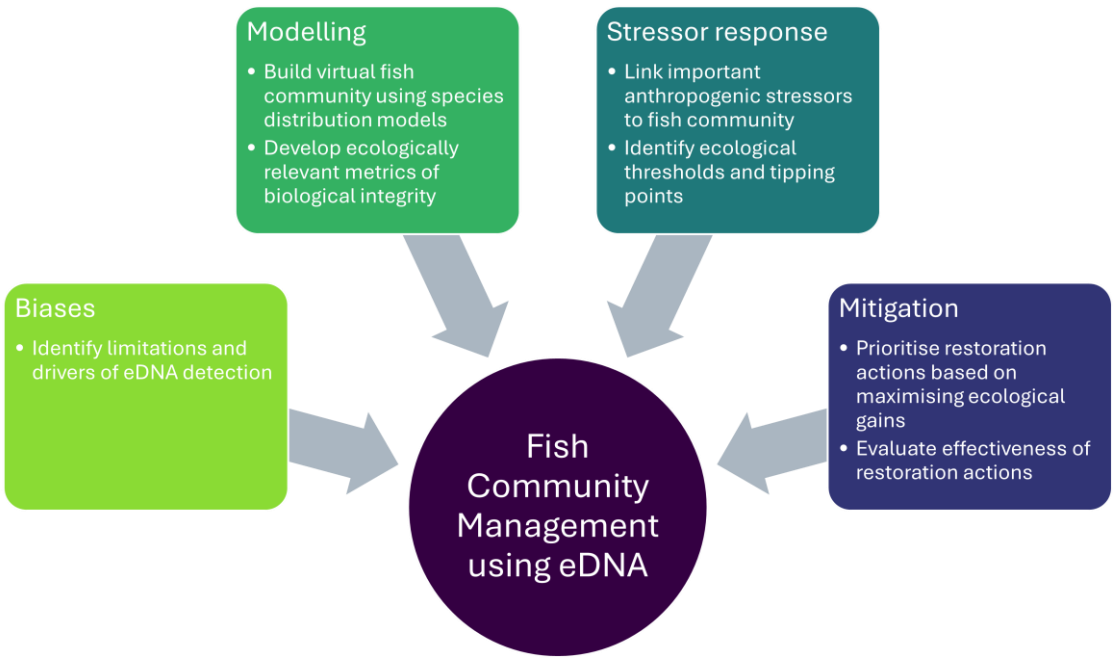

Figure 1-1: Workflow and key opportunities for future development of eDNA as a tool for fish community management.

1 Introduction

Fish populations are increasingly exposed to a suite of anthropogenic stressors, including habitat modification, water abstraction, pollution, and climate-driven changes to thermal and hydrologic regimes, making the timely identification of threatened populations a central challenge for aquatic management and conservation (Reid et al. 2019). Environmental DNA (eDNA) has fundamentally altered freshwater biomonitoring by enabling rapid, non-invasive detection of aquatic taxa, thereby expanding the ability of managers, consultants, and other practitioners to characterise fish communities across broad spatial and temporal scales. However, the usefulness of eDNA for management depends not only on its sensitivity, but also on whether sampling designs are practical, affordable, and robust enough to support real-world decision-making.

As with any survey method, the reliability of eDNA for describing fish assemblages depends strongly on sampling design. Species detection using eDNA is influenced by a range of environmental factors (e.g., discharge, temperature, and pH), methodological factors (e.g., sample volume and the number of replicate samples; Guillera-Arroita et al. 2017; Burian et al. 2021), and assay type and laboratory protocols (Shu et al. 2020; Melchior and Baker 2023). In New Zealand, Smith et al. (2024) found that six-sample replicates were required to consistently detect 90 % of the fish species that eDNA could detect at a given site. Based on this evidence, Melchior and Baker (2023) recommended a six-replicate sampling design for fish eDNA surveys, particularly for compliance and regulatory applications. To further limit the risk of false positives in these contexts, detections are often considered valid only when a species is present in at least two- of six-replicates (Melchior and Baker 2023).

While this sampling design provides confidence in species detection, it is currently costly and logistically demanding. These constraints may limit the feasibility or spatial extent of eDNA surveys for users such as smaller agricultural enterprises, community groups, or consultancies operating outside of formal regulatory frameworks. In many cases, the primary objective of sampling may simply be to understand which fish species are present, rather than to generate data suitable for statutory reporting. Under these circumstances, it is unclear whether six-replicate samples are necessary, or whether reduced sampling effort could provide sufficiently robust information while substantially lowering costs. The growing availability of large eDNA datasets spanning broad spatial extents in New Zealand now provides an opportunity to quantify the bias and uncertainty associated with using fewer replicate samples.

In addition to species richness, there is increasing interest in using eDNA data to support broader assessments of ecological condition. Indices of biological integrity (IBIs) are widely used to summarise fish community condition from conventional survey data (Fore et al. 1994; Ganasan and Hughes 1998; Mebane et al. 2003; Joy and Death 2004; Pont et al. 2007), and more recently have been adapted for use with eDNA detections (Curtis et al. 2026). In New Zealand, this includes the development of an eDNA-based Fish Index of Biotic Integrity (eFIBI; White et al. 2025) derived from the original Fish IBI (Joy and Death 2004). These indices provide a more interpretable way to assess fish community condition than species lists alone and are increasingly relevant in management and reporting contexts. While absolute differences in IBI scores are likely to occur under different eDNA sampling designs, it is unclear whether these differences translate into changes in the broader condition categories commonly used to interpret stream health. Thus, it is important to understand how changes in eDNA sampling effort may influence raw eFIBI scores and the condition classifications derived from them.

Here, we evaluate the bias and usefulness of alternative eDNA sampling regimes by comparing reduced replicate designs (one- to five-replicates) against the standard six-replicate eDNA protocol. Specifically, we assess how reductions in replicate sampling influence:

1. estimated species richness,
2. eFIBI scores, and
3. eFIBI quality classifications.

1.1 Approach

To evaluate the bias of alternative eDNA sampling regimes, we used the following approach:

1. A simulation framework based on observed six-replicate samples was used to quantify how reduced replication affects these metrics. We then evaluated the degree of agreement between reduced-replicates and six-replicate outcomes and identified environmental conditions that influenced the likelihood of agreement.
2. We then quantified the uncertainty and bias associated with reduced replication, particularly in relation to eFIBI scores and the risk of misclassification across management thresholds (e.g., eFIBI quality bands).
3. We then synthesised these results to provide practical guidance on when reduced replication may be appropriate, and when full six-replicate sampling is required, based on stream characteristics and monitoring objectives.
4. Finally, because the interpretation of any fish community survey depends on understanding what species could reasonably be expected at a given location, we review the complementary tools available for estimating potential fish presence in the context of catchment connectivity assessments. Tools focused on desktop assessments including empirical records from the Wilderlab eDNA database and the New Zealand Freshwater Fish Database, and the predictive distribution models. We discuss how these can be integrated to support the design and interpretation of eDNA surveys across a range of management contexts.

2 Methods

2.1 Existing environmental DNA Data

Environmental DNA data were extracted from the Wilderlab database. Only samples that included a six-replicate eDNA protocol were used in the study and the samples needed to include GPS coordinates within 200 m of the digital river network (DN2) and the water volume of each replicate. This resulted in 1049 samples across the North and South Island (Figure 2-1)

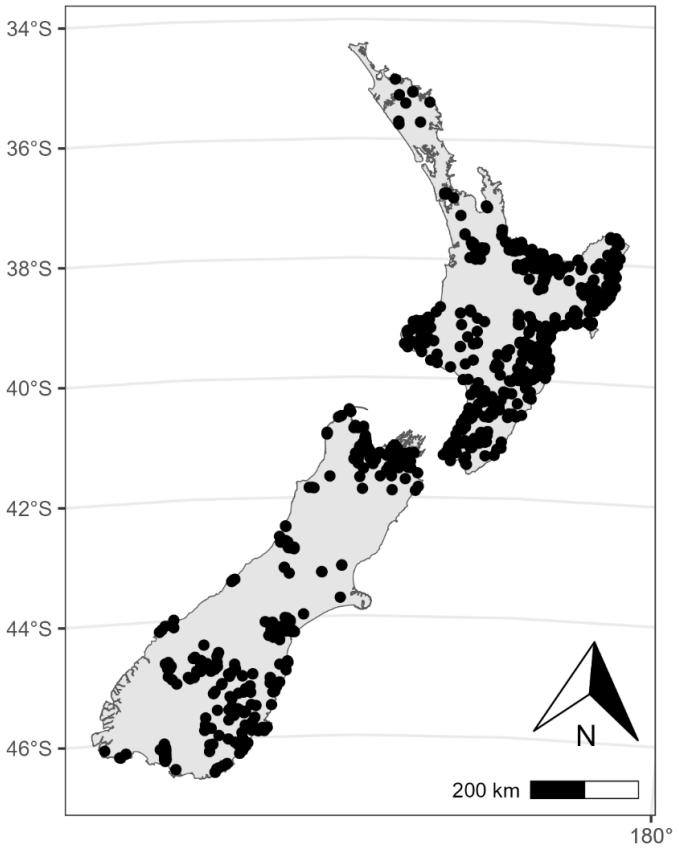

Figure 2-1: Location of the six-replicate environmental DNA across New Zealand.

Environmental DNA samples with six-replicates were considered the standard recommended sampling protocol and for the purposes of the simulation were considered the 'true' fish community detected within a sample. Based on the recommendations of Melchior and Baker (2023), for a fish species to be considered a 'true' detection in the sample, it must be detected in two- out of the six-replicates. However, this threshold is not feasible for the inclusion of one- to five-replicate comparisons. Therefore, for simulations using one- to five-replicate samples, a 'true' detection consisted of a species being present in at least 30% of the replicates. The number of reads required in each replicate was set at sequence count ≥ 5 following the protocol of Smith et al. (2024).

2.1.1 eDNA Simulations

Using a six-replicate eDNA sample, a total of 62 replicate combinations are possible across the different sampling scenarios (Table 2-1). Instead of randomising which replicate is included in a

scenario, all possible replicate scenarios were used and a species richness and corresponding eFIBI was calculated for each replicate scenario.

Table 2-1: Number of unique combinations for each replicate using a 6-replicate eDNA sample.

