BENTHIC MACROINVERTEBRATE MONITORING PLAN FOR LARGE TRANSBOUNDARY RIVERS IN THE ALBERTA-NWT REGION

ASSESSMENT OF RESULTS FROM THE FIFTH YEAR OF SAMPLING

2021

Alberta Government

Canadian Rivers Institute

UNB

Government of Northwest Territories / Gouvernement des Territoires du Nord-Ouest

Prepared by:

Jennifer Lento, MSc, PhD
Research Scientist
Canadian Rivers Institute and Department of Biology
University of New Brunswick
Fredericton, NB, Canada

Prepared for:

Alberta-Northwest Territories Bilateral Management Committee
Department of Environment and Parks, Government of Alberta, and
Department of Environment and Natural Resources, Government of the Northwest Territories

Revised March 2023

Executive Summary

This report provides an assessment of the first five years of sampling data in the Slave River from the Government of the Northwest Territories and the Government of Alberta benthic macroinvertebrate monitoring plan for large transboundary rivers. The goal of this sampling program is to collect baseline data to support the development of normal range criteria that can be used for the detection of any potential future impacts to the river system. Data were collected from the Slave River from 2017-2021, although sampling was limited in 2020 due to high water levels. Sampling of benthic macroinvertebrate (BMI) assemblages and supporting environmental variables took place in shoreline sites grouped within seven reaches in the Slave River. For a second year, water levels were too high to sample the Hay River, and it was not included in this report.

2021 was an important year for sampling in the Slave River because it followed the high water levels of 2020 and offered the opportunity to assess recovery from extreme flow conditions. Water chemistry and sediment chemistry results were similar to previous sampling years, and did not indicate any unusual or highly elevated results. BMI assemblage composition in 2021 appeared similar to that observed prior to 2020. Following the dominance of the cnidarian *Hydra* during high flow conditions in 2020, 2021 marked a return to more typical assemblages in Slave River reaches, though abundances of Ephemeroptera, Plecoptera, and Trichoptera (EPT; mayflies, stonesflies, and caddisflies) were elevated in most reaches compared to previous years. Total abundance and abundances of Chironomidae (midges) and *Hydra* were similar to those observed prior to 2020. Total taxonomic richness showed a decline from 2017-2021 in most reaches due to declining richness of EPT, Diptera (true flies, including midges), and Oligochaeta (segmented worms).

The development of normal range estimates and critical effect size (CES) boundaries for biotic metrics has been an ongoing process, with the addition of new data allowing for the refinement of these preliminary estimates. With five years of data for the Slave River, it was possible to critically evaluate what constitutes natural variability in metrics and what constitutes noise or a response to extreme conditions (for example, the high flows in 2020). Normal range estimates were refined with the exclusion of particular years that appeared to inflate temporal variability estimates. For example, total abundance and *Hydra* abundance metrics were tested with the exclusion of data from 2020, which resulted in more precise preliminary normal range estimates. For Chironomidae abundance metrics, the exclusion of data from 2017 (when there was high abundance and diversity of Chironomidae relative to 2018-2021) reduced variability and generated more narrow CES boundaries. Two sets of normal range boundaries were tested for EPT abundance, as most reaches showed low abundance from 2017-2019 and high abundance from 2020-2021; however, this approach was exploratory and should be reassessed with the collection of more data. Genus richness of EPT had a narrow normal range with all years of data included, and was identified as having strong potential diagnostic power. While the normal range for total richness was wider, it should continue to be monitored to determine whether the loss of taxonomic richness in the river is part of a longer-term trend.

Temporal assessment of full assemblage structure was completed using multivariate analysis. An ordination of data from 2017-2021 with 95% probability ellipses highlighted the loss of diversity from 2017 to 2021. However, the spatial arrangement of sites in multivariate space (based on assemblage structure) indicated that the most similar years were 2017, 2018, and 2021. Furthermore, change trajectories in multivariate space indicated the strong similarity of 2017 and 2021. Pairwise comparisons of ordinations among years through Procrustes analysis were used to refine normal range estimates developed from site residuals. The normal range for site Procrustes residuals was narrow, and most sites were within or below the CES boundaries, which indicated that this approach might be useful to detect

when one or more sites changes to an unusual degree relative to other sites (and thus has a high Procrustes residual) in future years.

Generalized Procrustes Analysis (GPA) was explored as a new method to capture temporal variability among sites based on assemblage data. The test develops a consensus ordination that is the average of multiple ordinations. By applying the test to temporal data, the consensus ordination summarizes temporal variability among BMI ordinations, and acts as a reference point, capturing the variability in assemblage structure across sample years. An ordination of new data can then be compared with the consensus ordination to identify any strong differences in the spatial arrangement of sites. The approach was tested by developing a consensus ordination using data from 2017-2019 and comparing it with an ordination from 2021, and the results found statistically significant similarity between the two ordinations, consistent with what was expected based on earlier analyses. A consensus ordination combining data from 2017-2019 and 2021 was then developed to act as a reference point to be compared with new data in future sampling years. Though not a direct test of assemblage structure, preliminary analysis suggests that it may be an effective technique to detect assemblage-level changes that lead to a shift in one or more sites relative to the others. As more data are added to the consensus ordination, it should become a more accurate and precise reflection of typical relationships among sites and reaches in the Slave River.

Table of Contents

1. Introduction

The Government of the Northwest Territories (GNWT) and the Government of Alberta (GOA) are working to establish a monitoring program for the bioassessment of large transboundary rivers (MacDonald Environmental Sciences Ltd. 1995, Lento 2017). Transboundary rivers provide unique challenges to assessment, as monitoring designs must meet the objectives of multiple jurisdictions that may differ with respect to economic and social goals as well as environmental management strategies (MacDonald Environmental Sciences Ltd. 1995). However, the potential for upstream development within one jurisdiction to cause downstream impacts within another jurisdiction emphasizes the need for cooperation in the monitoring of transboundary waters to ensure the detection of changes to ecosystem health (Flotemersch et al. 2011). Establishment of long-term monitoring and assessment supports the future detection of impacts that may arise from human development, but also supports the detection of ecological changes in response to a warming climate.

1.1. General Approach of the Monitoring Program

Monitoring questions related to assessing ecosystem health may be focused on comparison of reference sites with test sites in the presence of a known stressor, or they may be focused on characterizing the contemporary status of biotic and abiotic ecosystem components and evaluating whether any temporal changes have occurred (e.g., Environment Canada 2011, Culp et al. 2012b). One approach used in biological monitoring, particularly in the case of detecting future evidence of impairment, is to estimate the normal range of community composition based on natural variability in the system, and to detect any shifts in the diversity or abundance of organisms that occur over time (Munkittrick et al. 2009, Munkittrick and Arciszewski 2017). Where there is not a clear stressor in place, determining the range of "normal" variation in the data can be used to establish a baseline ecological condition, providing information that can be used in future years (with continued monitoring) to begin to address targeted questions as stressors increase (Munkittrick et al. 2009, Munkittrick and Arciszewski 2017). Quantification of variation that might be expected in the absence of impairment can support the development of "trigger" levels, or levels at which the magnitude of observed change is greater than expected, necessitating additional monitoring or management action (Arciszewski and Munkittrick 2015). Future assessments could focus on examining relationships of natural and anthropogenic drivers of change with ecosystem health, and detecting evidence of cumulative impacts (e.g., from a combination of climate change, development, resource exploration, or other stressors; Dubé 2003, Dubé et al. 2013, Somers et al. 2018). Establishing a strong baseline for comparison is a vital step in this process to allow for future detection of ecosystem responses to change (Culp et al. 2012b).

Part of the initial focus of the GNWT and GOA transboundary monitoring program is on benthic macroinvertebrate (BMI) assemblages, which are an important ecosystem component to monitor in northern rivers as an integrated measure of water quality and habitat condition (Culp et al. 2012b, Buss et al. 2015, Lento et al. 2022b). BMIs are commonly chosen for biomonitoring because they are widespread, easy to sample and identify, species-rich, have limited mobility, and have known tolerances and sensitivities to habitat conditions that can support the detection of anthropogenic impacts (Bonada et al. 2006, Resh 2008, Buss et al. 2015). Because they have generally low mobility, BMI respond to local-scale changes in water chemistry and habitat quality and are an excellent indicator of the physical and chemical impacts of disturbance. BMI provide a more time-integrated measure of change than spot measurements of water chemistry, which only describe conditions at the time of sampling. Moreover, BMI diversity at northern latitudes is strongly linked with temperature as a result of taxon-specific thermal tolerances (Culp et al. 2019, Lento et al. 2022b, Lento et al. 2022a). With climate change, it is predicted that biodiversity in northern regions will begin to more closely resemble those of lower-

latitude temperate systems through the northward movement of eurythermic and cold-intolerant species (Culp et al. 2012a, Lento et al. 2019). Thus, the long-term assessment of BMI assemblages has the potential to detect changes in response to a warming climate in addition to detecting future impacts from human development.

Within the Alberta-Northwest Territories transboundary river regions, there is relatively little information about the current state and composition of benthic assemblages. Assessments of BMI assemblages in the large transboundary rivers of the Alberta-Northwest Territories region have been limited (but see Paterson et al. 1991, Paterson et al. 1992 for baseline assessments of Slave River BMIs, and , Golder Associates 2010 for an overview of existing assessments), and (Dagg 2016) noted that this lack of background knowledge has made it difficult to identify water quality concerns and potential for impairment during local community discussions of potentially vulnerable ecosystem components. Therefore, it is vital that routine monitoring of large transboundary rivers be established to secure information about baseline conditions in these assemblages and to provide sufficient information to allow for future detection of trends.

1.2. Establishing Normal Ranges

In biomonitoring, the concept of the normal range is based on the idea that it is not always possible to access data from before any perturbation occurred in a region (Arciszewski and Munkittrick 2015), nor is it necessarily desirable to use such historical data if they do not accurately represent attainable water quality levels (Stoddard et al. 2006, Munkittrick and Arciszewski 2017). Instead, if sufficient contemporary data are collected to allow estimation of the range of variability that is acceptable given current conditions in a system (i.e., the limit of how variable a sample can be before it is considered to be different from expected), then this information can be used to detect any future deviations and pinpoint where targeted sampling should take place to identify causes of impacts (Kilgour et al. 2017, Munkittrick and Arciszewski 2017).

The normal range quantifies the range of variability in a community metric that is expected and acceptable for a system, given its current conditions. Values falling outside that range indicate that more monitoring is required or that management action must be taken. Quantifying the normal range for a system requires characterization of spatial variability within the system, but the ultimate goal is to describe temporal variability, to determine whether changes in metric values in subsequent monitoring years fall outside the range of acceptable variability for a site. Repeated sampling at the same location across multiple years allows for the characterization of a site-specific normal range of variation. Initially, temporal normal range estimates for a site will be imprecise as they encompass short-term, inter-annual variability in the systems. But as more years of data are collected for a site, the estimated temporal normal range of variation will become more precise and allow for the detection of potentially subtle changes happening over a longer time scale (e.g., 10+ years; Arciszewski and Munkittrick 2015).

Baseline data must be collected for multiple reference sites over multiple years, with sampling taking place in a single season (e.g., fall), and subsequent monitoring activities must continue at multiple sites for many years to allow for effective detection of change (Arciszewski and Munkittrick 2015). In the first two years of collecting baseline data, spatial variability among sites is described, and in subsequent years the natural temporal variability is quantified, and measures of temporal and spatial variability are refined. At least three years of baseline data must be collected before temporal variability can begin to be estimated, including the characterization of the regional normal range (as only two years of data may represent two different extremes). However, measurements based on three years of data are only initial estimates, and additional sampling beyond three years is recommended to achieve greater accuracy and precision in estimates of temporal variability and to detect any shifts in normal range due to climate

change (Arciszewski and Munkittrick 2015). In their analysis of long-term fish monitoring data from the Moose River, Arciszewski and Munkittrick (2015) noted that the precision of their estimates of variability improved as additional years of data were added, and they recommended a minimum of 12 years of data to capture the variability in the system, though the number of years required will vary among systems and may differ among target organism groups. In their global review of long-term freshwater monitoring studies, Jackson and Füreder (2006) suggested that five years of monitoring was the minimum number required to capture the range of ecological variability in BMI assemblages in response to short-term climatic cycles, but noted that at least 10 years of monitoring was required to capture the response to longer decadal cycles. Long-term data (> 10 years) for freshwater BMI in the Arctic are rare (Lento et al. 2019), but a recent study by Milner et al. (2023) examined changes in diversity of Alaskan stream BMI over a span of 22 years, and identified high inter-annual variability in diversity that was related to short-term climatic cycles and longer-term trends in diversity that appeared to relate to long-term climatic shifts in temperature and precipitation.

1.3. Quantifying Meaningful Change: Critical Effect Sizes

The concept of the normal range applies well to the situation where a monitoring program is being established in anticipation of potential future impacts, because it allows for quantification of the current status in the system as well as the level of change that would be deemed significant enough to warrant concern, termed the critical effect size (CES; Munkittrick et al. 2009, Arciszewski et al. 2017, Munkittrick and Arciszewski 2017). The CES is the magnitude of difference between sites or change across time (within a site) that is considered to be meaningful and to have ecological implications (Munkittrick et al. 2009). The CES forms the lower and upper boundaries of the normal range, indicating values below and above which there is meaningful change among sites or over time. It can act as a trigger point in adaptive monitoring plans to identify when additional sampling is necessary to investigate potential drivers of change (Somers et al. 2018).

In ongoing monitoring, the CES identifies the magnitude of change that is required before management action is taken, but in the development of monitoring programs, CES can also be used to ensure sampling designs are sufficient to detect impairment (Munkittrick et al. 2009). For example, as the normal range of variability across systems is quantified in pilot sampling years, the CES (values at the upper and lower limits or boundaries of the normal range) can be determined and used in power analysis to estimate the number of samples that would be required to detect a meaningful difference among sites. Initial establishment of variation among all sites in a river, as a measure of spatial variability, can be done with pilot-year monitoring data, but as more data are collected, it is important to refine the spatial CES to account for short-term temporal variability that is likely to be observed within systems (Arciszewski and Munkittrick 2015). Once at least three years of data have been collected (the minimum required to calculate CES), the CES can begin to be refined to capture site-specific temporal variability and quantify confidence intervals that can be used in future years to detect deviations from normal range. Exceedance of the CES by any site in future years would then act as a trigger to increase sampling efforts and determine if impairment has occurred.

A number of different approaches have been used to determine CES for different groups of organisms (see review in Munkittrick et al. 2009); however, studies of BMI assemblages that assess natural variability within and among sites have generally relied on standard deviation units or similar approaches (e.g., confidence intervals or probability ellipses) to set CES. For example, the CES for invertebrate abundance might be set to 2 SDs above and below the mean abundance observed in baseline data. In a normal distribution of data, a distance of 2 SDs from the mean encompasses 95% of the data, and any values that fall outside that range have a high probability of representing a different

population of data (e.g., an assemblage in an impaired or otherwise altered state). Such an approach can be easily applied to the calculation of normal range and CES for biotic metrics (summary indices of abundance and diversity), allowing the comparison of metric values with a range of expected values.

1.4. Temporal variability at the assemblage scale

Assessment of biotic metrics can provide meaningful and comprehensive summaries of community structure; however, the use of multivariate techniques can provide complementary information about compositional patterns and biotic interactions that cannot be captured by univariate assessments alone (Reynoldson et al. 1997, Bowman and Somers 2006). Multivariate analyses consider the presence or abundance of all taxa simultaneously (rather than individual groups of taxa), and use this information to identify differences in community composition among samples. Comparison of multivariate ordinations of samples between years could provide a measure of the change in community composition at a site relative to other sites from one year to the next. However, there is little work that has been done to establish multivariate measures of normal range and CES across temporal data. Multivariate techniques are used in the national CABIN program and in national programs outside Canada for comparison of test sites with reference sites, using probability ellipses to identify samples that fall outside of the multivariate normal range for reference sites (Bailey et al. 2004). Such approaches are generally built on assessing spatial datasets, with a large set of reference sites compared with test sites after grouping them based on environmental conditions (e.g., geology, climate). Extending multivariate approaches to consider temporal variability in a single river, where many sites and reaches are repeatedly re-sampled, does not easily fit with existing reference condition approach models, where reference sites are expected to be from different rivers, covering a wide range of habitat conditions. In addition, assessing temporal change in the full assemblage requires consideration of the non-independence of samples across time, to ensure that temporal data are compared within locations over time to detect changes. These challenges must be considered in the development of multivariate approaches to define normal range and CES.

Procrustes analysis, which is based on the concept of assessing the degree of similarity among spatial arrangements of points, provides an opportunity to examine and quantify temporal change in BMI assemblages as a whole through comparison of multivariate ordinations. Procrustes analysis can be used to determine whether the spatial position of sample points in multivariate space (based on BMI assemblage structure) is more similar between two ordinations than could occur by chance (Jackson 1995). The use of pair-wise Procrustes analyses to compare ordinations among pairs of years allows for the estimation of temporal change trajectories, by determining which years were most similar and which were most different (e.g., Lento et al. 2008). As an extension of this concept, Generalized Procrustes Analysis (Gower 1975) can be used to create a “consensus” ordination that represents the spatial positioning of points averaged across multiple ordinations. Though this test has most commonly been used in applications such as the social sciences (combining survey results to find a consensus), food sciences (combining judging scores on different food quality categories), and shape analysis (comparing shapes of objects using reference points), there is the potential to explore its use to develop a multivariate reference ordination (a consensus ordination that is the average of several years, thus incorporating temporal variability) with which ordinations in future years can be compared. Such an approach would estimate the deviation of future samples based on their relative assemblage composition, with changes in the spatial positioning of sites relative to the consensus ordination reflecting temporal differences in the similarity of sites.

1.5. Purpose and Objectives

The purpose of this report series is to assess spatial and temporal variability within the Hay and Slave rivers based on data collected as part of the GNWT and GOA large transboundary river BMI monitoring program. However, in 2021, it was not possible to sample the Hay River due to high water levels (see section 3.1.1 for details). Therefore, this report includes an assessment of spatial patterns in only the Slave River sampling data from September 2021 and temporal patterns in Slave River data from 2017 to 2021. Water chemistry, sediment chemistry, physical habitat, and BMI kick samples were collected using the methods described by Lento (2018), and data were analyzed to characterize spatial and temporal variability within the river, including quantification of CES for a number of biotic metrics. In addition, this report is focused on further developing measures of normal range and temporal variability for the Slave River based on the full assemblage. The assessment of temporal variability using multivariate methods is expanded to include an exploration of the use of consensus ordinations to summarize baseline patterns and develop a multivariate-based reference point for future sampling.

As this report summarizes five years of sampling in the Slave River, particular emphasis is placed on assessing the quality of normal range estimates while recognizing that conditions in some years may have contributed to a great deal of variability in these estimates. Normal range estimates developed with subsets of years are explored to provide greater diagnostic power.

2. Methods

2.1. Study Area and Sample Timing

The pilot program of the GNWT and GOA large transboundary river monitoring program is focused on the Slave River and the Hay River. Both rivers originate in Alberta flowing north into the Northwest Territories and terminating in Great Slave Lake (Figure 1), but they differ with respect to size, flow, and upstream land use (see overview in Golder Associates 2010). The Slave River is a large, fast-flowing river, with a mean annual discharge rate of 3,400 m^3/s (Sanderson et al. 2012) and a drainage basin of over 616,000 km^2 (Golder Associates 2010). The Hay River is narrower, more shallow, and slower-flowing, with a drainage basin of 48,100 km^2 (Golder Associates 2010), though water levels in recent years have been exceedingly high in this river. Details on the geology, climate, land cover, and land use history of both river catchments can be found in state of knowledge reports for the Hay River (Stantec Consulting Ltd. 2016) and Slave River (Pembina Institute 2016). Both rivers have the potential to be impacted by a

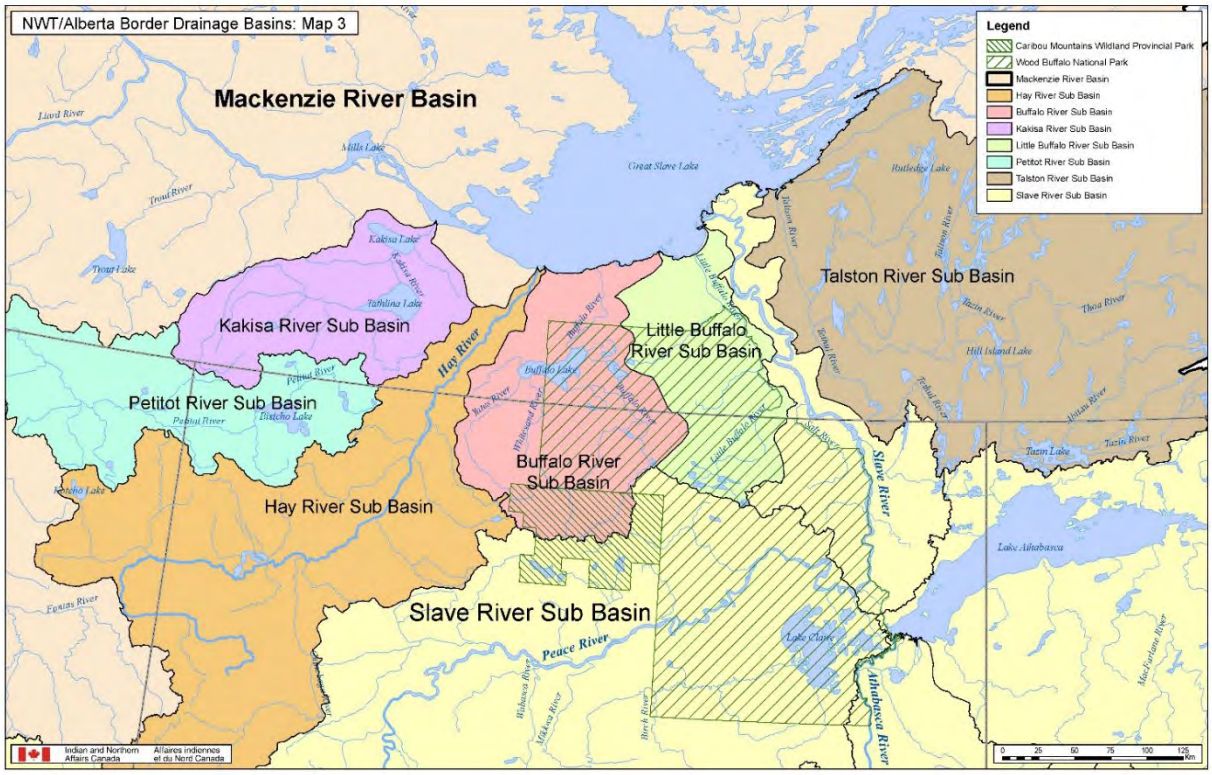

Figure 1. Drainage basins at the NWT/Alberta border, including the Hay River Sub Basin and Slave River Sub Basin. Map created by Indian and Northern Affairs Canada.

variety of human activities in the upstream basin, including oil and gas development and pulp and paper mills. Though change may have already occurred in these systems due to upstream activities, lack of historical baseline data precludes the assessment of such changes. The current program is aimed at characterizing the current ecological condition of these rivers as a baseline for future assessments.

The differences between these rivers with respect to size, depth, and flow initially required logistical considerations when planning and conducting BMI sampling. Sampling is designed to occur in the fall in part to take advantage of increased access to the shoreline that is gained when water levels recede, but the exact timing for sampling of each river was chosen to maximize accessibility for kick sampling. Low flows in the Hay River in 2017 and 2018 required earlier sampling and the use of a low-profile boat to maneuver through sand bars in some areas, but high water levels in 2019, 2020, and 2021 made sampling difficult or impossible in this river. Sampling was possible in the Slave River in all five years (though site access was limited in 2020), but additional safety equipment (e.g., belay and dry suits) was required to safely sample the deep, fast-flowing river. In 2021, it was not possible to sample the Hay River, and sampling took place in the Slave River from September 8-10 (see Table 9 in Appendices).

2.1. Site selection

The BMI monitoring plan for large transboundary rivers (described briefly here, but see Lento 2018 for details) prescribes a sampling design with 5-10 approximately 500-m-long reaches sampled in a river. The number of reaches depends on how variable the reaches are, and how many would be required to

Figure 2. Map of Hay River and Slave River, showing kick-sampling reaches (red points) selected within the rivers, and an overlay of the stream network. In 2021, no sampling took place in the Hay Rive due to high water levels. Stream network layer from National Hydro Network (NHN) GeoBase Series (open.canada.ca).

Table 1. Approximate coordinates in decimal degrees (DD) for each kick-sampling reach in the Hay River and Slave River, with indication of the site numbers (1-5) at which water chemistry and BMI samples were collected in 2021. Only the Slave River was sampled in 2021, because water levels were too high in the Hay River. High water levels in the Slave River did not allow access to all sites/reaches. Reach codes are explained in text.

River	Reach	Latitude (DD)	Longitude (DD)	Chemistry sites sampled in 2021	BMI sites sampled in 2021
Hay River	HR-KS1	59.9321	-116.9524	None	None
	HR-KS2	59.9465	-116.9565	None	None
	HR-KS3	59.9885	-116.9304	None	None
	HR-KS4	60.0026	-116.9713	None	None
	HR-KS5	60.0113	-116.9218	None	None
	HR-KS6	60.0279	-116.9216	None	None
Slave River	SR-KS1	59.4085	-111.4620	3	1,2,3,4,5
	SR-KS2	59.4276	-111.4629	3	1,2,3,4,5
	SR-KS3	59.5350	-111.4577	3	1,2,3,4,5
	SR-KS4A	59.5912	-111.4195	3	3,4,5
	SR-KS4B	59.5903	-111.4225	3	1,2,3,4,5
	SR-KS6	59.6766	-111.4856	3	1,2,3,4,5
	SR-KS5	59.7182	-111.5058	3	1,2,3,4,5

characterize the river and achieve adequate power to detect biologically-meaningful differences among reaches, if they were to exist (with this number refined through the assessment of baseline monitoring data). Reaches are selected to have similar substrate composition throughout the reach. The goal is to

select reaches with rocky substrate, as these will have the most diverse BMI assemblages, though soft sediments are deemed acceptable if comparable substrates can be sampled in additional reaches (see Lento 2017, 2018 for more details). Within each reach, five replicate kick-sites are sampled, approximately 50-125 m apart. If access to both banks of the river is possible, a total of 10 kick-sites is sampled within a reach (five on each river bank). This design allows for the application of multiple statistical analyses to characterize variability within a river. For example, sites can be compared directly along a longitudinal gradient, or sites can be treated as replicates in a statistical comparison of reaches. This design was applied during the first five years of sampling, though some adjustments were made to reflect local conditions.