Replicate	Number of unique combinations
1	6
2	15
3	20
4	15
5	6

2.1.2 eDNA Fish Index of Biological Integrity (eFIBI)

The eDNA Fish Index of Biological Integrity (eFIBI; White et al. 2025) is an updated index for eDNA based on the Fish Index of Biological Integrity developed by Joy and Death (Fish IBI; 2004). The Fish IBI was developed using electric fishing and netting/trapping data held in the New Zealand Freshwater Fish Database. The index ranges from 0 to 60, with values closer to 0 indicating a poor fish community and 60 indicating a near-pristine fish community, relative to the sampling site's location (distance from the sea and elevation). We compared how eFIBI scores derived from different eDNA replication levels (1 – 5) aligned with the recommended six-replicate protocol.

eFIBI quality score

The IBI is a principal metric used in the National Policy Statement for Freshwater Management (NPSFM; MfE 2020). Within the NPSFM (2020) the eFIBI score can be classified into sites of high integrity ≥ 34, medium integrity < 34 and ≥ 28, low integrity < 28 and ≥ 18, and severe loss or poor integrity < 18. An important caveat when using eDNA data is that these bands were not derived from ecologically defined thresholds; rather, they reflect the distribution of Fish IBI scores based on physical capture data, with breakpoints set approximately at the 25th, 50th, and 75th percentiles. In species rich communities, White et al. (2025) found that eDNA sampling can inflate eFIBI scores to around 1.5 times those obtained using conventional electric fishing. For example, in stream reaches where electric fishing produced an IBI score of 40, eDNA sampling may generate an eFIBI score of 60 for the same reach.

Although the percentile thresholds may differ for eFIBI, comparison against the current classification bands provides a useful framework for assessing agreement at a broader management-relevant scale. It also allows results to be interpreted within the existing regulatory context to which stakeholders must adhere. Consequently, to assess eFIBI quality, we also evaluated the congruence of the broader NPSFM (2020) IBI integrity classifications across the five reduced replication categories relative to the standard six-replicate eDNA sampling protocol.

2.2 Probability of agreement

2.2.1 Covariates

The presence of a species in eDNA sampling is driven by a range of factors that affect our detection efficacy. This includes, limiting temporal sampling to periods when a species is present in the study reach (for migratory species) or during periods of increased activity (e.g., summer; Mahlum et al. 2025). The environmental conditions of the river can also play an important role in detecting species with eDNA. Therefore, it is expected that a range of factors associated with the sampling methodology (e.g., number of eDNA replicates and volume of water taken), as well as river conditions (e.g., river flow, turbidity and water temperature) will partly determine the effectiveness of eDNA sampling and whether a reduced number of replicates can reliably predict the fish community at a given site (Table 2-2). The selected covariates used here are based on the dendritic river network version 2.4 (DN2) and each variable is summarised at the segment level. Because eDNA is a summary of the upstream community, the metrics were also recalculated to take the mean of the upstream segments weighted by the segment length. Summaries of the upstream processes were found to better relate to sampling techniques that sample upstream populations (e.g., eDNA and POCIS sampling for lamprey; Mahlum et al. 2024).

Table 2-2: A priori selected covariates used to model how one- to five-replicate eDNA samples predict species abundance, eFIBI, and eFIBI quality.

Covariate	Mean (sd)	Type	Transformation	Description
Replicate	NA	Categorical	None	Number of eDNA replicates ranging from 1 - 5
Distance	58 (71)	Continuous	None	Distance from the stream segment to the ocean (km)
Elevation	140 (189)	Continuous	None	Mean elevation of the stream segment (m)
Sediment Load	0.167 (1.064)	Continuous	Log	Suspended sediment load (Mt/year) including the influence of lakes
Stream Flow	7.56 (41.72)	Continuous	Log	Mean flow over all time (m^3/s)
Stream Order	4 (1.6)	Continuous	None	Strahler stream order
Stream Slope	1.1 (1.8)	Continuous	None	Mean segment slope (m/m)
Stream Width	11 (13)	Continuous	None	Wetted width across the river channel (m) at mean flow
Sample Volume	893 (214)	Continuous	None	Mean volume sampled across replicates (ml)

2.2.2 Probability of agreement analysis

The subsampling analysis is entirely based on observed data using all replicate combinations (Table 2-1). This provides an empirical assessment of how reduced replication affects observed outcomes. Generalized mixed-effects models were then fitted to these simulation outputs to generalise patterns, estimate probabilities of agreement with six-replicate samples, and evaluate how environmental covariates influence performance that a one- to five-replicate eDNA sample can predict the species richness, eFIBI, and eFIBI integrity classification (quality). Using the six-replicate eDNA samples, a total of 62 unique trials were created (Table 2-1). Each model included three main terms, the first was a categorical variable distinguishing between one- through five-replicate samples drawn. Sampling protocols specify that each replicate

should consist of a 1000 ml water sample; however, it is not always possible to collect this volume due to factors such as elevated sedimentation. Therefore, each model also included a fixed effect of sample volume. Finally, to account for the repeated sampling at different temporal and spatial extents, the model also included two nested random terms for year/month and region/catchment/site.

A range of factors can affect the quality and likelihood of an eDNA sample detecting a species, and it is anticipated that these factors can drive the many differences observed between a standard six-replicate sample and samples that use fewer replications. A range of covariates were *a priori* selected to investigate factors that may influence fish detection across different numbers of eDNA replicates. The Akaike Information Criterion (AIC) was used to identify the combination of covariates that best explained the data (Burnham and Anderson 2002).

2.3 Quantifying uncertainty and determining replication requirements

Based on the bias in eFIBI scores associated with reduced replication and its potential impact on eFIBI quality classification, we also quantified the uncertainty in eFIBI scores for one- to five-replicate samples. Differences between reduced-replicate and six-replicate eFIBI values were first calculated for each simulated sample, and absolute mean differences per site/replicate were then modelled using a Gaussian generalised linear mixed effects model with replication level as a fixed effect. To account for repeated sampling across temporal and spatial scales, the model also included two nested random effects for year/month and region/catchment. Predicted median differences and 95% confidence intervals were extracted for each replication level.

We then generated a reference eFIBI range spanning all possible values (0–60) and, for each value, calculated an interval defined by adding and subtracting the predicted median difference and its confidence bounds. Values were flagged for potential verification where these intervals overlapped established NPSFM (MfE 2020) eFIBI category thresholds (A: ≥34, B: 28–33, C: 18–27, D: <18). This approach identified the combinations of eFIBI scores and replication levels where uncertainty from reduced replication could lead to misclassification of ecological condition, thereby informing when a six-replicate sample is required for confirmation.

Stream conditions can also play a role in eDNA detection. Therefore, using the top modelled probability of agreement eFIBI model, we predicted probabilities for each replication level across combinations of mean upstream stream width and stream order. By identifying the replication level that maximises probability for a given stream width and stream order, end-users can determine when additional replicates may be required to improve the likelihood of accurately predicting eFIBI for a given the size and order.

3 Results

3.1 Simulation and probability of agreement

3.1.1 Species richness

Simulation results for species richness

Across all sampling locations, the median number of species detected using a standard six-replicate sample was 5 (sd = 2.8; bottom right panel; Figure 3-1) with species richness increasing with stream order (Figure 3-3). Absolute differences in species richness showed that both the two-replicate and five-replicate eDNA samples had a median difference of zero relative to the six-replicate baseline (Figure 3-2). However, because the five-replicate approach still met the sample validity threshold (two-replicates) used for the standard six-replicate design, it was prone to underestimating site-level species richness (Figure 3-2).

The two-replicate sample exhibited an approximately normal distribution of differences centred on zero, with limited bias in either direction (lower quantile = -1.07, upper quantile = 0.93) when compared with the standard six-replicate baseline. In contrast, the one- and four-replicate samples showed a tendency toward negative bias, with median differences of -1.33 and -0.2, respectively. Only the three-replicate samples exhibited a slight positive bias, with a median difference of 0.28 (lower quantile = -0.4, upper quantile = 1.75; Figure 3-2). Occasional positive differences (i.e., higher richness) arise from differences in detection thresholds (e.g., true detection = being detected in 30% of the available replications) between the six-replicate and reduced-replicate scenarios, combined with stochastic variation in replicate composition. These differences do not indicate improved detection performance but reflect how threshold-based filtering interacts with subsampling.

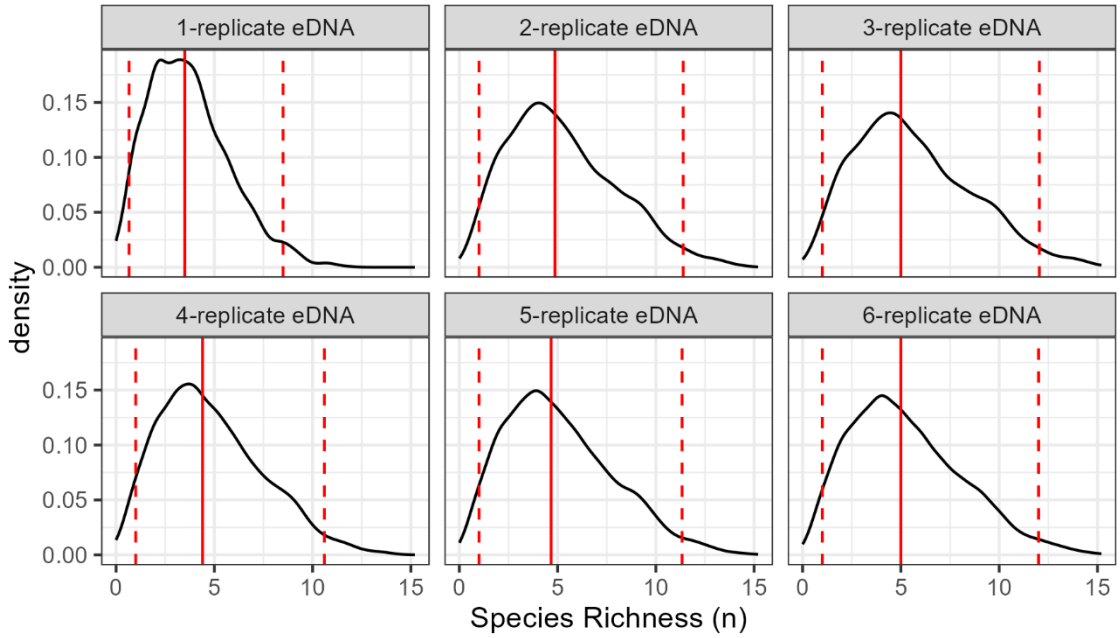

Figure 3-1: Density plot of species richness across the different replication levels. The solid red line is the median value across samples, and the dashed red lines are the 2.5% and 97.5% quantile.

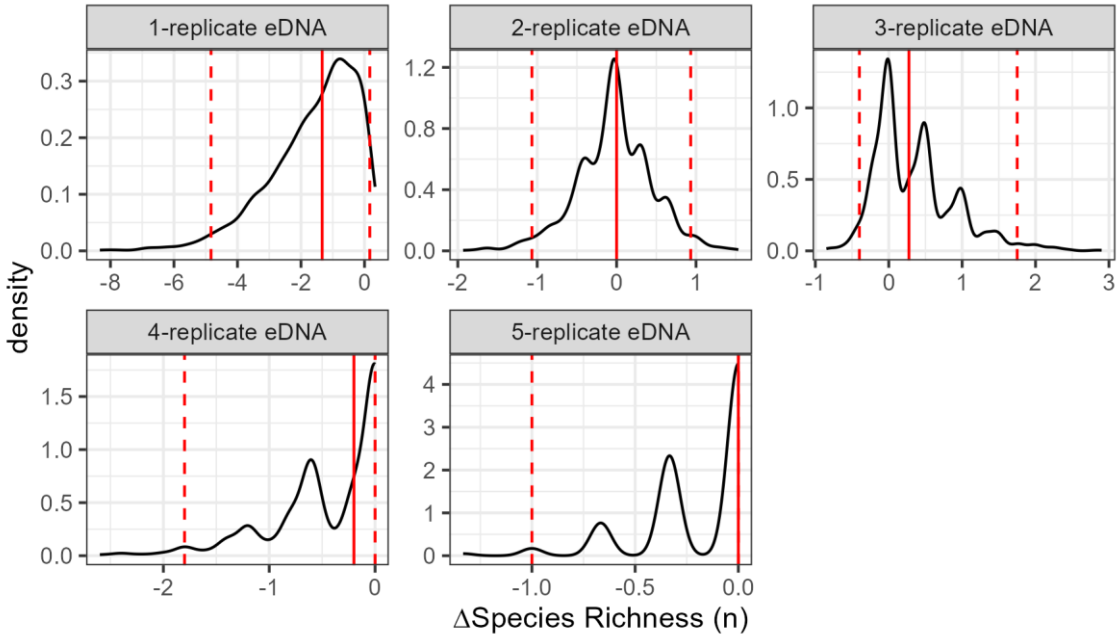

Figure 3-2: Difference in calculated species richness relative to the species richness based on a six-replicate eDNA. Solid line is the median difference, and the dashed lines are the 2.5 and 97.5 quantiles.

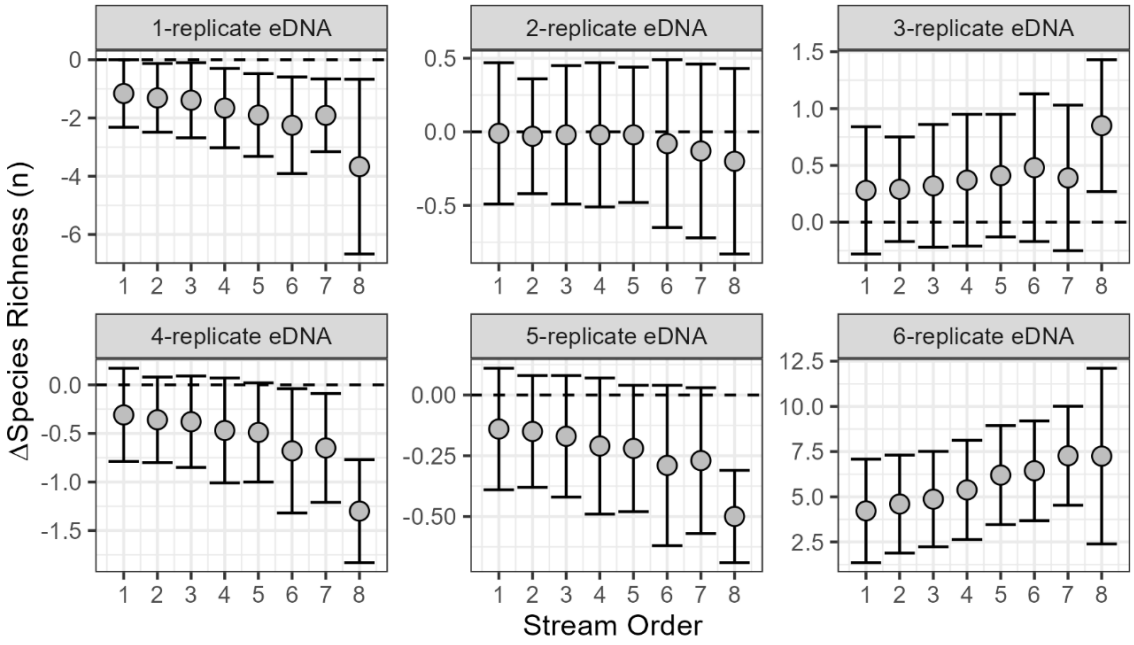

Figure 3-3: Difference in species richness across stream orders and eDNA replicates relative to a standard six-replicate eDNA sample. Points represent the mean difference, and error bars are ±1 standard deviation. The dashed line indicates no difference in species richness from the six-replicate eDNA sample baseline (bottom right panel).

Modelled probability of agreement for species richness

Model selection using AIC identified a single top-performing model (ΔAIC ≤ 2), which accounted for 99.6% of the total model weight (Table A-1). The selected model included replication, stream order, segment slope, stream width, and interactions between replication and each of these habitat covariates (Table 3-1).

Increases in stream order were associated with a reduced probability of achieving equivalent species richness (OR = 0.71, $p < 0.001$), whereas higher slope was associated with increased odds of equivalence (OR = 1.23, $p = 0.001$). Stream width showed a negative but non-significant main effect (OR = 0.94, $p = 0.376$).

Relative to a single replicate, the probability of achieving equivalent species richness increased markedly with additional replications, with odds ratios ranging from 8.51 for two-replicates to 37.7 for five-replicates (all $p < 0.001$). However, the effects of slope and stream width varied with replication, with the positive influence of slope and the negative influence of stream width diminishing as replicate number increased.

The number of eDNA replicates had a strong positive effect on detection performance (Figure 3-3). Compared with a single-replicate, all higher replicate levels significantly increased the probability of matching the six-replicate benchmark, with odds ratios increasing from 8.51 for two-replicates and 8.84 for three-replicates to 12.01 for four-replicates and 37.7 for five-replicates (all $p < 0.001$; Figure 3-4). While most replicate levels differed significantly from one another, no significant difference was observed between the two- and three-replicate samples in their ability to reproduce species richness equivalent to the six-replicate standard (Figure 3-4).

Table 3-1: Fixed-effect estimates from a binomial GLMM of detection probability of sample species richness for one- to five-replicate eDNA samples relative to the standard six-replicate sample. Odds ratios greater than 1 indicate an increased likelihood of matching the six-replicate reference sample, whereas values less than 1 indicate reduced likelihood, with effect sizes interpreted relative to the reference category or per unit increase in continuous predictors. Bold p-values indicate a significant effect.

Variable	Estimate	SE	Odd ratio	p-value
Intercept	-1.674	0.134	0.19	**< 0.001**
Volume	0.019	0.053	1.02	0.724
Two-replicate	2.141	0.043	8.51	**< 0.001**
Three-replicate	2.179	0.042	8.84	**< 0.001**
Four-replicate	2.486	0.043	12.01	**< 0.001**
Fiver-replicate	3.63	0.054	37.7	**< 0.001**
Stream order	-0.341	0.077	0.71	**< 0.001**
Slope	0.203	0.063	1.23	**0.001**
Stream Width	-0.067	0.075	0.94	0.376
Two-replicate: Stream order	0.065	0.053	1.07	0.226
Three-replicate: Stream order	0.047	0.052	1.05	0.371
Four-replicate: Stream order	0.049	0.054	1.05	0.359
Five-replicate: Stream order	0.071	0.065	1.07	0.28
Two-replicate: Slope	-0.123	0.042	0.88	**0.004**
Three-replicate: Slope	-0.104	0.041	0.9	**0.012**
Four-replicate: Slope	-0.086	0.043	0.92	**0.045**
Five-replicate: Slope	-0.069	0.059	0.93	0.24
Two-replicate: Stream width	-0.128	0.054	0.88	**0.017**
Three-replicate: Stream width	-0.117	0.053	0.89	**0.026**
Four-replicate: Stream width	-0.177	0.054	0.84	**< 0.001**
Five-replicate: Stream width	-0.157	0.062	0.85	**0.011**

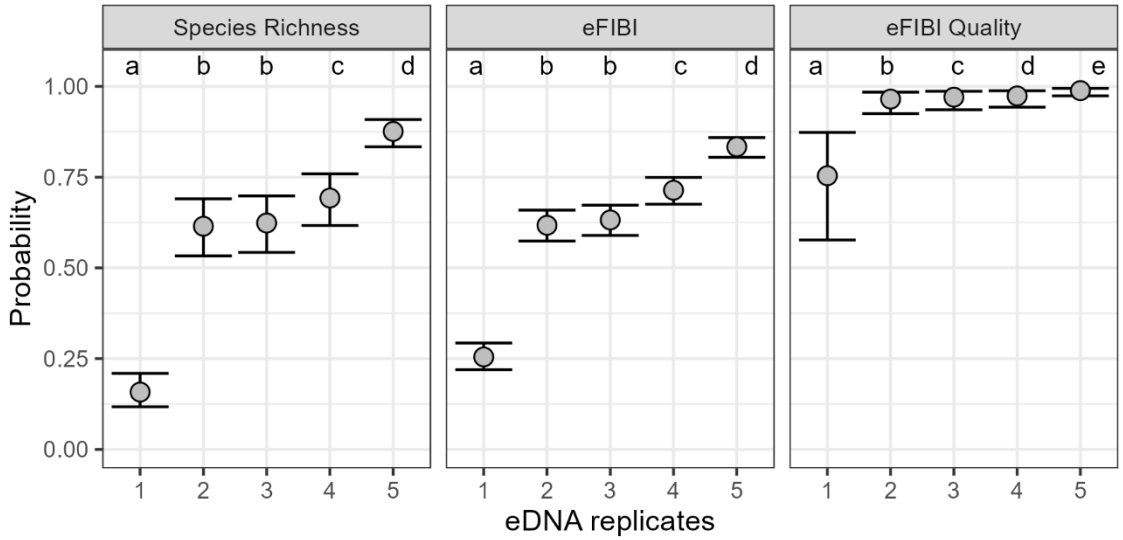

Figure 3-4: Predicted probability that a one- through five-replicate eDNA sample can accurately predict the species richness, eFIBI, and eFIBI quality based on the standard six-replicate eDNA sample. Points are the predicted probability with the error bars representing 95% confidence intervals. Replicates with the same letters indicate non-significant comparisons at α = 0.05.