Both rivers are accessed via boat launches on the Alberta side of the border (Figure 2). Five kick-sampling reaches were chosen within each river for the pilot year of sampling, and this number was increased to six in the Hay River in 2018 and to six in the Slave River in 2019 (Table 1; Figure 2). Sample reaches were selected to be approximately 500 m in length, though in some areas, the availability of suitable habitat limited the total length of reaches (e.g., in the Hay River, reaches were 250 m to 500 m in length, whereas in the Slave River, reaches were 250 m to 600 m in length). Sample reaches were numbered KS1 to KS5 or KS6 in each river, with KS1 representing the farthest upstream sampling location and KS5 or KS6 representing the farthest downstream sampling location (Figure 3). In the Slave River, the name for reach KS6 was assigned because it was added two years after the other reaches were chosen (KS1 to KS5), but it is located upstream of KS5 (Figure 3B). Reach 4 of the Slave River was the only location where sampling took place on both banks of the river, resulting in two sets of sites (SR-KS4A and SR-KS4B) in the same reach (Table 1). In the Hay River, reaches were 2.5 to 6.7 km apart, whereas in the larger Slave River, reaches were 1.9 to 11.8 km apart.

The Hay River is sinuous with slow flow in typical years. In the pilot year of sampling, reaches with rocky habitat were generally found at the bends of the river, typically on the erosional banks (Figure 3A). The depositional bank was generally a thick silty/muddy substrate that would not have allowed for access or for sampling (unlike sandy habitats, in which kick sampling can be conducted). Because of the shallow nature of some extents of the river, site selection was limited in some areas to reaches that could be accessed from the boat launch in a timely manner using a canoe with outboard motor. Analysis of reaches sampled in 2017 indicated that there were some differences between reaches upstream (HR-KS1 to HR-KS3) and downstream (HR-KS4 and HR-KS5) of the boat launch and inflow from tributaries, and a recommendation was made to sample an additional reach downstream of the boat launch to ensure adequate replication in the downstream portion of the river. Reach HR-KS6 was added in 2018 in response to this recommendation (Table 1; Figure 3A), and it was found to resemble the two other downstream reaches (Lento 2020).

The Slave River is wider than the Hay River with a straighter channel and faster flow (Figure 3B). Rocky substrates were generally found in areas of rocky outcrops along the shoreline. In the analysis of data from 2017 and 2018, substrate and flow appeared to play a large role in determining the BMI assemblage that was characteristic of a particular reach, and a recommendation was made to add another reach with rocky habitat and fast flow. In 2019, Reach SR-KS6 was added upstream of reach SR-KS5 (Figure 3B), and it was found to be a suitable addition to the sampling program (Lento 2021).

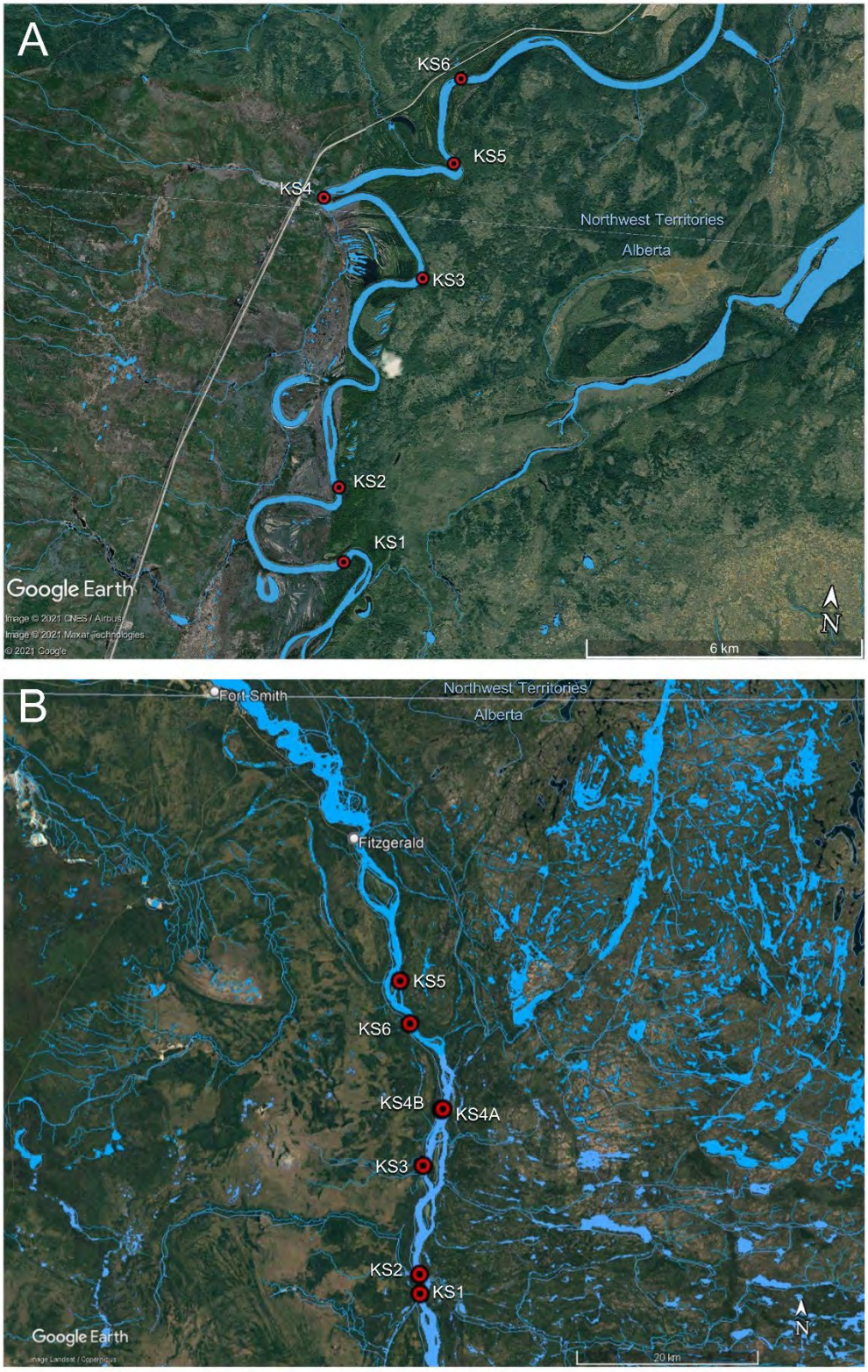

Figure 3. Kick-sample reaches (red points) in the (A) Hay River and (B) Slave River. Reaches are labeled in white text. No sampling was possible in the Hay River in 2021, but all reaches in the Slave River were sampled. Water body and stream layers overlain on maps are from the National Hydro Network (NHN) GeoBase Series (open.canada.ca).

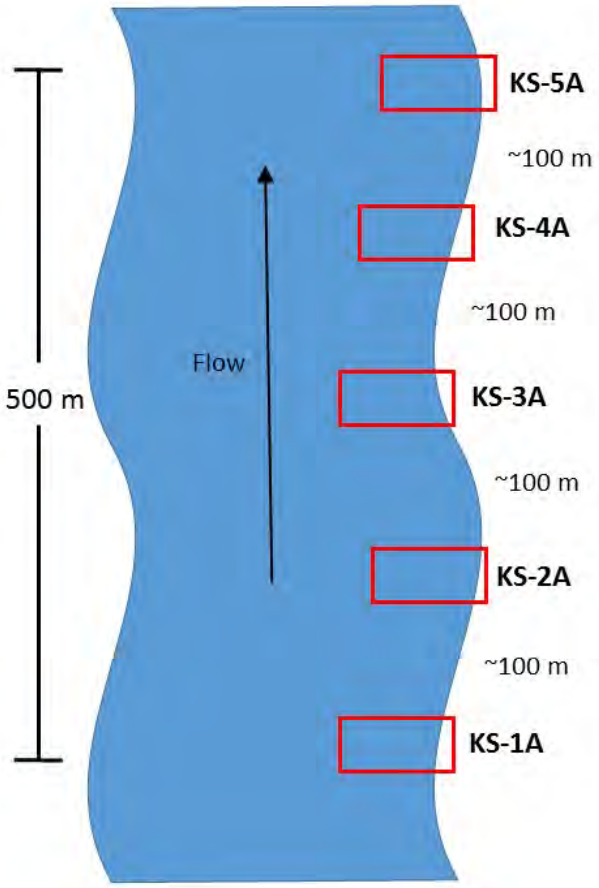

Figure 4. Example sampling design used for a single reach within the Hay River and Slave River, indicating the location of 5 sites within the 500 m reach and numbering of sites with respect to flow direction. Sampling of sites began downstream, at site KS-5A and worked upstream towards site KS-1A. Sites located on the opposite bank (left bank, when facing downstream) were numbered KS-1B through KS-5B. Sites were located approximately 100 m apart (50 m to 125 m) and sampling extended out into the river to a depth of approximately 1 m (maximum safe depth for kick sampling).

Sampling takes place in each reach on the bank where rocky habitat is located (e.g., see Figure 4 for an example of single-bank sampling design). Kick-sites within a reach are numbered 1-5, with site 1 as the farthest upstream site and site 5 as the farthest downstream site (consistent with the numbering of reaches); however, sampling is done at kick-site 5 first to avoid downstream contamination of samples. The right-hand bank while facing downstream (river right) is called the A bank and the left-hand bank (river left) is the B bank, and each site code is appended with A or B to indicate which side of the river was sampled. Reach KS4 in the Slave River is the only location (for either river) where sampling is feasible on both banks, and samples are collected from both the A and B banks in this reach to compare habitat conditions and BMI composition. Kick-sites were evenly spaced within reaches, when habitat availability allowed. Distance between chosen kick-sites was generally 50-125 m, as allowed by reach length. Kick-sites within each reach were generally of similar substrate composition, and were chosen to minimize differences in substrate composition. Based on data from 2017 and 2018, recommendations were made to shift some sites that appeared to be too silty (e.g., SR-KS2-1A, SR-KS4-1A, and SR-KS4-2A). These reaches were shifted to rockier habitat in 2019 to ensure data were more comparable with other reaches.

2.2. Sample Collection

Sample collection at kick-sampling locations followed the methods prescribed in the monitoring plan (Lento 2018), including collection of water chemistry samples, use of handheld meters for field chemistry, a habitat survey (modified from the Canadian Aquatic Biomonitoring Network - CABIN), a modified three-minute CABIN kick sample, and a modified rock walk (see details in Lento 2018). An overview of the full sampling scheme is provided here, with notes of modifications in 2021.

At kick-site 3 in each reach, water samples were collected for analysis of a standard suite of parameters, including nutrients, ions, and suspended solids. Water chemistry samples were reduced to a single site in each reach due to low variability within and among reaches. Additional water samples were collected for the analysis of metals (including mercury) at the same sites. These samples represented spot measurements of water chemistry, and were intended to characterize the chemical habitat at the time of sampling to provide supporting information that could help in understanding the distribution of BMI assemblages. Water chemistry samples were kept cool and sent to Taiga Environmental Laboratory for analysis. A handheld meter was used to record air and water temperature, pH, specific conductivity, dissolved oxygen, and turbidity on-site.

Sediment samples were collected at kick-site 3 in each reach to analyze metals and polycyclic aromatic hydrocarbons (PAHs) in soil. Because BMI live in contact with or burrow within the sediment, contaminant concentrations within the sediment may more accurately reflect their exposure levels. Sediment samples were taken from within the channel and placed into jars. Sediment samples were kept cool and sent to ALS Labs for analysis.

BMI kick samples were collected at each kick-site (see Table 1 and Table 9 for details on sites sampled in 2021) using a modified travelling kick method (Lento 2018). The operator held a 400-µm-mesh kicknet with an attached collection cup downstream while standing in the river near the shore at a wadeable depth (approximately 1 m). The operator then kicked and disturbed the substrate upstream of the net for a period of three minutes while moving upstream in a slight zig-zag fashion (maintaining the same approximate depth). Because of the size of each river, sampling remained in the nearshore habitat rather than attempting to cross the channel as in a standard kick sample method. Samples were retrieved from the net and collection cup and stored in 95% ethanol for transport to the lab for sorting and identification. Samples were sorted and identified following standard CABIN protocols (Environment Canada 2014) by Biologica Environmental Services Ltd. In brief, samples were sorted using a Marchant box to randomly sub-sample until at least 300 individuals were counted. BMI were identified to the lowest practical taxonomic level. In addition, a large/rare sort was completed following the sub-sampling procedure, with an abbreviated survey of the remaining cells in the Marchant box to pick out any large or particularly rare taxa that might have been missed as part of the sub-sampling process. Although a large/rare sort is not part of the standard CABIN laboratory procedures, the use of this approach recognizes that sub-sampling procedures may exclude large taxa that contribute a great deal to biomass and secondary production in the system, but that are fewer in number and thus less likely to be encountered than smaller, more common taxa. Inclusion of these organisms provides a more accurate measure of diversity. Individuals identified as part of a large/rare sort may include taxa from families of large-bodied dragonflies and stoneflies, as well as large molluscs.