3.1.2 eFIBI

Simulation results for eFIBI

The median eFIBI based on the standard six-replicate eDNA sample was 36, with 95% of values ranging from 16 to 56 (Figure 3-5). The two- and five-replicate eDNA samples showed distributions more evenly centred around zero compared to the one-, three-, and four-replicate samples (Figure 3-6). Relative to the six-replicate baseline, the three-replicate samples exhibited a right-skewed distribution, whereas the one- and four-replicate samples were left-skewed.

The eFIBI increased with stream order (Figure 3-7). Differences among replicate levels relative to stream order were generally close to zero, although the one-replicate samples showed the strongest negative bias (median ΔeFIBI = -5.33). The two-replicate samples were the least biased across stream order, while the three-replicate samples were the only level to exhibit a positive bias (Figure 3-7).

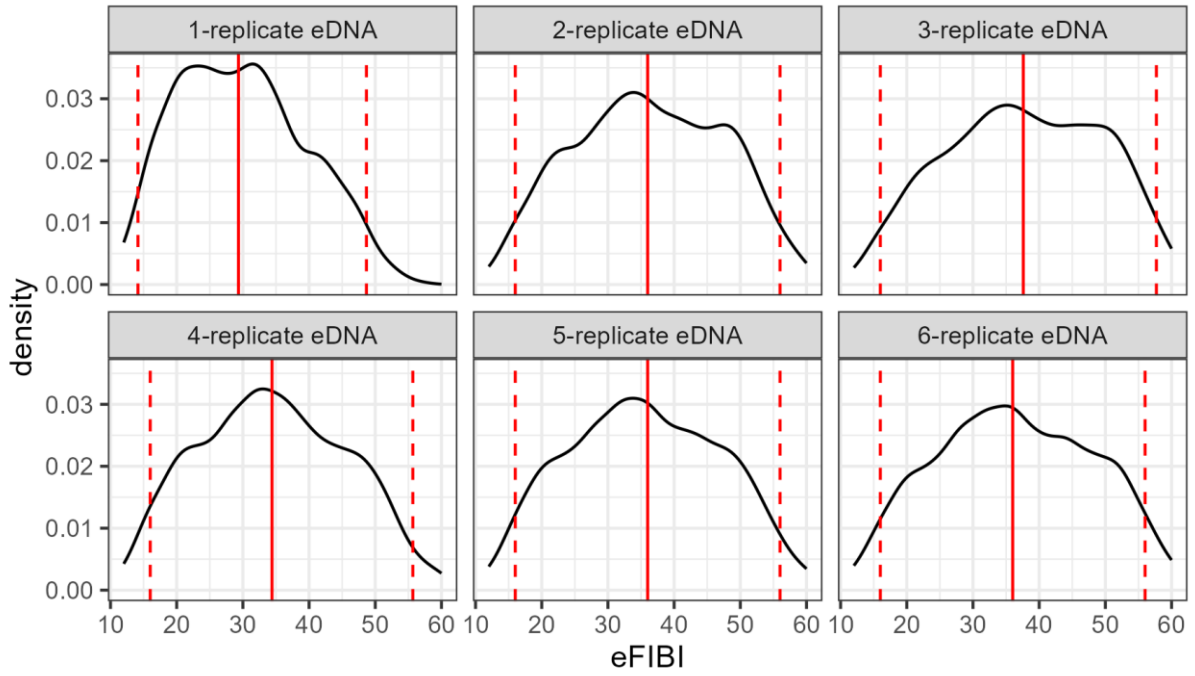

Figure 3-5: Density plot of eFIBI scores across the different replication levels. The solid red line is the median value across samples, and the dashed red lines are the 2.5% and 97.5% quantile.

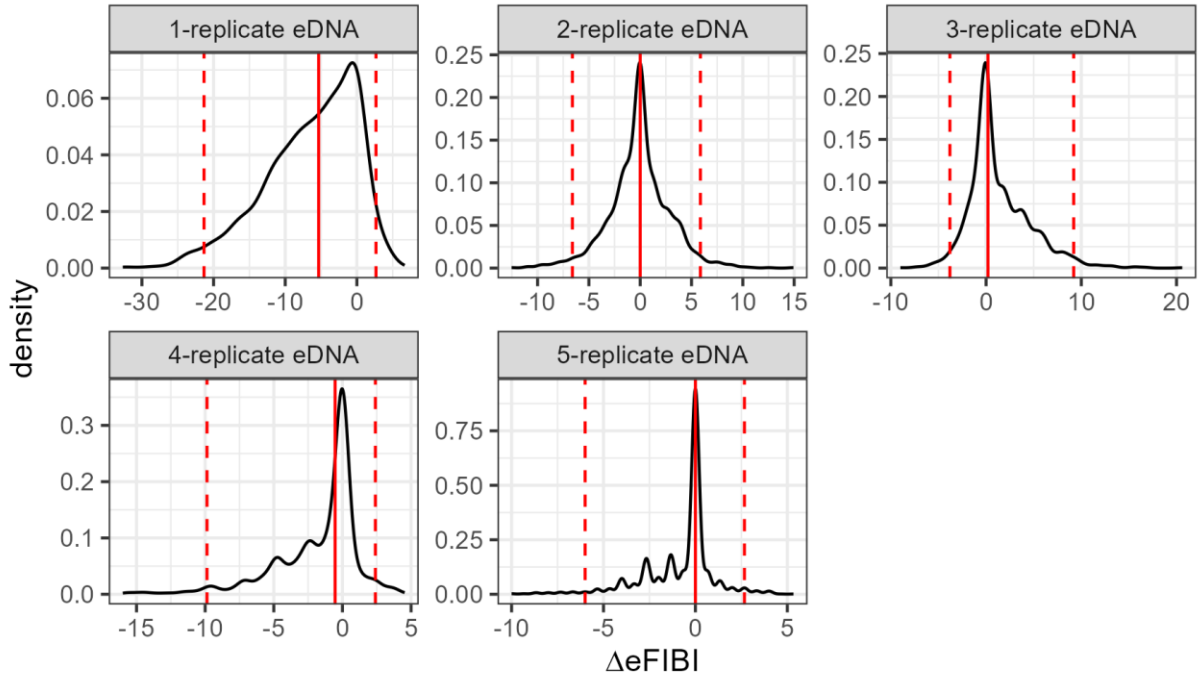

Figure 3-6: Difference in calculated eFIBI relative to the eFIBI based on a six-replicate eDNA. Solid line is the median difference, and the dashed lines are the 2.5 and 97.5 quantiles.

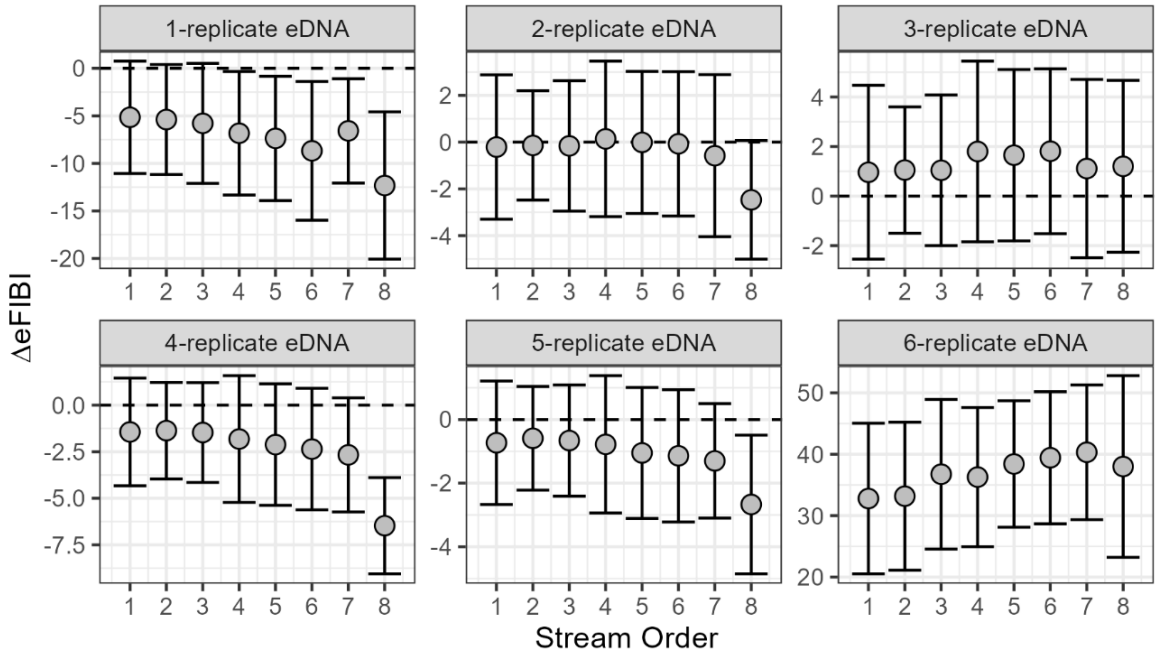

Figure 3-7: The difference in eFIBI scores across stream orders (replicates 1-5) relative to the standard six-replicate eDNA sample (bottom right panel). Error bars are ±1-standard deviation and horizontal line is for reference with values below the line indicating a mean under-representation of the eFIBI and points above the line indicating over-representation of the eFIBI.

Modelled probability of agreement for eFIBI

Model selection using AIC identified a single top-performing model (ΔAIC ≤ 2), which accounted for 78% of the model weight (Table A-2). The selected model included mean upstream stream width and slope, as well as interaction terms between replicate number and both stream width and slope, in predicting the probability that eFIBI matched the six-replicate eDNA reference sample (Table 3-2).

Consistent with the species richness results, the probability that eFIBI derived from reduced replication matched the six-replicate reference increased with the number of eDNA replicates (middle panel, Figure 3-4). Relative to a single-replicate, all replicate levels significantly increased the likelihood of agreement with the six-replicate benchmark, with odds ratios increasing from 4.72 for two-replicates to 14.66 for five-replicates (all $p < 0.001$; Table 3-2). No significant difference was detected between the two- and three-replicate samples, whereas higher replicate levels differed significantly in their ability to reproduce the six-replicate eFIBI (middle panel, Figure 3-3).

Interaction terms between replicate number and environmental covariates were generally non-significant, suggesting that the positive effect of increased replication on eFIBI performance was broadly consistent across gradients of and stream order, slope, and width.

Table 3-2: Fixed-effect estimates from a binomial GLMM of detection probability to estimate the eFIBI for one- to five-replicate eDNA samples relative to the standard six-replicate eDNA sample. Odds ratios greater than 1 indicate an increased likelihood of matching the six-replicate reference sample, whereas values less than 1 indicate reduced likelihood, with effect sizes interpreted relative to the reference category or per unit increase in continuous predictors. Bold p-values indicate a significant effect.

Variable	Estimate	SE	Odd Ratio	p-value
Intercept	-1.074	0.075	0.34	**< 0.001**
Volume	-0.02	0.055	0.98	0.714
Two-replicate	1.552	0.039	4.72	**< 0.001**
Three-replicate	1.615	0.038	5.03	**< 0.001**
Four-replicate	1.989	0.039	7.3	**< 0.001**
Five-replicate	2.685	0.049	14.66	**< 0.001**
Stream Order	-0.204	0.068	0.82	**0.003**
Slope.us	0.075	0.07	1.08	0.286
Stream width.us	-0.253	0.074	0.78	**< 0.001**
Two-replicate: Stream Order	-0.06	0.043	0.94	0.161
Three-replicate: Stream Order	-0.07	0.041	0.93	0.092
Four-replicate: Stream Order	0.002	0.043	1	0.956
Five-replicate: Stream Order	-0.063	0.053	0.94	0.229
Two-replicate: Slope.us	-0.063	0.042	0.94	0.135
Three-replicate: Slope.us	-0.04	0.041	0.96	0.326
Four-replicate: Slope.us	-0.096	0.042	0.91	**0.024**
Five-replicate: Slope.us	-0.173	0.053	0.84	**0.001**
Two-replicate: Stream width.us	0.064	0.047	1.07	0.174
Three-replicate: Stream width.us	0.053	0.045	1.05	0.242
Four-replicate: Stream width.us	0.063	0.047	1.06	0.182
Five-replicate: Stream width.us	0.138	0.057	1.15	**0.015**

3.1.3 eFIBI quality: Modelled probability of agreement

Model selection using AIC identified a single top-performing model for predicting eFIBI quality classes ($\Delta AIC \leq 2$), which accounted for 100% of the model weight (Table A-3). The selected model included mean segment slope, stream width, and a quadratic term for log-transformed sediment load at the sampled reach. Together, these results suggest that reach-scale geomorphic conditions influence the probability of correctly classifying eFIBI quality relative to the standard six-replicate reference (Table 3-3).

However, the main effects of slope, log-transformed sediment, its quadratic term, and stream width were not statistically significant, indicating limited evidence that these covariates independently affect classification accuracy (Table 3-3). In particular, the quadratic sediment term was non-significant (OR = 0.71, $p = 0.439$), providing little support for a non-linear relationship between sediment load and classification performance.

Despite this, several interaction terms suggested that environmental influences varied with replication level. The quadratic sediment interaction was significant for three-replicates (OR = 1.63, $p = 0.048$), while stream width interactions were significant for two-replicates (OR = 0.87, $p = 0.047$) and four-replicates (OR = 0.86, $p = 0.035$). Nonetheless, the dominant pattern was a strong increase in classification accuracy with increasing eDNA replication (right panel Figure 3-4), with odds of matching the six-replicate standard increasing from 9.02 for two-replicates to 27.8 for five-replicates (all $p < 0.001$).

Table 3-3: Fixed-effect estimates from a binomial GLMM of detection probability to estimate the eFIBI quality (high, medium, low, and poor) for one- to five-replicate eDNA samples relative to the standard six-replicate eDNA sample. Odds ratios greater than 1 indicate an increased likelihood of matching the six-replicate reference sample, whereas values less than 1 indicate reduced likelihood, with effect sizes interpreted relative to the reference category or per unit increase in continuous predictors. Bold p-values indicate a significant effect.

Variable	Estimate	SE	Odd Ratio	p-value
Intercept	1.12	0.315	3.06	**< 0.001**
Volume	0.188	0.085	1.21	**0.027**
Two-replicate	2.199	0.048	9.02	**< 0.001**
Three-replicate	2.37	0.047	10.69	**< 0.001**
Four-replicate	2.496	0.049	12.13	**< 0.001**
Five-replicate	3.325	0.07	27.8	**< 0.001**
Slope	0.137	0.09	1.15	0.127
Log(Sediment)	0.077	0.402	1.08	0.848
I(Log(Sediment)^2)	-0.339	0.439	0.71	0.439
Stream width	-0.097	0.134	0.91	0.469
Two-replicate: Slope	-0.008	0.053	0.99	0.877
Three-replicate: Slope	0.058	0.053	1.06	0.271
Four-replicate: Slope	0.04	0.057	1.04	0.483
Five-replicate: Slope	0.102	0.089	1.11	0.256
Two-replicate: Log(Sediment)	-0.036	0.237	0.96	0.879
Three-replicate: Log(Sediment)	-0.092	0.23	0.91	0.69
Four-replicate: Log(Sediment)	0.209	0.241	1.23	0.386
Five-replicate: Log(Sediment)	-0.045	0.346	0.96	0.897
Two-replicate: I(Log(Sediment)^2)	0.427	0.253	1.53	0.092
Three-replicate: I(Log(Sediment)^2)	0.486	0.246	1.63	**0.048**
Four-replicate: I(Log(Sediment)^2)	0.073	0.256	1.08	0.777
Five-replicate: I(Log(Sediment)^2)	0.306	0.363	1.36	0.399
Two-replicate: Stream width	-0.14	0.07	0.87	**0.047**
Three-replicate: Stream width	-0.083	0.069	0.92	0.233
Four-replicate: Stream width	-0.149	0.071	0.86	**0.035**
Five-replicate: Stream width	-0.063	0.095	0.94	0.509

3.2 Quantifying uncertainty for the eFIBI

Modelled absolute differences in eFIBI scores decreased with increasing eDNA replication, indicating improved precision with additional replicates. Single-replicate samples exhibited the greatest uncertainty (mean absolute difference = 7.32 eFIBI units; 95% CI: 6.52–8.08; Figure 3-8), whereas uncertainty was substantially reduced for two- to four-replicate samples (means = 2.3–2.8 units; Figure 3-8). Five-replicate samples showed the lowest uncertainty (mean = 1.47 units; 95% CI: 0.72–2.23; Figure 3-8).

Mixed-effects modelling confirmed that, after accounting for temporal and spatial variation, all multi-replicate designs significantly reduced uncertainty relative to single-replicate samples ($p < 0.001$). Overall, these results indicate that more than single-replicate samples provides moderate precision for most applications, while higher replication substantially reduces the risk of error leading to overlap with management thresholds in eFIBI scores.

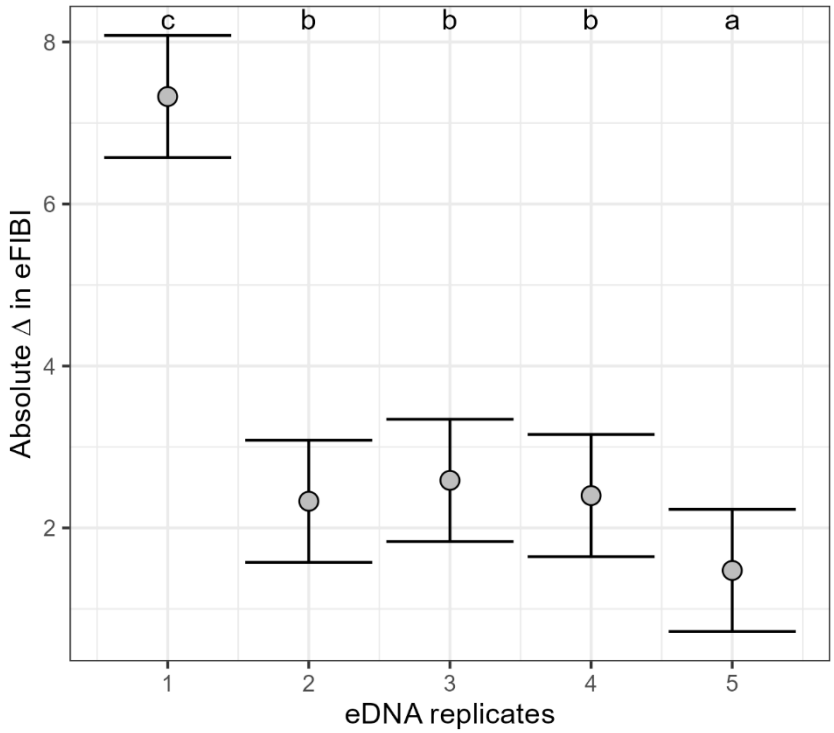

Figure 3-8: Modelled absolute difference in eFIBI calculations relative to the standard six-replicate eDNA sample. Points indicate the modelled mean for each replicate category, and error bars indicate the upper and lower confidence interval at alpha 0.05. Replicates with the same letters indicate non-significant comparisons.

4 Discussion

Analyses showed that increasing the number of replicates improved the ability of both species richness and eFIBI scores to approximate estimates derived from the standard six-replicate eDNA sample. However, eFIBI quality classifications based on two to five replicates showed high agreement with those from six-replicate samples. Although some variation occurred among replicate treatments, absolute differences were generally small, resulting in similar eFIBI scores and, consequently, comparable overall eFIBI quality classifications. As a result, eDNA sampling using two to five replicates would likely support similar management decisions under NPSFM integrity classifications. In contrast, single-replicate eDNA sample consistently showed reduced performance relative to multi-replicate approaches, indicating that caution is needed when relying on single replicates for ecological monitoring or management decision-making.

The six-replicate eDNA sample threshold is based on being able to detect approximately 90% of the freshwater fish species present at a sampling location (Smith et al. 2024). As expected, the likelihood of detecting the full complement of species decreased as the number of replicates decreased. Species most likely not to be detected under lower replication are those that are rare or occur at low abundance. This has important implications for invasive species management, where early detection is essential for preventing spread and establishment. Similarly, threatened species are often naturally rare making it critical to maximise detection probability to inform effective management and conservation strategies.

The modelling results indicate that the performance of reduced-replication sampling may vary across the river network. For both species richness and eFIBI, certain environmental covariates influenced the likelihood of matching the six-replicate reference samples, although these effects were generally weaker than the influence of replicate number itself. Notably, reduced performance in higher-order or wider streams suggests that larger channels or more hydraulically complex environments may require greater replication to achieve comparable confidence to smaller streams. This aligns with the expectation that eDNA in larger or more heterogeneous systems may be more diluted or unevenly distributed, increasing variability among individual water samples. While these findings do not alter the overall conclusion that increased replication improves performance, they do suggest that the optimal level of replication may depend on site conditions and monitoring objectives.

While clear declines in agreement were evident among replicate sampling approaches for both species richness and eFIBI scores, the broad categorisations showed strong alignment when using two or more replicates. This likely reflects the use of relatively wide score bands in the eFIBI framework where small changes in species detection or raw scores are often insufficient to shift a site into a different condition class. As a result, despite reduced sampling effort, the categorical eFIBI outputs remain robust for assessing ecosystem condition in relations to fish communities. However, caution is still required, as eFIBI can be sensitive to the presence of exotic species, which may disproportionately influence condition bands. Sampling protocols should therefore account for this limitation when interpreting results.