Modified CABIN field survey sheets (Environment Canada 2012) were completed at each site in order to characterize the in-stream and surrounding habitat. This survey included a description of riparian vegetation, surrounding land use, and % cover of macrophytes and % cover of periphyton in the river at each site. In addition, water velocity was measured, and a modified rock walk was completed at each site (though water velocity measurement was not possible in all reaches in 2021). For the rock walk,

operators selected substrate particles at random and measured the intermediate axis (b-axis) of each particle to the nearest mm to characterize substrate composition. This was completed for 20 substrate particles at each site. Rock walk data were summarized as percent composition in each particle size class.

2.3. Data Analysis

2.3.1. 2021 Hydrologic Conditions

High flows in the Hay and Slave Rivers in 2021 made it impossible to sample the Hay River and made access to sites in Reach 4A of the Slave River difficult. To characterize the flow conditions in 2021, the annual hydrographs for both rivers were examined, and simple flow metrics were compared between years. To reflect recent changes experienced by the BMI assemblage prior to sampling, antecedent conditions were summarized as the median flow in the 30 days preceding sampling and in the 60 days preceding sampling. The coefficient of variation (CV; calculated as the mean divided by the standard deviation, and converted to a percentage) was also calculated for each time period and for each year, in order to quantify variability in antecedent flow conditions. Although the Hay River was not sampled, the 2021 sampling dates for the Slave River were used as a reference point for the Hay River to compare conditions between years.

2.3.2. 2021 Slave River Assessment

Data from the Slave River were analyzed in a similar manner to previous reports, including a spatial analysis of variability in BMI assemblages within and among reaches, and temporal analysis of variability in BMI composition within sites and reaches to define the normal range and CES. A decrease in the number of water chemistry and sediment chemistry samples (only one sample collected per reach in 2021) limited the analysis of spatial variability in chemical parameters as well as the assessment of biotic-abiotic relationships, as noted below. However, variability in chemical parameters within and among reaches was low from 2017-2020. If changes to water chemistry, sediment chemistry, or BMI composition are noted in future years, the number of chemistry samples collected per reach can be increased to better capture spatial variability, and assessment of biotic-abiotic relationships can resume.

2.3.2.1. Spatial and temporal variation in the chemical and physical habitat

Spatial variation in the chemical and physical habitat of the Slave River was presented visually to characterize the BMI habitat at the time of sampling. Variability in water chemistry, physical habitat (e.g., substrate size, velocity, etc.), and sediment chemistry was summarized for the Slave River in a series of tables showing the results for chemical and physical habitat parameters for each reach. Water chemistry and sediment chemistry results were compared with CCME water and sediment quality guidelines, respectively (Canadian Council of Ministers of the Environment 2001b, a)(Canadian Council of Ministers of the Environment 2001b, a). However, it should be noted that as chemistry samples represented only spot measurements, any exceedances of guidelines should be interpreted with caution, as they may not reflect long-term trends.

Although water chemistry data represent spot measurements, patterns in concentrations of select parameters over time were visually presented in bubble graphs to provide an overview of the degree of temporal variability. Bubble graphs plot values as bubbles of different sizes, with sizes scaled to parameter values, for a high-level summary of spatial and temporal patterns. Bubble graphs were created for ions, nutrients, and physical parameters from water chemistry samples. Bubble graphs were also created for metals that have typically exceeded the CCME long-term exposure guidelines for the protection of aquatic life (Canadian Council of Ministers of the Environment 2001b), total aluminum and

total iron. A bubble plot was also created for arsenic from sediment chemistry samples, as this metal has typically exceeded CCME interim guidelines for sediments. Bubble plots were created using the ggplot2 package (Wickham 2016) in R Version 4.1.3 (R Development Core Team 2022).

2.3.2.2. Spatial variation in BMI assemblages

2.3.2.2.1. Biotic metrics

Spatial variability in BMI assemblage composition was summarized for the Slave River in a table showing the mean ± standard deviation of biotic metrics for each reach. Biotic metrics included many compositional metrics that are commonly used in biomonitoring (see background on metric development and diagnostic testing in Barbour et al. 1999 and references cited therein), including those that describe general patterns in diversity and abundance, and those that characterize diversity and abundance of dominant taxonomic groups (total abundance; total taxonomic richness; abundance, relative abundance, and richness of Ephemeroptera, Plecoptera, and Trichoptera (EPT; mayflies, stoneflies, and caddisflies), Chironomidae (midges), and Diptera (true flies, including midges) + Oligochaeta (segmented worms)). In 2017-2019, the abundance, relative abundance, and richness of Mollusca was included, but as these taxa have recently made up only a small portion of the assemblage, this metric was excluded from analysis. Because *Hydra* was such a dominant taxon in 2020, the metrics abundance of *Hydra* and relative abundance of *Hydra* were again included in the analysis. Calculations of richness metrics (total taxonomic richness, EPT richness, Chironomidae richness, and Diptera + Oligochaeta richness) were based on the number of unique taxa identified at the lowest practical taxonomic level, and calculation of abundance metrics was based on all individuals within the specified taxonomic group.

Box plots were used to present summaries of variation in BMI metrics within and among reaches. Box plots present the median, 25th and 75th quartiles, and the range of the data outside the lower and upper quartiles. Box plots were created using the ggplot2 package in R Version 4.1.3.

2.3.2.2.2. Multivariate analysis of composition

Multivariate analysis was used to fully characterize the biotic assemblage of each river using data for all identified taxa (not biotic metrics). This analysis was intended to assess correlations and variability within and among reaches. BMI relative abundance data were summarized for multivariate analysis at the family/subfamily level, with Chironomidae at subfamily and all other taxa at family or higher (as this level has been shown to be sufficient to characterize northern river BMI data while reducing noise from more detailed taxonomy; Lento et al. 2013, Culp et al. 2019, Lento et al. 2022b). Taxa identified to genus level were combined at the family/subfamily level, and those identified to a coarser level (e.g., order or higher) were retained if they were unique (i.e., not identified at family/subfamily or genus level in any sample from the river). Indirect gradient analysis (eigenanalysis-based multivariate approach) was used instead of a distance-based method (e.g., non-metric multidimensional scaling) in order to simultaneously represent sites and taxa relationships in low-dimensional space and easily attribute site differences to particular taxa. Spatial variation in assemblage structure (relative abundance) among sites was assessed using PCA because there was low turnover among samples, which indicated that assemblage variance was best described by a linear model (Hirst and Jackson 2007). PCA with centering/standardization by taxa (PCA of the correlation matrix) was run in CANOCO (Version 4.05; ter Braak and Šmilauer 2002).

Variability in multivariate assemblage structure among reaches was assessed statistically to determine whether there were significant differences in composition among reaches. PERMANOVA (Permutational Multivariate Analysis of Variance; McArdle and Anderson 2001, Anderson 2017), a rank-based

multivariate approximate to ANOVA, was used to test whether there were significant differences in assemblage composition among reaches based on a dissimilarity measure (Sørenson dissimilarity index, calculated for pairwise comparisons of assemblage data for each sample, to focus on differences in taxa presence across sites). Pairwise tests, analogous to post-hoc tests in univariate ANOVA, were used to identify differences among reaches when the PERMANOVA results indicated a significant effect of reach on composition. Variability within reaches was assessed using a test for homogeneity of multivariate dispersions (Anderson et al. 2006). This analysis used the site dissimilarity matrix to calculate the distance to centroid (in multivariate space) for each reach, as a measure of variability among reaches (the farther the distance to centroid, the greater the dissimilarity among sites in a reach). A permutational pairwise test was used to identify significant differences in the distance to centroid among reaches to compare the magnitude of within-reach variability. To control for an increased rate of Type I error, a false discovery rate (FDR) correction was applied to α for all pairwise comparisons (Benjamini and Hochberg 1995). Distance to centroid was plotted with a box plot to visualize within-reach variability across reaches for each river. PERMANOVA and homogeneity of multivariate dispersions were run in R version 4.1.3 using the packages vegan version 2.5-7 (Oksanen et al. 2020) and pairwiseAdonis (Martinez Arbizu 2020).

2.3.2.3. Temporal characterization of BMI assemblages

Analysis of temporal variation in monitoring data from 2017 to 2021 began with a general assessment of changes to composition, including taxonomic richness and abundance. Pie charts of the average relative abundance of major invertebrate groups (e.g., numerically abundant insect orders and orders or classes of non-insects) across all reaches were used to compare composition between years (2017, 2018, 2019, 2020, and 2021) for the Slave River. These plots were used for a visual assessment of major changes that occurred between sampling years. Bubble plots were used to visualize the degree of change in metric values among reaches and among years, with separate bubble plots created for taxonomic groups that appeared particularly dynamic (EPT, Chironomidae, and *Hydra*). Temporal line plots were created for each biotic metric, with mean metric values for each reach plotted for each sampling year. Data from all reaches were overlain on the same plot for each metric to examine general patterns of change over time.

2.3.3. Normal range and CES for BMI metrics

The CES approach makes use of the variation among samples to determine if test samples are impaired (i.e., if they fall outside the normal range, or range of natural variability). The CES is based on variability in the data, and changes in habitat conditions that result from natural variability (i.e., due to shifts in flow, timing of the spring freshet, water temperature, etc.) may lead to different normal ranges from one year to the next. The greater the number of years of data that can be used to develop normal range estimates and set CES, the closer the estimates will be to accurately and precisely capturing natural variability in the system. In this report, CES is used to assess within-year variability as well as variability across the five years of sampling in the Slave River. Where particular years appeared to differ with respect to one or more metrics, the normal range was also estimated and examined for subsets of years.

2.3.3.1. Within-year variability

The normal range and CES were initially developed using 2021 data to assess within-year variability among sites (Arciszewski and Munkittrick 2015). CES limits were determined for the Slave River by calculating the mean and standard deviation of each BMI metric using 2021 data, and setting bounds of CES equal to the mean ± 2 SD, following the approach of previous BMI monitoring programs (see Munkittrick et al. 2009). BMI data from 2021 were also compared with CES limits calculated from the combined 2017-2021 data, to look at variation in the current year relative to all years of sampling (multi-year CES).

2.3.3.2. *Temporal variability*

The report on 2019 sampling results (Lento 2021) provided the first opportunity to assess temporal variability in normal range and CES for the rivers. This approach estimates the normal range of variability over time at a specific location (Arciszewski and Munkittrick 2015), here at the site scale and at the reach scale. For the BMI monitoring plan in the Hay and Slave rivers, where the end goal is to be able to detect impacts from upstream land use when they occur, reach-specific temporal CES will allow for the determination of the magnitude of change required at that location to trigger additional sampling or investigation of possible impacts. These location-specific normal ranges will capture the natural inter-annual variability within the system, and can be adjusted with the addition of new data and with shifts in normal range that occur as a result of climate change.

Critical Effect Size (upper and lower boundaries of the normal range) can be determined using different measures of variability (see Munkittrick et al. 2009 for an overview of approaches). For univariate metrics, the temporal normal range is calculated using a grand mean (the mean of means for all sample years) and standard deviation (the standard deviation of means for all sample years), with CES calculated as the grand mean ± 2SD (Arciszewski and Munkittrick 2015). For the Slave River, the normal range was calculated at the river scale and at the reach scale. At the river scale, the grand mean was calculated as the mean of annual means across all sites in the river, and SD was calculated from the same annual means. At the reach scale, the grand mean was calculated as the mean of annual means across all sites in the reach, and SD calculated from the same annual means.

Temporal CES was plotted to assess site-scale temporal variability relative to the normal range for the river, and to assess reach-scale temporal variability relative to the normal range for the reach. At the site scale, BMI metrics were plotted as the multi-year mean (2017-2021 data) ± SE (standard error) for each site, and they were compared with the temporal CES for the river (grand mean ± 2SD for the river). At the reach scale, BMI metrics were plotted as the mean (across sites) ± SE for each year (2017, 2018, 2019, 2020, and 2021), and they were compared with the temporal CES for the reach (grand mean ± 2SD for the reach).

With five years of data in the Slave River, it is possible to begin to characterize the normal range of variability in metrics while identifying and potentially omitting years in which extreme values were recorded, for example, due to high water levels in 2020. Such a critical assessment of the baseline data can be used for adaptive monitoring to refine the normal range and eliminate the effect of noise (Arciszewski and Munkittrick 2015). It also allows for a better understanding of what changes might be expected under particular conditions, such as those observed in 2020. In this way, assessment of variability in the normal range and changes across the five years of sampling can be used to support conclusions and recommendations for future years of sampling.

2.3.4. Multivariate Normal Range and CES

Multivariate temporal patterns were assessed for the Slave River (2017-2021) to further test the application of normal range and CES in the context of the full assemblage. Initially, 95% normal probability ellipses were used as a measure of normal range, to evaluate the degree of assemblage-level change across sampling years. For this analysis, a single PCA was run for each river with all years of data included (2017-2021) and 95% normal probability ellipses were created for each sampling year, allowing for a visual assessment of inter-annual variability. The degree of overlap of probability ellipses was indicative of the similarity in assemblage structure between years. The normal probability ellipses indicated the area of multivariate space in which there was a 95% probability that samples would fall if they were part of the same population (i.e., representative of samples from the year that was used to create the ellipse). Samples falling outside the probability ellipse for one year were therefore deemed

to have a different assemblage composition from sites within the ellipse. This approach follows that of the Reference Condition Approach utilized by CABIN, which makes use of probability ellipses around reference sites to determine whether test sites are impaired. However, the use of probability ellipses does not recognize the non-independence of samples that results from re-sampling the same sites across years, and though it captures general variability in composition at a river scale, it does not accurately assess the degree of temporal variation within sites or reaches.

To quantify temporal variability at the site scale, Procrustes analysis was used to compare the spatial arrangement of samples in multivariate space between ordinations from different years. Procrustes analysis can be used to determine whether two ordinations (e.g., PCAs) are more similar than could occur by chance. One ordination (the rotational ordination) is rotated and stretched to best match the other ordination (the target ordination) and the fit of the two ordinations is assessed using the sum of squared residuals (m_{12}^2) for sample points (Jackson 1995). A randomization test is run with the analysis by comparing 999 random configurations of the sample points with the target ordination, and a significant result (at $\alpha = 0.05$) indicates that the target and rotational ordinations are more similar than could occur by chance. When two ordinations are found to be statistically significantly similar, it indicates that the spatial arrangement of sites in relation to each other in multivariate space did not change significantly from one year to the next, which speaks to temporal stability in assemblage composition.

Pairwise Procrustes analyses of ordinations among all possible combinations of years were used to assess the degree of similarity in assemblage structure over time. The goal in this assessment was to use Procrustes residuals as a measure of inter-annual variability in assemblage structure at the site scale. Calculation of the sum of squared residuals for the PCoA dissimilarity matrix was done through pairwise comparison of years with Procrustes analysis, with each pairwise comparison including only the sites that were sampled in both years. The number of sites contributing to each m_{12}^2 value in the dissimilarity matrix therefore differed depending on the pairwise comparison. The m_{12}^2 (sum of squared residuals) from each Procrustes analysis was extracted and used as a dissimilarity measure for each pairwise comparison of years. Because the number of sites differed, there were natural differences in the magnitude of m_{12}^2 among comparisons (i.e., comparisons with 2020 had naturally lower sum of squared residuals because there were fewer site residual values to contribute to m_{12}^2). To account for this and more accurately represent differences among years, m_{12}^2 values were divided by the number of sites in the pairwise comparison, to scale the values based on sample size. A dissimilarity matrix of those values was constructed, with values of 0 indicating complete similarity and values larger than 0 indicating increasing dissimilarity in the spatial arrangement of sites between years. The dissimilarity matrix of scaled Procrustes residuals was used in a Principle Coordinates Analysis (PCoA) to evaluate similarities among years and visualize change trajectories (following Lento et al. 2008). Years that plotted close to each other in the PCoA were more similar, while those that plotted farther apart were more dissimilar.

Procrustes residuals were also extracted for each site from each pairwise comparison of years. These site-scale residuals give a measure of the degree of shift in the relative position of samples in the ordination from one year to the next. While not a direct measure of BMI assemblage change, residuals indirectly quantify such change. The degree to which a site changes position in ordination space from one year to the next is a measure of how the BMI composition of the site has changed relative to other sites. For example, if a site becomes more strongly associated with a different set of taxa, it might change position in the ordination. Changes in the position of many sites in relation to the previous year indicate a greater amount of change in assemblage composition relative to other sites that did not shift position. Site-scale Procrustes residuals were used to build CES plots, with the normal range defined as the grand mean of residuals (mean of mean annual residuals) across all year comparisons ± 2 SD, and

with each site plotted as the mean residual ± SE. For calculation of site-scale normal range and CES, Procrustes analysis was run on a subset of sites that was sampled in all years, to ensure each site mean was based on the same number of pairwise comparisons. For the Slave River, two sets of Procrustes analyses were run: (1) analysis of 2017-2019 and 2021 (excluding 2020), using the 33 sites sampled in all four years (i.e., excluding Reach 6 and Reach 4A sites 1 and 2), and (2) analysis of 2017-2021, using the 17 sites sampled in all five years (i.e., sites sampled in 2020, excluding Reach 6). The multi-year PCA and Procrustes analysis were run with the vegan package in R, probability ellipses were created with the package ggfortify version 0.4.14 (Tang et al. 2016), and the PCoA was run with the ape package version 5.6.1 (Paradis and Schliep 2019).

2.3.5. Test of Generalized Procrustes Analysis

Finally, the use of Generalized Procrustes Analysis (GPA; Gower 1975) was explored for temporal comparisons. GPA creates a consensus ordination from multiple sets of multivariate data, with the consensus ordination representing the average of all ordinations (Matteucci and Pla 2010). The test is typically used to combine different sets of values for the same individuals/locations; for example, when a group of judges is scoring based on several different sets of criteria, GPA can be used to create a consensus matrix of scores for each judge based on all criteria groups. Matteucci and Pla (2010) applied GPA to combine environmental quality scores and scores from social surveys on land quality for the same locations to develop an integrated summary of land quality that could be used as a reference point for management. In the same way, creating a consensus ordination using multiple years of BMI assemblage data for all sites in a river can be used to create a summary of baseline temporal variability among sites, and provide a reference spatial arrangement of sites against which future data can be compared. For example, if the consensus ordination summarizes the spatial arrangement of sites in ordination space in the first five years of sampling, an ordination of data from year six could be compared with the consensus ordination using Procrustes analysis to determine whether it differed significantly. Significant differences in this case would indicate that sites changed relative to each other in multivariate space, reflecting changes in assemblage structure in one or more sites relative to the consensus ordination.

In this report, GPA was initially tested by creating a consensus ordination using data from 2017-2019 (2020 was omitted because it included so few sites, and because assemblages were shown to differ in some respects due to the high water levels). The coordinates of the consensus matrix were extracted and Procrustes analysis was used to compare the consensus matrix with the results from the PCA ordination of 2021 data. This was intended as an initial exploration of the potential utility of this approach. The coordinates of the consensus matrix using data from 2017-2019 were also used in a Procrustes analysis with the PCA ordination of 2020 data, to get a sense of how different that year was for the 18 sites that were sampled. A consensus ordination using data from 2017-2019 and 2021 combined was also created as a reference point for comparison with future data collected in the river. GPA was run in R version 4.1.3 using the package FactoMineR (version 1.34; Le et al. 2008).

3. Results and Discussion

3.1. 2021 Hydrologic Conditions

River flow has been a significant source of variability in habitat conditions and a constraint on sampling efforts through the first five years of the GNWT and GOA large transboundary river BMI monitoring program. Water levels have determined the timing of sampling in the Hay and Slave rivers each year, but have also limited the extent to which sampling could take place. In addition to causing logistical

constraints for sampling, high variability in river flow from one year to the next can also have noticeable impacts on BMI assemblage composition. The timing, magnitude, duration, and variability of flows within and among years are known to be significant drivers of the structure and function of river communities (Bunn and Arthington 2002, Monk et al. 2008, Peters et al. 2014, Monk et al. 2018). River flow affects the availability of suitable habitat for organisms, including substrate composition and stability and the presence and distribution of riffle, run, and pool habitat types, all of which affect the composition of benthic communities (Bunn and Arthington 2002). The timing and duration of low/peak flows, ice on/off, and rise rates/fall rates (rates of increasing flow and decreasing flow) have implications for life history processes, including recruitment and spawning of fish, the timing of dispersal, and the timing of insect emergence (Bunn and Arthington 2002, Peters et al. 2014). The magnitude of flows can affect connectivity, including access to floodplains (Bunn and Arthington 2002, Peters et al. 2014). In addition, higher flow years may favour BMI taxa that have adaptations for fast velocities, and low flow years may result in a dominance of taxa that are well-adapted to slower velocities (Monk et al. 2008). When inter-annual changes in flow are severe enough, they may cause a shift in the benthic community if high flows and benthic scouring wash out some individuals, or if increases in water depth alter habitats from riffles to runs or pools. Analysis of flow data in relation to BMI assemblage data for 2017-2020 did not identify strong indicator taxa or community change points in response to flow velocity for the Slave River, which suggested that flow-based CES may not be possible for that river (Lento 2022). However, patterns in biotic metrics and assemblage composition should still be interpreted in relation to our knowledge of broad-scale variation in water level among years and the potential influence on biota.

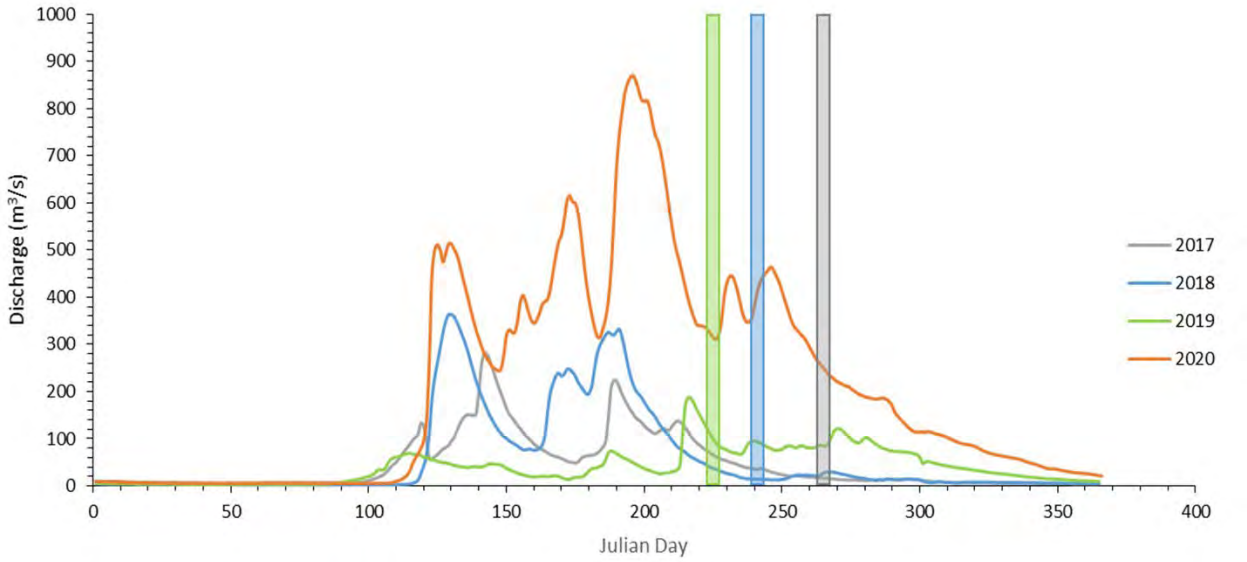

Figure 5. Hydrographs for the Hay River in 2017 (grey), 2018 (blue), 2019 (green), and 2020 (orange), with vertical shaded bars indicating the timing of sampling in each year (no vertical shaded bar for 2020 because it was not possible to sample). Data for Hay River near ALTA/NWT boundary (station 07OB008) from wateroffice.ec.gc.ca. Data were not available for this gauge were not available for 2021.

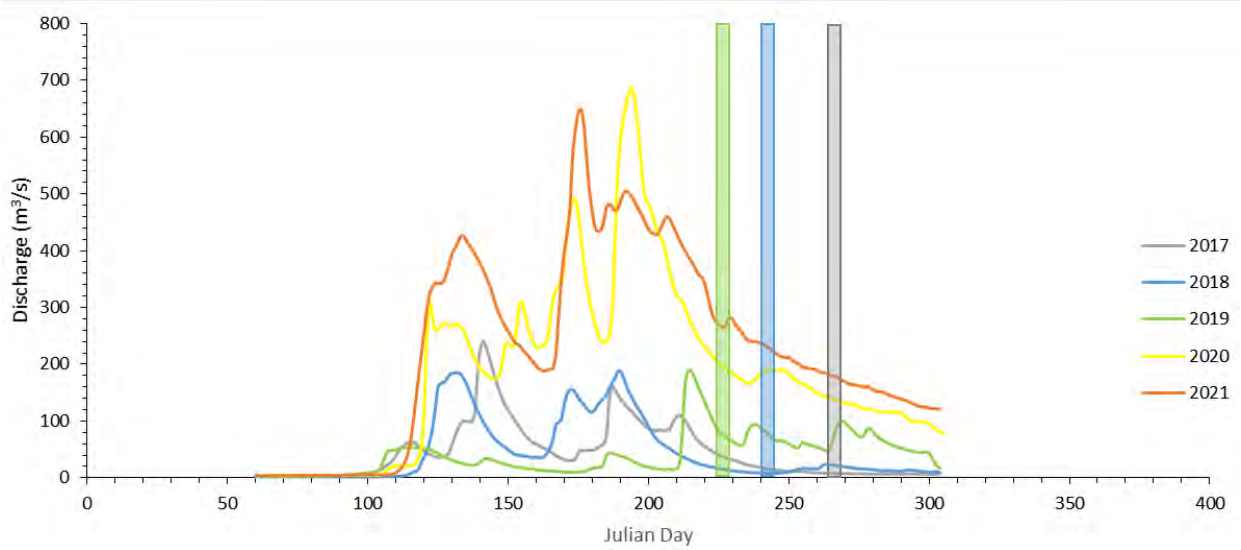

Figure 6. Hydrographs for the Hay River in 2017 (grey), 2018 (blue), 2019 (green), 2020 (yellow), and 2021 (orange), with vertical shaded bars indicating the timing of sampling in each year (no vertical shaded bar for 2020 or 2021 because it was not possible to sample). Data for Hay River near Meander River (station 07OB003) from wateroffice.ec.gc.ca. Reporting for Hay River near ALTA/NWT boundary (station 07OB008) ended in 2020.