5 Summary and recommendations

The results of this study support a tiered approach to eDNA sampling design, where replication requirements are determined by monitoring objectives, acceptable uncertainty, and stream characteristics. Table 5-1 details the expected error in calculating the eFIBI and the average number of units expected to be missed when reduced replicate sampling is carried out.

Table 5-1: The probability and mean error that reduced replicate sampling will misrepresent the eFIBI relative to six-replicate eDNA surveys. Values in parenthesis indicate the upper and lower confidence intervals.

Replicate number	Likelihood of non-agreement with the eFIBI relative to six-replicate sampling	Mean eFIBI error (±) relative to six-replicate sampling
1	75% (71-78)	7.3 (6.6 – 8.1)
2	39% (34 - 43)	2.3 (1.6 – 3.1)
3	37% (33 - 42)	2.6 (1.8 – 3.3)
4	29% (26 - 33)	2.4 (1.6 – 3.2)
5	17% (14 - 20)	1.5 (0.7 – 2.2)

Given the expected error in reducing replicates from six, the following recommendations are intended to guide practitioners in selecting replication effort for freshwater fish eDNA surveys:

1. **Six-replicate eDNA samples are necessary when monitoring is mandated by the NPSFM (MfE 2020) and consents or any central and local government context.** In these situations, adherence to the six-replicate protocol and the two-of-six detection threshold outlined in Melchior and Baker (2023) remains best practice.

2. **Single replicate sampling is not recommended** to calculate fish community metrics (e.g., species richness or indexes of biological integrity), as it shows substantially lower agreement with reference conditions and has a high likelihood of underestimating species richness and eFIBI scores.

3. **Two- to three-replicates may be sufficient for broad-scale assessments of ecosystem condition**, where the primary objective is to classify sites into eFIBI quality bands rather than precisely estimate community composition. Agreement in condition classifications was high for two or more replicates, reflecting the width of the NPSFM (MfE 2020) classification bands. However, this robustness also means that agreement in classification may mask underlying variability in ecological condition, particularly where sites lie near band thresholds.

4. **Full replication (six samples) should be retained when monitoring targets rare, threatened, or invasive species.** Reduced replication disproportionately affects the detection of low-abundance taxa, increasing the risk of false absences. For invasive species, where early detection is critical to management response, and

for threatened species, where confirming presence informs conservation prioritisation, this risk is unacceptable.

5. **Replication should be increased in larger or more complex stream systems.** The probability that reduced-replicate samples match the six-replicate reference decreases with stream characteristics, particularly stream order and channel width. In larger or higher-order systems, increased hydraulic complexity and dilution reduce sampling consistency, requiring greater replication to achieve comparable confidence.

   Figure 5-1 a decision-support framework for selecting replication effort across environmental gradients. Sites falling below the 0.5 probability contour (upper right of red line) of matching the six-replicate sample should preferentially increase -replicates, whereas reduced replication may be appropriate in smaller, lower-order systems where agreement is higher (lower left of red line; Figure 5-1).

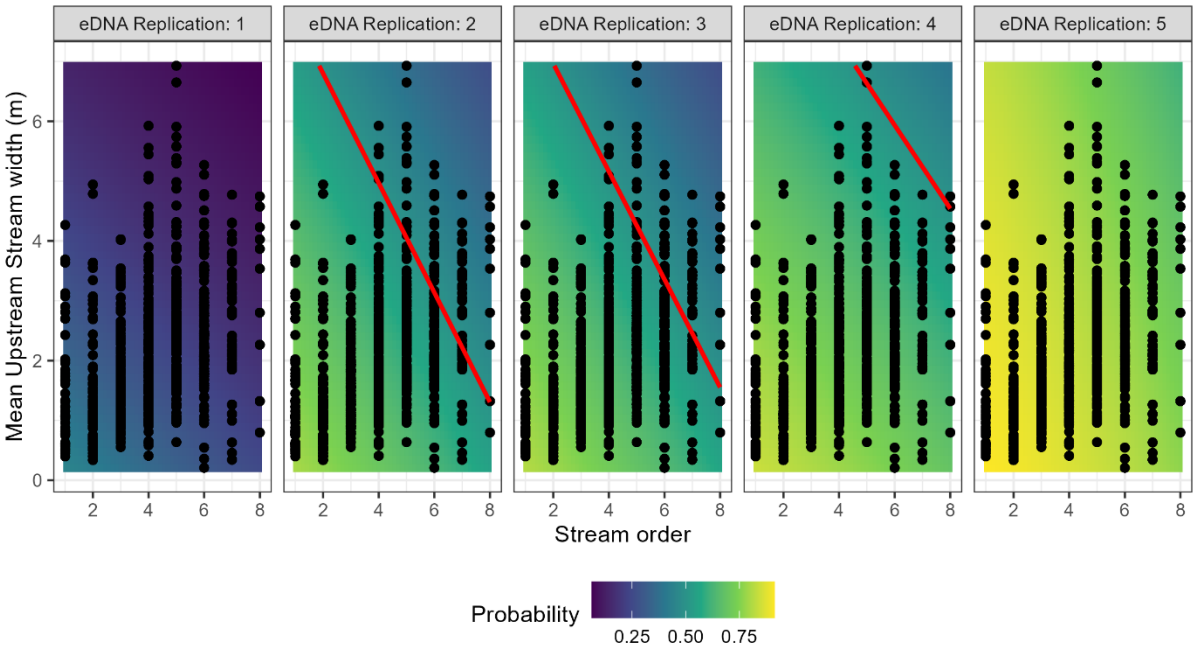

Figure 5-1: Predicted probability that an eDNA sample matches the six-replicate reference across gradients of stream order and mean upstream stream width, shown for replication levels from one to five-replicate samples. Colours represent model-predicted probabilities, with warmer colours (yellow) indicating increased probability of agreement. The red contour line denotes the 0.5 probability threshold. Points show observed mean upstream stream width versus stream order of six-replicate samples taken across New Zealand. Predictions are based on a fitted model with other covariates sample volume and channel slope held constant (continuous variables at their mean values; random effects excluded). The predicted probability for one-replicate samples does not exceed the probability threshold of 0.5. Whereas all prediction probabilities for a five-replicate sample exceed the probability threshold of 0.5.

6. **Results near eFIBI classification thresholds should be treated with caution and may require additional replication.** Reduced-replication sampling introduces uncertainty in eFIBI scores that may be sufficient to shift a site across a condition band boundary.

Figure 5-2 identifies the score ranges where this risk of crossing band boundaries is greatest for each replication level. When observed scores fall within these uncertainty zones (purple points), there is an increased risk that classification differs from the six-replicate reference. In such cases, it is recommended to increase replication or confirm results with additional sampling.

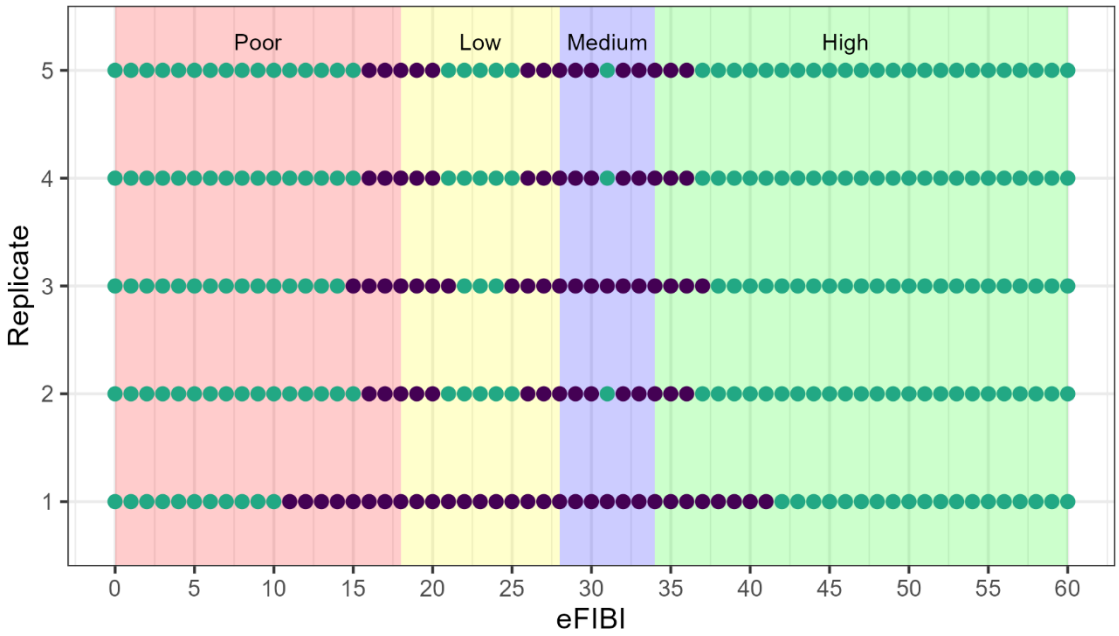

Figure 5-2: Uncertainty range of eFIBI scores based on modelled difference between the standard six-replicate eDNA sample and reduced one- to five-replicate eDNA samples. Purple points indicate values where the uncertainty overlaps eFIBI quality thresholds outlined in the NPSFM (2020). Each circle represents one eFIBI score between 0 and 60.

Overall, the results indicate that the suitability of reduced-replicate eDNA sampling depends on the intended application and acceptable level of uncertainty. While six-replicate sampling remains the preferred approach for regulatory monitoring and surveys targeting rare, threatened, or invasive species, reduced replication can provide a practical alternative for broader assessments of fish community condition. Of the reduced replication designs evaluated, two- to three-replicate sampling provided the greatest reduction in cost while maintaining relatively high agreement with the six-replicate benchmark. However, sites located near eFIBI quality thresholds or within larger, higher-order systems may require additional replication to ensure confidence in ecological condition assessments.

6 Estimating potential fish presence

To further provide context for end users into the availability of species at a sampling location, it is important to understand what species could potentially be present at the sampling site. This will provide necessary information into the fish community of a waterway by providing context into what species are expected to reside in the area. Below we discuss a range of tools that are available to assist in this process (Table 6-1).