*Table 2. Antecedent hydrology metrics for the Hay River for 2017-2021, including discharge (Q (m^3/s)) at time of sampling (sample date for the Slave River used for 2020 and 2021 for context, as Hay River discharge was too high for sampling), median discharge, and the coefficient of variation (CV) of flow, calculated for 60 days and 30 days prior to sampling in each year. Data for 2017-2020 are from Hay River near ALTA/NWT boundary (station 07OB008) and data from 2021 are from Hay River near Meander River (station 07OB003), with the change in gauge indicated in the table with *. Data from wateroffice.ec.gc.ca. Reporting for Hay River near ALTA/NWT boundary (station 07OB008) ended in 2020.*

Year	At Sampling	60 Days Prior to Sampling		30 Days Prior to Sampling	
	Q (m^3/s)	Median Q (m^3/s)	CV (%)	Median Q (m^3/s)	CV (%)
2017	16.9	42.6	67.6	23.7	31.1
2018	14.6	97.6	80.9	37.5	57.0
2019	100	37.6	84.0	44	74.8
2020	192	339	24.3	264	26.1
2021*	208	322	29.5	241	11.6

3.1.1. Hay River

Water levels in the Hay River have been extremely variable across the five years of the sampling program. In 2017, water levels were low enough to make it difficult to access the reaches downstream of the boat launch, and sandbars throughout the river added to the challenges of sampling. In 2018, sampling was shifted earlier in the year to ensure higher water levels, but water levels in the Hay River were at or below record minimum levels at the end of August 2018 (ECCC gauge Hay River near ALTA/NWT boundary, station 07OB008; Figure 5), which resulted in lower water levels for sampling than observed the previous year. The timing of sampling was shifted earlier in August in 2019 because water levels were low during the usual spring freshet (Figure 5), and there were concerns that many sites would be inaccessible. However, a surge in discharge prior to sampling led to very high water levels at the time of sampling compared to previous years (discharge of approximately 100 m^3/s, compared with 16.9 m^3/s and 14.6 m^3/s in 2017 and 2018, respectively; Table 2). In 2019, some aspects of sampling (e.g., rock walk) could not be completed at some sites where water levels were too high. Flow conditions became even more extreme in 2020, as discharge in the summer of 2020 peaked at more than twice the maximum discharge observed in 2019, and flows remained high throughout the summer

and fall (Figure 5). On October 5, 2020 (the first day of sampling in the Slave River), discharge in the Hay River was 192 m^3/s, nearly twice the discharge observed when the river was sampled in 2019 (Table 2). As a result, it was not possible to access the sample sites in the Hay River in 2020, and sampling could not take place.

Water levels in the Hay River remained high through 2021, and were generally higher than observed in 2020 (Figure 6). Discharge measurements from the gauge near the Alberta/NWT border were discontinued in 2021, and the next nearest gauge (Hay River near Meander River, station 07OB003) recorded lower peak flows from 2018-2020 compared with those recorded near the border. However, comparisons among years for the Hay River near Meander River gauge indicate the magnitude of difference in flows across all 5 years, showing the highest fall flows in 2021 (Figure 6). At the time of sampling for the Slave River in 2021 (September 8), discharge in the Hay River near Meander was 208 m^3/s, making it impossible to access sample sites in the Hay River.

Antecedent hydrologic conditions in the 60 days and 30 days prior to sampling were compared among years by calculating two metrics of flow: the median discharge and the coefficient of variation of discharge, the latter of which provides a standardized measure of variation in flow. When compared

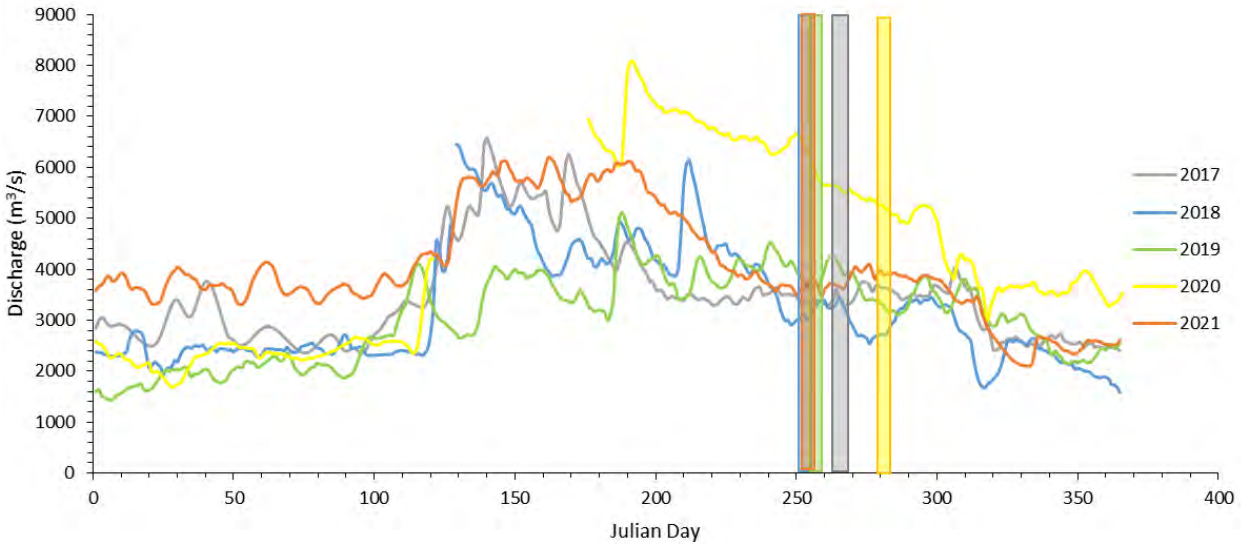

Figure 7. Hydrographs for the Slave River in 2017 (grey), 2018 (blue), 2019 (green), 2020 (yellow), and 2021 (orange), with vertical shaded bars indicating the timing of sampling in each year. Note that sample periods for 2018 and 2021 overlap. Data for Slave River near Fort Fitz (station 07NB001) from wateroffice.ec.gc.ca.

across the period of 60 days prior to sampling, median discharge estimates in the Hay River in 2020 and 2021 were clearly much higher than in previous years, differing from antecedent median flow in 2019 by an order of magnitude (Table 2). However, discharge was much less variable in 2020 and 2021 than in earlier years, which reflected the consistently high water levels in these two years. A similar pattern was observed when antecedent conditions in the 30 days prior to sampling were compared among years. Median flow in 2020 and 2021 was much higher than in previous years (an order of magnitude higher than in 2017), but variability was lower, particularly so in 2021 (Table 2). It is important to note that the lower discharge in 2021 compared to 2020 shown in Table 2 reflects a change in discharge gauge location; Figure 6 indicates that measurements at the same gauge show higher discharge in 2021. These results indicate that the extreme changes to flow conditions in the Hay River in 2020 continued into 2021, and make it clear that sampling was again not possible in 2021, nor would it have likely yielded representative samples of BMI assemblages in the river under such extreme conditions.

3.1.2. Slave River

The Slave River is a large, fast-flowing river with high discharge, but flows in this river have also been variable since 2017, with the greatest change evident in 2020. In 2018, there was a late peak in water levels, occurring only 45 days prior to sampling (Figure 7), and this peak appeared to have influenced the biotic assemblages of the river (Lento 2020). The hydrograph in 2019 also differed from what was observed in 2017, this time showing a flatter profile during the typical spring freshet, with a more gradual increase in water levels across the summer, and a more gradual and flashy decline (Figure 7). In both 2018 and 2019, water levels were higher at the time of sampling than in 2017, but in 2019 the hydrograph was generally flatter across the spring/summer than in 2018, with less seasonality to flows. In 2020, water levels in the Slave River peaked well above previous years and remained extremely high at the time of sampling in early October (Figure 7). Although there was a gradual decline from peak flows, water levels remained high through the end of the year. Whereas discharge at the time of sampling ranged from 3220 to 3900 m^3/s between 2017 and 2019, the discharge was 5260 m^3/s at the time of sampling in 2020. This led to difficulties accessing all sample sites in the river, and many sites could not safely be sampled. In 2021, water levels remained high through the winter, but the peak at the spring freshet was similar in magnitude to that observed in 2017, and there was a general decline in flow following the freshet that was more similar to previous years (Figure 7). Discharge at the time of sampling in 2021 (3560 m^3/s) was similar to that recorded in 2017 (Table 3).

Table 3. Antecedent hydrology metrics for the Slave River for 2017-2021, including median discharge (Q (m^3/s)) and the coefficient of variation (CV) of flow, calculated for 60 days and 30 days prior to sampling in each year. Data for Slave River near Fort Fitz (station 07NB001) from wateroffice.ec.gc.ca.

Year	At Sampling	60 Days Prior to Sampling		30 Days Prior to Sampling	
	Q (m^3/s)	Median Q (m^3/s)	CV (%)	Median Q (m^3/s)	CV (%)
2017	3480	3430	2.7	3490	3.2
2018	3220	4100	19.0	3730	14.7
2019	3900	4070	7.7	4070	6.1
2020	5260	6360	8.1	5640	7.4
2021	3560	4340	17.3	3820	4.6

Antecedent hydrologic conditions were compared among years using metrics summarizing the periods 60 days and 30 days prior to sampling. Over the period 60 days prior to sampling, median flows in 2020 were nearly twice the median discharge observed in 2017, though variability was low (Table 3). In 2021, antecedent conditions in the 60 days prior to sampling were more similar to those observed in 2018 and 2019, and higher variability reflected the continual decline in flow over that period. Similar patterns were observed when flow metrics for the Slave River were compared among years for the 30 days prior to sampling, with much higher median discharge in 2020 than in previous years, and median discharge in 2021 that was more comparable to previous sampling years (Table 3). Flow conditions in 2021 appeared to show a return to more typical flows for this river, although the effects of the high flow in 2020 (which continued into the winter and early spring of 2021) may still be evident in the system.

3.2. 2021 Slave River Assessment

3.2.1. Chemical and physical habitat

3.2.1.1. Water chemistry

Water chemistry samples were collected in the Slave River to act as supporting variables for the BMI data. These samples represented spot measurements of water chemistry conditions at the time of

sampling, and were collected at a single site per reach (Table 1) to broadly describe the chemical environment. The Slave River is a large river (wetted width at reaches in 2019 was > 100 m on average), and habitat conditions and assemblages are generally expected to vary somewhat among reaches, as they are located far apart geographically. However, flow also plays a large role in water chemistry conditions in the river. Discharge in the river has been highly variable among sampling years, and the differences in peak flow magnitude as well as hydrograph seasonality have the potential to lead to variability in water chemistry between sampling years (Table 3, Figure 7). In their analysis of long-term trends in water quality of the transboundary waters of the Slave River, Sanderson et al. (2012) found

Table 4. Summary of ion, nutrient, and physical water chemistry parameters sampled in the Slave River at six sample reaches in 2021. Each value represents a single sample taken at a single site within each reach, with the exception of Reach KS3b, where values indicate the mean of two replicate samples ± standard deviation. The detection limit is presented where samples were below detection. Reaches are ordered from upstream (KS1) to downstream (KS5). No parameters exceeded the Canadian Water Quality Guidelines for the Protection of Aquatic Life (Canadian Council of Ministers of the Environment 2001b).

Parameter	SR-KS1B	SR-KS2A	SR-KS3B	SR-KS4A	SR-KS4B	SR-KS6B	SR-KS5A
Alkalinity (mg/L)	65.7	63.5	64.2 ± 0.3	64.1	64.1	63.3	64.1
Ammonia as N (mg/L)	0.010	0.007	0.010 ± 0.001	0.006	0.009	0.009	0.010
Calcium (mg/L)	20.5	19.6	19.6 ± 0.0	20.3	19.4	19.6	19.9
Chloride (mg/L)	9.10	6.00	7.85 ± 0.07	5.70	5.80	5.90	5.60
Specific Conductivity (µS/cm)	182.0	172.0	179.5 ± 0.7	173.0	175.0	172.0	172.0
Hardness (mg/L)	74.6	70.7	71.5 ± 0.6	72.5	69.8	70.5	70.9
Magnesium (mg/L)	5.69	5.28	5.48 ± 0.14	5.30	5.18	5.23	5.16
Nitrate (mg/L)	0.020	0.020	0.020 ± 0.000	0.010	0.020	<0.01	0.020
Nitrite (mg/L)	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Nitrate+Nitrite (mg/L)	0.020	0.020	0.020 ± 0.000	0.010	0.020	<0.01	0.020
Dissolved N (mg/L)	0.220	0.200	0.220 ± 0.057	0.160	0.220	0.190	0.180
Total N (mg/L)	0.340	0.300	0.320 ± 0.014	0.310	0.330	0.300	0.290
DOC (mg/L)	6.20	5.80	5.85 ± 0.07	5.60	5.70	5.50	5.50
TOC (mg/L)	6.10	5.90	5.80 ± 0.00	5.80	5.70	5.60	5.70
Ortho-Phosphate (mg/L)	0.0040	0.0030	0.0030 ± 0.0000	0.0030	0.0040	0.0030	0.0030
pH	8.03	8.05	8.06 ± 0.01	8.07	8.07	8.07	8.07
Dissolved P (mg/L)	0.005	0.003	0.003 ± 0.000	0.003	0.003	<0.002	0.003
Total P (mg/L)	0.054	0.047	0.044 ± 0.001	0.051	0.053	0.037	0.039
Potassium (mg/L)	1.09	1.08	1.07 ± 0.01	1.04	1.01	0.99	1.00
Sodium (mg/L)	8.26	6.48	7.14 ± 0.08	6.02	6.15	6.01	6.07
TDS (mg/L)	126.0	122.0	114.0 ± 2.8	116.0	112.0	112.0	104.0
TSS (mg/L)	36.0	28.0	32.0 ± 5.7	46.0	36.0	26.0	38.0
Sulphate (mg/L)	16.0	16.0	16.0 ± 0.0	16.0	16.0	16.0	16.0
Turbidity (NTU)	26.4	23.7	21.2 ± 0.6	23.7	23.2	17.9	18.3

that some temporal trends in water chemistry parameters reflected temporal changes in flow (with summer/fall flows decreasing over time in the river), and correction for flow resulted in the removal of temporal trends in those parameters. Changes in flow and water chemistry patterns over the long term in this river are partially a reflection of the impacts of the William A. C. Bennett dam in the upstream Peace River basin in northern British Columbia (Glozier et al. 2009, Sanderson et al. 2012). In the short

term, interannual flow variability from 2017 to 2021 likely has also contributed to variation in water chemistry parameters between years. This is difficult to capture through annual spot measurements of water chemistry, and is better monitored through temporal trend analysis of routine sampling data. The assessment here is therefore primarily focused on characterizing the water chemistry conditions at the time of sampling and on looking broadly at the magnitude of change in these spot measurements across sample years.

Water samples were collected in each river reach at one site (see Table 1 for details) and analyzed for ions, nutrients, and physicals. A duplicate sample was collected in Reach 3. Mean levels of water chemistry parameters (Table 4) were compared with Canadian guidelines for short-term and long-term exposure to identify any reaches where water chemistry was indicative of poor water quality (Canadian Council of Ministers of the Environment 2001b). Short-term water quality guidelines have generally not been derived for the protection of aquatic life; therefore, most comparisons were with long-term exposure guidelines. Of the parameters that were tested (see Table 4), guidelines were available for ammonia, chloride, nitrate, pH, TSS, and turbidity, and there were no exceedances of these guidelines in any reach. Most water quality parameters had similar values to those observed in previous sampling years, including alkalinity, conductivity, and nutrients (Table 4). Some parameters like TDS and turbidity that were observed to be lower in 2020 remained low in 2021. TSS levels were generally low and well below the long-term (1982-2010) mean of less than 100 mg/L reported for August and September at Fort Smith (Sanderson et al. 2012). Discharge in the Slave River showed a gradual decline with a low frequency of reversals throughout the summer following the spring freshet (Figure 7), and this likely contributed to a more steady rate of sediment transport throughout this period than was observed in earlier years with more frequent reversals.

Estimates of mean TP in the Slave River were all between 0.037-0.054 mg/L, and reaches were classified as eutrophic based on the Canadian Guidance Framework (Canadian Council of Ministers of the Environment 2001b), though TP levels at Reach 6 and Reach 5 approached the upper limit of meso-eutrophic (0.035 mg/L). Slave River reaches have been classified as eutrophic to hyper-eutrophic in all five sample years, consistent with the results of Glozier et al. (2009), who found long-term trends of elevated total and dissolved phosphorus in the Slave River relative to the Athabasca and Peace Rivers that flow into the Slave. Glozier et al. (2009) determined the long-term median TP to be 0.078 mg/L for the Slave River, higher than was observed in the spot measurements taken in 2021.

Bubble plots were created to visually represent broad-scale patterns in water chemistry parameters over the five years of sampling, recognizing the limitations of assessing long-term trends in spot measurements of water chemistry. Values of parameters were averaged across all reaches for each year to give an overview of the degree of change across the 5 sample years. Although the plots do not capture variability among reaches, such variability has generally been low within sample years, and the simplicity of these plots allowed for the representation of a great deal of information in a compressed space.

Bubble plots indicated that there generally appeared to be low variability in river-scale means of water chemistry parameters over time (Figure 8; Figure 9). Dissolved and total organic carbon were both elevated in 2019 relative to other years, as was nitrate (Figure 8). Magnesium appeared to be highest in 2018, whereas chloride, alkalinity, conductivity, and total dissolved solids were highest in 2020 (Figure 8; Figure 9). TSS appeared to decrease across all sample years from 2017 to 2021 (Figure 9), which is consistent with long-term trends of declining TSS noted by Sanderson et al. (2012). But overall, variability among years was low for most parameters, and temporal changes in spot measurements were generally of low magnitude.

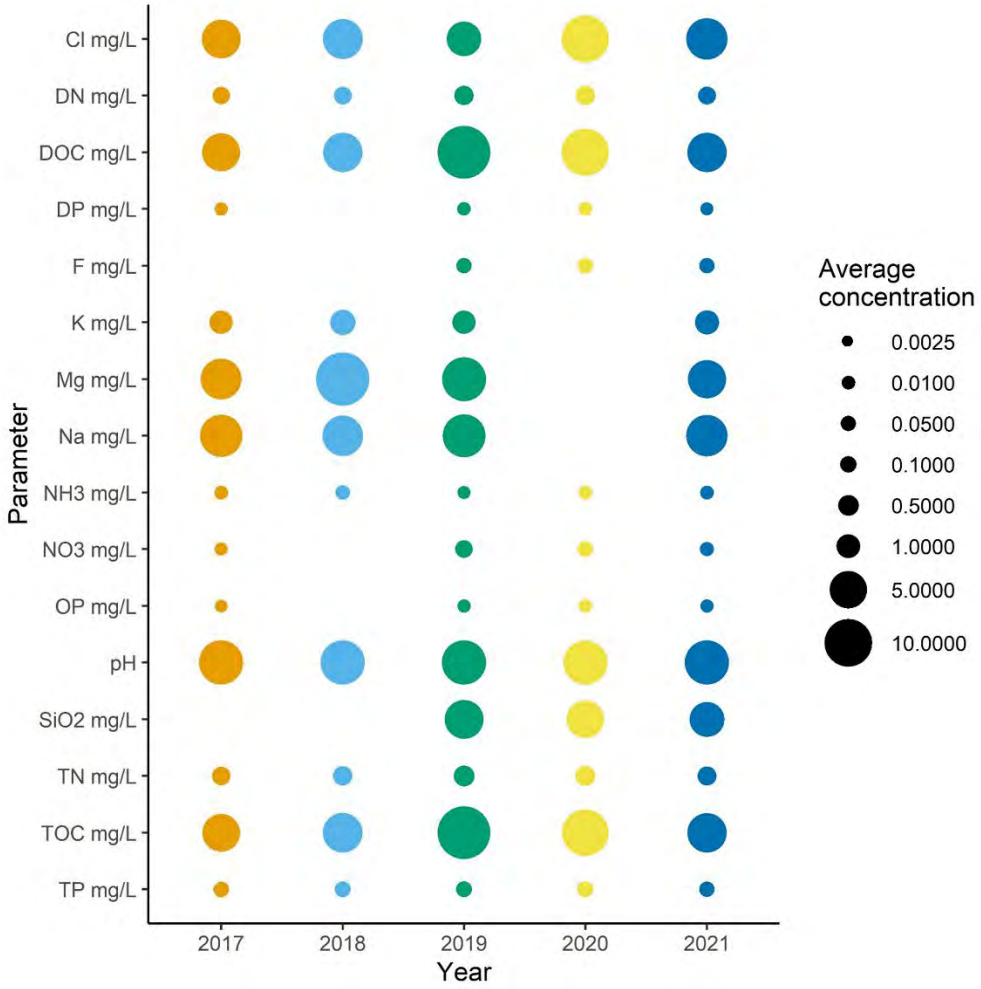

Figure 8. Bubble plot of water chemistry concentrations in the Slave River for parameters with low values, with size-scaled bubbles representing the average of all reaches for each year. Variability in size from left to right provides an indication of the degree of change over time in the Slave River. No bubble is shown for parameters that weren't measured in a particular year.

Total and dissolved metals were also measured in water chemistry samples to quantify the levels to which BMI were exposed at the time of sampling. Dissolved metals provide a more accurate estimate of the relevant exposure of biota than total metals because they are generally more biologically available than the particulate forms, which are included in estimates of total metals (Sanderson et al. 2012). Concentrations of metals were compared with water quality guidelines for the protection of aquatic life, which generally only include guidelines for long-term exposure (Canadian Council of Ministers of the Environment 2001b). Guidelines exist for total aluminum, total arsenic, total boron, total cadmium, total copper, total iron, total lead, dissolved manganese, total mercury, total molybdenum, total nickel, total selenium, total silver, total thallium, total uranium, and dissolved zinc.

Dissolved metal concentrations were generally low in Slave River reaches, with many dissolved metals below detection limit (Table 10 in Appendices). As a result, no dissolved metals exceeded long-term exposure water quality guidelines for the protection of aquatic life for those parameters that had guidelines available (Canadian Council of Ministers of the Environment 2001b).

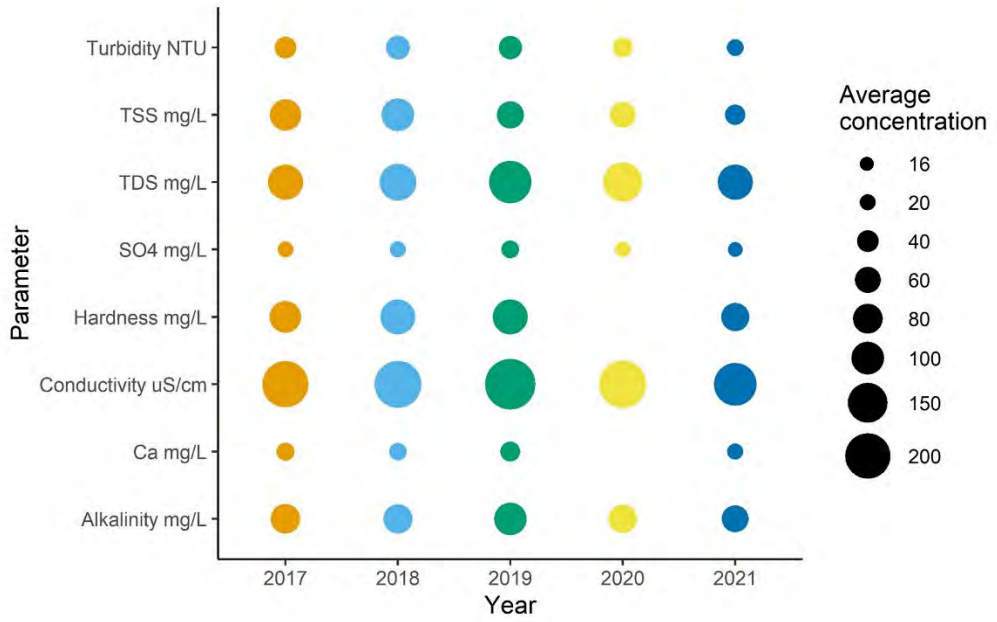

Figure 9. Bubble plot of water chemistry concentrations in the Slave River for parameters with high values, with size-scaled bubbles representing the average of all reaches for each year. Variability in size from left to right provides an indication of the degree of change over time in the Slave River. No bubble is shown for parameters that weren't measured in a particular year.