Previous survey data are the most robust method to assess the species composition within a given stream reach. Here, the eDNA database from Wilderlab (wilderlab) and the New Zealand Freshwater Fish Database (NZFFD) curated by Earth Sciences New Zealand are the primary publicly available sources of information on species presence derived from eDNA detections and physical captures, respectively. However, both spatial and temporal considerations limit their potential usefulness. For a given site, previous samples within the stream reach (both physical captures and eDNA) should be considered the best estimate of potential species at a location. While the NZFFD remains a valuable and widely used resource for establishing species presence, physical capture methods are generally less efficient at detecting the full assemblage of species. This limitation arises from the relatively short stream lengths typically sampled (e.g., 50–150 m reaches) and differences in catchability among fishing methods and species (Crow et al. 2014), which can result in incomplete detection compared with eDNA-based approaches.

Beyond empirical records, predictive modelling frameworks can be used to estimate the potential presence of fish species at sites where survey data are limited or absent. In New Zealand, the most widely used resource for this purpose is the predictive fish distribution modelling by Crow et al. (2014), which are accessible through NZ River Maps, developed by the National Institute of Water and Atmospheric Research (NIWA, now Earth Sciences New Zealand). The modelling approaches used by Crow et al. (2014) relates known fish occurrence records from the NZFFD to environmental characteristics of river networks such as temperature, stream order, elevation, distance to the sea, and hydrological attributes.

Using these environmental relationships, models estimate the probability that a species could occur at unsampled locations across the river network. This allows predictions of potential species distributions even in areas where no direct surveys have been conducted. Such modelling frameworks are particularly useful in New Zealand because large portions of the river network remain unsampled or have limited survey coverage.

While predictive models are a valuable tool, it is important to recognise that they provide estimates of potential presence rather than confirmed occurrences. Model predictions are influenced by the quality and spatial coverage of the training data, as well as the environmental variables included in the modelling framework. As a result, predictions may overestimate or underestimate species occurrence in some locations, particularly in regions with limited historical sampling or in systems where local barriers or habitat conditions restrict fish movement.

A further limitation of predictive distribution modelling based on the NZFFD is that the underlying occurrence data originate from multiple agencies and are subject to biases in target species, differing survey methods, varying sampling periods, and substantial variability in sampling intensity and species detectability across habitats. Consequently, absence records within the database often reflect non-detection rather than true absence, which can introduce

uncertainty into model outputs and reduce confidence in predictions for rare, cryptic, or low-abundance species. These challenges are particularly relevant for contemporary management applications where fine-scale, current distribution information is required to assess ecological condition, identify biodiversity hotspots, or detect distributional changes associated with land use and climate pressures.

Environmental DNA approaches provide a promising pathway to address many of these limitations by enabling standardised and highly sensitive detection of aquatic species across broad spatial scales. Importantly, eDNA datasets are well suited to modern hierarchical modelling frameworks, including site occupancy models, which explicitly account for imperfect detection and sampling uncertainty. By separating the ecological process of species occurrence (i.e., where species are found) from the observation process (i.e., are we detecting the species), occupancy-based eDNA modelling can produce more robust and defensible estimates of species distributions than approaches relying solely on historical occurrence records. Integrating eDNA surveys with occupancy modelling therefore represents a substantial opportunity to improve the resolution, reliability, and scalability of freshwater biodiversity assessments in New Zealand, while also providing a strong foundation for long-term monitoring and adaptive management frameworks.

6.1 Recommended workflow

For a given sampling location, species assessments should first prioritise records from the same reach or stream as the focal river section (Point A top two figures Figure 6-1), as these data provide the strongest evidence of species presence at a site. Where no information exists for the focal reach or catchment, additional insight can be gained from surveys in adjacent river sections (Point B Figure 6-1). Predictive models are most effective when used alongside empirical datasets, such as eDNA detections or physical survey records. Together, these approaches provide a practical framework for estimating species presence: empirical observations confirm current or historical occurrence, while predictive models identify additional species that may plausibly occur based on habitat suitability and regional distribution patterns. Where available, consideration of anthropogenic pressures can further help explain species absences. For example, site B in Figure 6-1 is highly fragmented (i.e., the ability of a fish to move between any two habitat patches) and some species may be absent from this part of the catchment because barriers restrict upstream passage.

In practice, end users should interpret the available sources of species presence as an indication of species that *may* occur at a location and use them to guide sampling design, prioritise monitoring efforts, or provide ecological context for interpreting survey results. When empirical observations conflict with model predictions, direct survey data should generally be considered the more reliable indicator of species presence.

Table 6-1: Comparison of available methods to assess species presence in New Zealand Rivers.

Method	What it is	Pros	Cons	Best use
Wilderlab eDNA (metabarcoding)	Empirical detection of DNA from water samples processed by Wilderlab's laboratory and pipeline. Detects species presence from shed DNA. (wilderlab.com)	• Detects species non-invasively including rare/cryptic taxa. • Can detect taxa traditional methods miss. • Scalable / relatively fast turnaround. • Allows collection where physical surveys are difficult (deep/fast waters).	• Detection does not always equate to local residence (DNA transport/legacy signal). • Sensitive to lab/field contamination without rigorous QA/QC. • Semi-quantitative, difficult to infer abundance reliably. • Requires good reference databases for taxon matching (gaps can cause false negatives).	• Detect what *is actually shed* into water in your sampling window. • Good for presence/absence at fine spatial scale.
NZ Freshwater Fish Database (NZFFD)	National database of historical fish surveys (electrofishing, nets, documented records). (nzffdms.niwa.co.nz)	• Long-term, multi-method empirical record. • Includes metadata on sampling method, environment, sometimes abundance. • Broad spatial coverage accumulated over decades.	• Sampling effort, methods, and intensity are inconsistent across records. • Biases in where/when surveys occurred (detection gaps) and often the agency carrying out the survey. • Often presence-only or asymmetric effort, hard to infer true absence reliably.	• Good baseline for *historical occurrence* and large spatial patterns. • Useful for context, regional trends, and SDM training data.
Modelled presence (NZ River Maps / predictive models)	Spatially predicted fish presence/absence/habitat suitability across river network using NIWA/NIWA-linked models built from NZFFD and environmental covariates. (shiny.niwa.co.nz/nzrivermaps/)	• Fills gaps where empirical data are absent. • Consistent, standardised across the entire river network. • Useful for large-scale planning and hypothesis generation.	• Predictions are model outputs not confirmed occurrences. • Spatial/temporal resolution limited by model structure and input data quality. • Risk of over/under-predicting in regions with poor training data. • Known to predict poorly for certain taxa	• Good for regional planning, identifying *potential* habitat occurrence, and prioritising sampling.

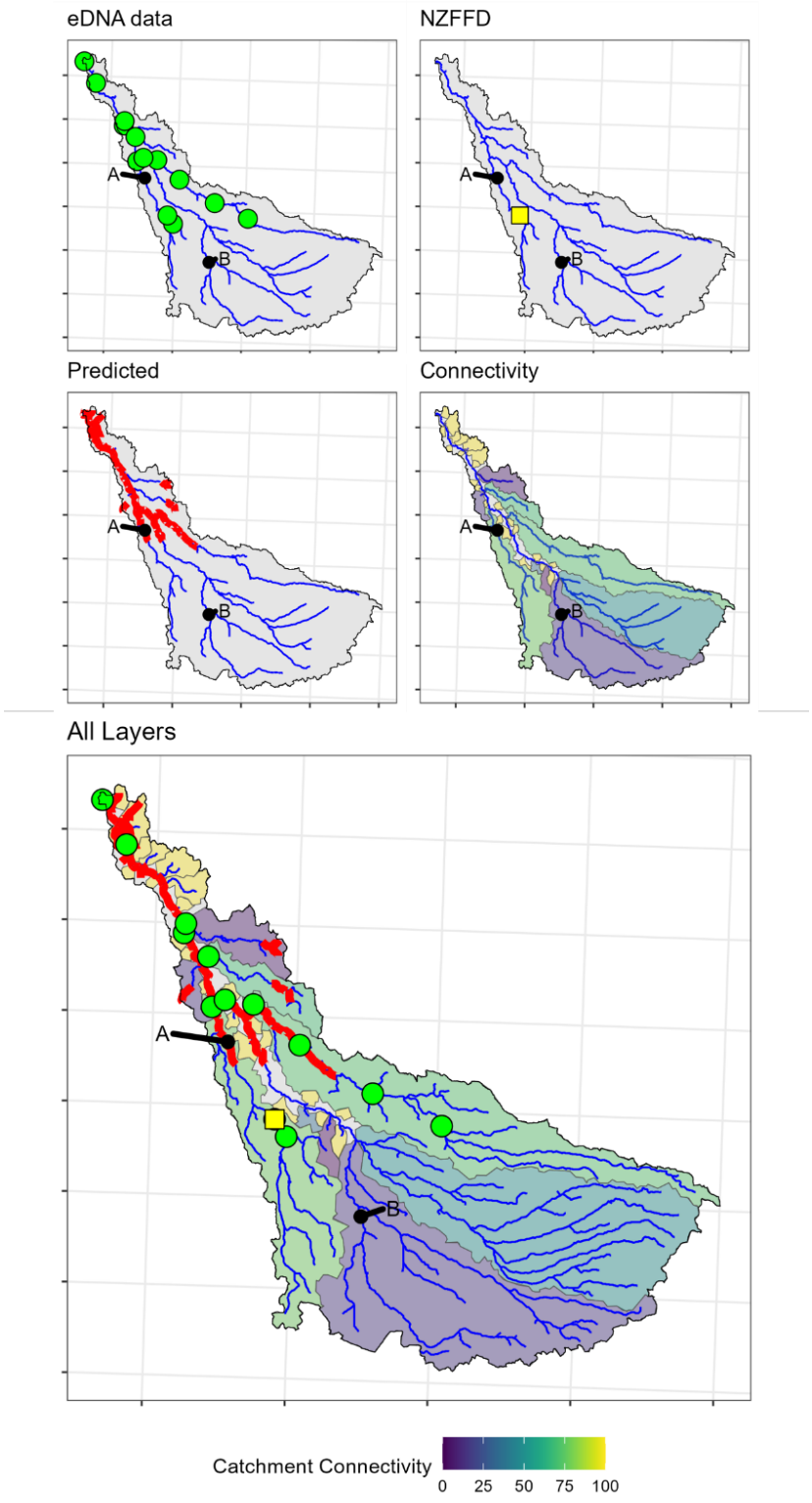

Figure 6-1: Example of Cran's bully (*Gobiomorphus basalis*) presence based on eDNA, catch records, modelled occupancy relative to stream connectivity. Points A and B are theoretical monitoring locations. Environmental DNA data for the figure was obtained from Dairy NZ, catch records were taken from the New Zealand Freshwater Fish Database (NZFFD), predicted data are from the New Zealand Rivers Map (Crow et al 2014), and river connectivity was obtained from the Barrier Assessment and Prioritisation Tool (BART).

7 References

Burian, A., Mauvisseau, Q., Bulling, M., Domisch, S., Qian, S., Sweet, M. (2021) Improving the reliability of eDNA data interpretation. *Molecular Ecology Resources*, 21(5): 1422–1433. https://doi.org/10.1111/1755-0998.13367

Burnham, K.P., Anderson, D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.

Crow, S., Booker, D., Sykes, J., Unwin, M., Shankar, U. (2014) Predicting distributions of New Zealand freshwater fishes. *Christchurch: National Institute of Water and Atmospheric Research Ltd*, 102.

Curtis, A.N., Harper, L.R., Davis, M.A., Larson, E.R. (2026) Applying Environmental DNA Metabarcoding to Calculate an Index of Biotic Integrity for Freshwater Fish. *Environmental DNA*, 8(2): e70251. https://doi.org/10.1002/edn3.70251

Fore, L.S., Karr, J.W., Conquest, L.L. (1994) Statistical properties of an index of biological integrity used to evaluate water resources. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(5): 1077–1087.

Ganasan, V., Hughes, R.M. (1998) Application of an index of biological integrity (IBI) to fish assemblages of the rivers Khan and Kshipra (Madhya Pradesh), India. *Freshwater Biology*, 40(2): 367–383.

Guillera-Arroita, G., Lahoz-Monfort, J.J., van Rooyen, A.R., Weeks, A.R., Tingley, R. (2017) Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, 8(9): 1081–1091. https://doi.org/10.1111/2041-210X.12743

Joy, M.K., Death, R.G. (2004) Application of the Index of Biotic Integrity Methodology to New Zealand Freshwater Fish Communities. *Environmental Management*, 34(3): 415–428. 10.1007/s00267-004-0083-0

Lacoursière-Roussel, A., Rosabal, M., Bernatchez, L. (2016) Estimating fish abundance and biomass from eDNA concentrations: variability among capture methods and environmental conditions. *Molecular Ecology Resources 16(6)*: 1401–1414. doi:10.1111/1755-0998.12522.

Mahlum, S., Baker, C., Lake, M., White, E., Williams, A., Williams, P., Franklin, P. (2025) The spatiotemporal movement ecology of shortfin eels and entrainment risk of yellow eels at a flood relief pump station with gravity bypass. *Ecological Solutions and Evidence*, 6(4): e70130. https://doi.org/10.1002/2688-8319.70130

Mahlum, S., Baker, C., Williams, P., Tieman, G., Melchior, M., Crawford, R. (2024) Integrating data sources of pouched lamprey detections to estimate their spatial distribution. 154th American Fisheries Society Annual Meeting, September 13-20, Hawaii 2024.

Mebane, C.A., Maret, T.R., Hughes, R.M. (2003) An index of biological integrity (IBI) for Pacific Northwest rivers. *Transactions of the American Fisheries Society*, 132(2): 239–261.

Melchior, M., Baker, C. (2023) Environmental DNA Guidelines and Field Protocols for Lotic Systems.

MfE (2020) National Policy Statement for Freshwater Management 2020. New Zealand Government.

Nakagawa, H., Fukushima, K., Sakai, M., Wu, L., Minamoto, T. (2022) Relationships between the eDNA concentration obtained from metabarcoding and stream fish abundance estimated by the removal method under field conditions. *Environmental DNA 4*: 1369–1380.

Pont, D., Hugueny, B., Rogers, C. (2007) Development of a fish-based index for the assessment of river health in Europe: the European Fish Index. *Fisheries Management and Ecology*, 14(6): 427–439.

Reid, A.J., Carlson, A.K., Creed, I.F., Eliason, E.J., Gell, P.A., Johnson, P.T.J., Kidd, K.A., MacCormack, T.J., Olden, J.D., Ormerod, S.J. (2019) Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biological reviews*, 94(3): 849–873.

Shu, L., Ludwig, A., Peng, Z. (2020) Standards for Methods Utilizing Environmental DNA for Detection of Fish Species. Genes 11, 296; doi:10.3390/genes11030296.

Smith, J., David, B., Hicks, A., Wilkinson, S., Ling, N., Fake, D., Suren, A., Gault, A. (2024) Optimizing eDNA Replication for Standardized Application in Lotic Systems in Aotearoa, New Zealand. *Environmental DNA*, 6(5). 10.1002/edn3.70017

White, R., Graham, E., Stoffels, R., Baker, C. (2025) *Environmental DNA indicators for supporting implementation of the National Objectives Framework (NOF) in the NPS-FM*. Earth Sciences New Zealand, Presentation to the Ministry of the Environment.

Appendix A Akaike Information Criterion Tables

Table A-1: Model selection for comparing the accuracy of one- through to five-replicate eDNA sampling to predict species richness relative to the standard six-replicate eDNA sample. Each model also included the following nested random intercept structure (1 | year/month) + (1 | region/catchment/site). LogLik is the log likelihood, K is the number of terms in the model, AIC is Akaike Information Criterion value, ΔAIC is the difference between the smallest (top model) and the corresponding model, and Weight is the weight of evidence in support of the model. A value of 1 indicates 100% support of the model where 0 indicates no support of the model given the data.

Model	LogLik	K	AIC	ΔAIC	Weight
Response ~ volume + replicate * stream order + replicate * slope + replicate * stream width	**-12411.254**	**22**	**24866.508**	**0**	**0.996**
Response ~ volume + replicate * stream order + replicate * Log(flow.us) + replicate * (Log(sediment) + I(Log(sediment)2))	-12412.638	27	24879.277	12.769	0.002
Response ~ volume + replicate * stream order + replicate * slope + replicate * Log(flow.us)	-12418.092	22	24880.184	13.676	0.001
Response ~ volume + replicate * slope + replicate * stream width	-12423.132	17	24880.264	13.756	0.001
Response ~ volume + replicate * stream order + replicate * Log(flow.us)	-12424.384	17	24882.768	16.26	0
Response ~ volume + replicate * stream order	-12430.779	12	24885.558	19.05	0
Response ~ volume + replicate * slope + replicate * Log(flow.us) + replicate * stream width	-12421.074	22	24886.148	19.64	0
Response ~ volume + replicate * stream width	-12434.278	12	24892.556	26.048	0
Response ~ volume + replicate * Log(flow.us) + replicate * stream width	-12430.381	17	24894.763	28.255	0
Response ~ volume + replicate * slope + replicate * Log(flow.us)	-12430.602	17	24895.204	28.696	0
Response ~ volume + replicate * Log(flow.us)	-12439.692	12	24903.385	36.877	0
Response ~ volume + replicate * slope	-12448.986	12	24921.971	55.464	0
Response ~ volume + replicate	-12467.064	7	24948.128	81.62	0

Table A-2: Model selection for comparing the accuracy of one- through to five-replicate eDNA sampling to predict the eFIBI relative to the standard six-replicate eDNA sample. Each model also included the following nested random intercept structure (1 | year/month) + (1 | region/catchment/site). LogLik is the log likelihood, K is the number of terms in the model, AIC is Akaike Information Criterion value, ΔAIC is the difference between the smallest (top model) and the corresponding model, and Weight is the weight of evidence in support of the model. A value of 1 indicates 100% support of the model where 0 indicates no support of the model given the data.

Model	LogLik	K	AIC	ΔAIC	Weight
Response ~ volume + replicate * stream order + replicate * slope.us + replicate * stream width.us	**-12420.151**	**22**	**24884.301**	**0**	**0.776**
Response ~ volume + replicate * stream order + replicate * slope.us + replicate * Log(flow.us)	-12421.803	22	24887.606	3.305	0.192
Response ~ volume + replicate * stream order + replicate * Log(flow.us)	-12427.891	17	24889.783	5.482	0.065
Response ~ volume + replicate * stream order	-12433.629	12	24891.259	6.958	0.031
Response ~ volume + replicate * (Log(sediment) + I(Log(sediment)2)) + replicate * stream width.us	-12427.337	22	24898.674	14.373	0.001
Response ~ volume + replicate * slope.us + replicate * Log(flow.us)	-12433.703	17	24901.407	17.106	0
Response ~ volume + replicate * slope.us + replicate * stream width.us	-12433.872	17	24901.745	17.444	0
Response ~ volume + replicate * Log(flow.us)	-12439.881	12	24903.763	19.462	0
Response ~ volume + replicate * slope.us + replicate * Log(flow.us) + replicate * stream width.us	-12430.507	22	24905.014	20.713	0
Response ~ volume + replicate * stream width.us	-12441.442	12	24906.884	22.583	0
Response ~ volume + replicate * Log(flow.us) + replicate * stream width.us	-12438.383	17	24910.767	26.466	0
Response ~ volume + replicate * slope.us	-12451.427	12	24926.853	42.552	0
Response ~ volume + replicate	-12457.643	7	24929.286	44.985	0

Table A-3: Model selection for comparing the accuracy of one- through to five-replicate eDNA sampling to predict eFIBI quality relative to the standard six-replicate eDNA sample. Each model also included the following nested random intercept structure (1 | year/month) + (1 | region/catchment/site). LogLik is the log likelihood, K is the number of terms in the model, AIC is Akaike Information Criterion value, ΔAIC is the difference between the smallest (top model) and the corresponding model, and Weight is the weight of evidence in support of the model. A value of 1 indicates 100% support of the model where 0 indicates no support of the model given the data.

Model	LogLik	K	AIC	ΔAIC	Weight
Response ~ volume + replicate * slope + replicate * (Log(sediment) + I(Log(sediment)2)) + replicate * stream width	**-8372.575**	**27**	**16799.149**	**0**	**1**
Response ~ volume + replicate * stream width	-8396.377	12	16816.755	17.606	0
Response ~ volume + replicate * slope + replicate * stream width	-8392.471	17	16818.943	19.793	0
Response ~ volume + replicate * flow.us	-8400.305	12	16824.611	25.462	0
Response ~ volume + replicate * flow.us + replicate * stream width	-8395.59	17	16825.18	26.03	0
Response ~ volume + replicate * slope + replicate * flow.us	-8395.888	17	16825.775	26.626	0
Response ~ volume + replicate * slope + replicate * flow.us + replicate * stream width	-8391.661	22	16827.323	28.174	0
Response ~ volume + replicate * stream order + replicate * flow.us	-8397.247	17	16828.493	29.344	0
Response ~ volume + replicate * stream order + replicate * slope + replicate * flow.us	-8393.784	22	16831.567	32.418	0
Response ~ volume + replicate * stream order	-8407.569	12	16839.139	39.989	0
Response ~ volume + replicate * stream order + replicate * slope	-8403.656	17	16841.312	42.163	0
Response ~ volume + replicate * slope	-8409.751	12	16843.503	44.353	0
Response ~ volume + replicate	-8416.337	7	16846.674	47.525	0