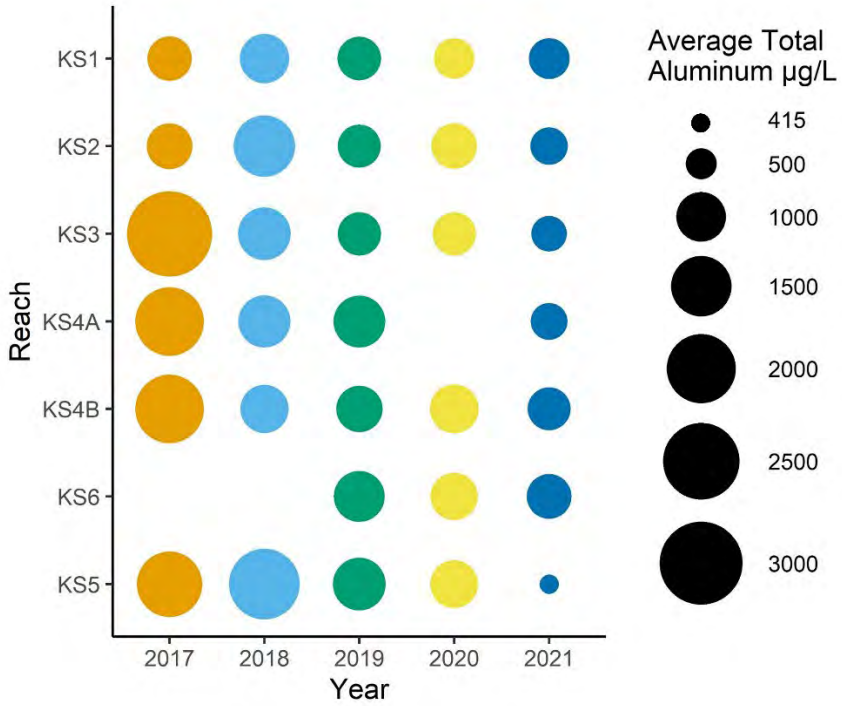

Figure 10. Bubble plot of total aluminum concentrations in water in the Slave River, with size-scaled bubbles representing the average concentration for each reach in each year. Reaches are arranged in order from upstream (top) to downstream (bottom). Variability in size from top to bottom indicates spatial variability among reaches, and variability from left to right indices of the degree of change over time in the Slave River. No bubble is shown for reaches that weren't sampled in a particular year.

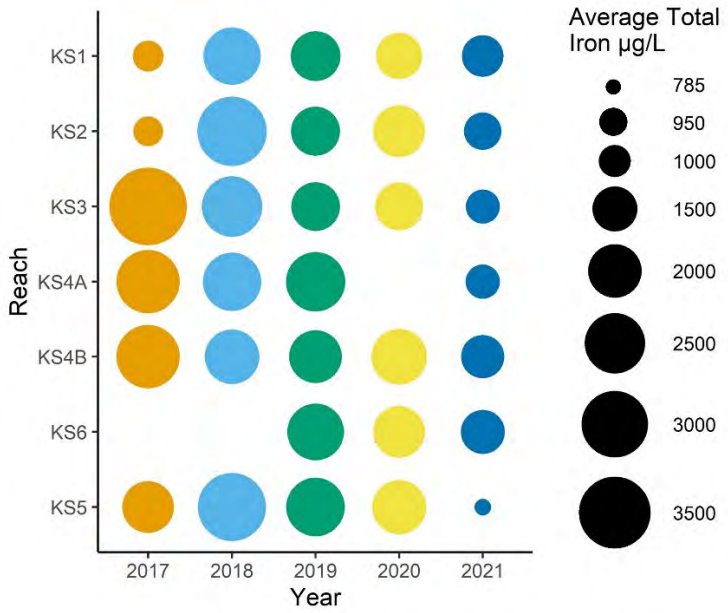

Figure 11. Bubble plot of total iron concentrations in water in the Slave River, with size-scaled bubbles representing the average concentration for each reach in each year. Reaches are arranged in order from upstream (top) to downstream (bottom). Variability in size from top to bottom indicates spatial variability among reaches, and variability from left to right indices of the degree of change over time in the Slave River. No bubble is shown for reaches that weren't sampled in a particular year.

Total metal concentrations were generally low, with most metals near or below detection limits (Table 10). As a result, there were few exceedances of long-term water quality guidelines (Table 10). Total aluminum concentrations exceeded long-term exposure water quality guidelines (100 µg/L), as they have in each year of sampling. However, concentrations of total aluminum were similar to or lower than those observed in previous sampling years, particularly when compared with values recorded in 2017 (Figure 10). Furthermore, the range of values observed in 2021 (414-831 µg /L) remained much lower than the long-term median value of 4360 µg/L reported for the Slave River at Fort Smith (Sanderson et al. 2012).The largest spatial variability among reaches was observed in 2017, and concentrations were generally more similar among reaches in each year that followed (Figure 10). In 2021, the highest concentration of total aluminum (Reach 6) was twice as high as the lowest recorded concentration (Reach 5; Table 10), but this still represented less of a difference among reaches than was observed in 2017. Total iron also exceeded long-term water quality guidelines, as concentrations in all reaches ranged from 784-1450 µg/L (Table 10), which was above the CCME long-term exposure guideline of 300 µg/L (Canadian Council of Ministers of the Environment 2001b). Total iron concentrations in the Slave River have exceeded CCME long-term exposure guidelines in each year of sampling (Figure 11). Similar to total aluminum, concentrations in 2021 were near to or lower than concentrations observed in previous years in all reaches, and the highest values and greatest variability was observed in 2017 (Figure 11). Concentrations observed in 2021 were lower than the long-term median value of 3526 µg/L reported for the Slave River at Fort Smith (Sanderson et al. 2012). Furthermore, the Federal Environmental Quality Guideline for iron (Environment and Climate Change Canada 2019) suggests a water quality guideline of 3442 µg/L based on a DOC concentration of 5.8 mg/L and a pH of 8.06 (as observed in the Slave River samples in 2021). Following the federal guideline from ECCC, all concentrations of total iron were well within acceptable limits.

Table 5. Physical habitat variables measured in the Slave River in 2021, summarized by reach. Velocity (spot measurement) is presented as mean ± standard deviation for reaches where > 1 site was sampled (velocity measurements were not available for three reaches); dominant streamside vegetation and periphyton coverage are presented as the most common category in each reach across sampled sites; substrate composition is presented as the sum of rock counts for each reach (20 rocks measured per

site), adjusted to a percentage where fewer than 5 sites were sampled. Sites are ordered from upstream (KS1) to downstream (KS5).

Variable	SR-KS1	SR-KS2	SR-KS3	SR-KS4A	SR-KS4B	SR-KS6	SR-KS5
Velocity (m/s)	0.56 ± 0.29	0.25 ± 0.07	NA	NA	NA	0.02 ± 0.35	0.44 ± 0.06
Wetted width (m)	346.0	196.0	667.0	175.0	173.0	582.0	205.0
Dominant streamside vegetation	deciduous trees	deciduous trees	coniferous trees	coniferous trees	deciduous trees	deciduous trees	deciduous trees
Periphyton coverage	< 0.5 mm thick	< 0.5 mm thick	< 0.5 mm thick	< 0.5 mm thick	< 0.5 mm thick	< 0.5 mm thick	< 0.5 mm thick
Substrate - sand (%)	0	17	11	3	28	15	2
Substrate - gravel (%)	1	0	3	2	0	0	3
Substrate - pebble (%)	61	34	45	53	21	46	64
Substrate - cobble (%)	38	44	39	37	45	39	28
Substrate - boulder (%)	0	2	1	2	3	0	1
Substrate - bedrock (%)	0	3	1	3	3	0	2

3.2.1.2. Physical Habitat

Measurements were taken at each site to characterize the physical habitat in BMI sampling locations, including variables such as velocity, wetted width, streamside vegetation, in-stream periphyton cover, and substrate composition (Table 5). Velocity ranged from 0.02 to 0.56 m/s on average across reaches (Table 5), though velocity was not measured at three reaches. Reaches KS1, KS2, and KS5 had similar velocity estimates to previous years, which generally ranged between 0.2-0.6 m/s, but velocity estimates at KS6 were comparatively low, although variable among sites. Wetted width varied across reaches, ranging from 173 m to 667 m. Substrate composition in reaches was predominantly a combination of pebble and cobble size classes (Table 5), consistent with previous sampling years. Periphyton coverage was recorded as < 0.5 mm thick at all sites, which is typical for a high-discharge river and similar to observations in previous years.

3.2.1.3. Sediment chemistry

Sediment chemistry samples were collected from one site in each Slave River reach (with a duplicate collected in Reach 3) and analyzed for metals and polycyclic aromatic hydrocarbons (PAHs). PAHs are common organic compounds that have natural sources such as forest fire, but that also result from human activities, and enter waterways from sources such as urban/industrial runoff, wastewater effluent, and coal and oil combustion (McGrath et al. 2019). PAHs cycle through aquatic ecosystems and can become incorporated into sediments in the benthic habitat due to sorption to particulate matter

Table 6. Summary of sediment chemistry parameters sampled in the Slave River in 2021 at seven reaches, indicating site mean ± standard deviation for Reach SR-KS3 where duplicate samples were collected, and the single sample value for all other reaches. For Reach SR-KS3, the detection limit (DL) is presented for parameters for which both samples were below DL. Values were compared with CCME sediment quality guidelines for the protection of aquatic life (Canadian Council of Ministers of the Environment 2001a), and values in bold were greater than interim freshwater sediment quality guidelines (ISQGs) whereas values in red were greater than probable effect levels (PELs). Reaches are ordered from upstream (KS1) to downstream (KS5).

Parameter	SR-KS1	SR-KS2	SR-KS3	SR-KS4A	SR-KS4B	SR-KS6	SR-KS5
Particle Size/Physicals							
% Clay (<2um)	5.2	7.8	11.6 ± 1.3	6.6	9.6	6.1	7.5
% Silt (2um - 0.05mm)	17.8	38.2	46.9 ± 5.0	25.4	27.0	22.5	44.5
% Sand (0.05mm - 2.0mm)	77.0	54.0	41.6 ± 3.7	68.0	63.4	71.4	48.0
Moisture %	20.2	36.8	37.4 ± 8.9	37.9	32.7	33.3	35.6
Metals (mg/kg)							
Antimony (Sb) (mg/kg)	0.270	0.310	0.375 ± 0.007	0.300	0.360	0.270	0.300
Arsenic (As) (mg/kg)	4.81	5.35	**6.27 ± 0.18**	5.51	**6.57**	4.87	5.27
Barium (Ba) (mg/kg)	28.2	310.0	255.0 ± 28.3	183.0	214.0	252.0	322.0
Beryllium (Be) (mg/kg)	0.230	0.330	0.360 ± 0.028	0.420	0.340	0.280	0.330
Cadmium (Cd) (mg/kg)	0.106	0.310	0.448 ± 0.017	0.261	0.377	0.255	0.315
Chromium (Cr) (mg/kg)	9.8	13.1	13.4 ± 0.78	13.9	13.3	11.6	13.4
Cobalt (Co) (mg/kg)	3.63	6.16	7.08 ± 0.29	6.81	7.47	5.97	6.15
Copper (Cu) (mg/kg)	14.70	9.52	12.50 ± 0.42	7.77	10.60	6.55	8.70
Lead (Pb) (mg/kg)	5.20	5.27	6.70 ± 0.28	5.44	6.18	4.51	5.32
Mercury (Hg) (mg/kg)	0.089	0.037	0.053 ± 0.009	0.027	0.039	0.029	0.037
Molybdenum (Mo) (mg/kg)	2.580	0.580	0.765 ± 0.021	0.590	0.690	0.500	0.600
Nickel (Ni) (mg/kg)	13.1	16.8	20.3 ± 0.28	18.1	20.1	16.6	17.4
Selenium (Se) (mg/kg)	0.260	0.260	0.365 ± 0.021	0.260	0.340	<0.2	0.240
Silver (Ag) (mg/kg)	<0.1	<0.1	0.115 ± 0.007	<0.1	<0.1	<0.1	<0.1
Thallium (Tl) (mg/kg)	<0.05	0.092	0.125 ± 0.006	0.084	0.100	0.071	0.095
Tin (Sn) (mg/kg)	<2	<2	<2	<2	<2	<2	<2
Uranium (U) (mg/kg)	1.140	0.765	0.777 ± 0.076	0.691	0.848	0.633	0.783
Vanadium (V) (mg/kg)	13.1	23.7	23.2 ± 1.98	24.2	24.2	20.9	23.8
Zinc (Zn) (mg/kg)	16.9	50.4	62.8 ± 2.05	49.7	58.9	48.3	52.8
Polycyclic Aromatic Hydrocarbons (PAHs) (mg/kg)							
1-Methylnaphthalene (mg/kg)	<10	14.0	29.0 ± 9.899	<10	12.0	10.0	25.0
2-Methylnaphthalene (mg/kg)	<10	16.0	**32.5 ± 10.607**	<10	14.0	12.0	**28.0**
Acenaphthene (mg/kg)	<5	<5	<5	<5	<5	<5	<5
Acenaphthylene (mg/kg)	<5	<5	<5	<5	<5	<5	<5
Anthracene (mg/kg)	<4	<4	<4	<4	<4	<4	<4
B(a)P Total Potency Equivalent (mg/kg)	<20	<20	<20	<20	<20	<20	<20
Benz(a)anthracene (mg/kg)	<10	<10	<10	<10	<10	<10	<10
Benzo(a)pyrene (mg/kg)	<10	<10	<10	<10	<10	<10	<10
Benzo(b&j)fluoranthene (mg/kg)	<10	11.0	17.0 ± 2.828	<10	<10	<10	15.0
Benzo(b+j+k)fluoranthene (mg/kg)	<15	<15	17.0 ± 2.828	<15	<15	<15	<15
Benzo(g,h,i)perylene (mg/kg)	<10	10.0	18.5 ± 3.536	<10	<10	<10	15.0
Benzo(k)fluoranthene (mg/kg)	<10	<10	<10	<10	<10	<10	<10
Chrysene (mg/kg)	<10	13.0	19.5 ± 3.536	<10	<10	10.0	17.0
Dibenz(a,h)anthracene (mg/kg)	<5	<5	<5	<5	<5	<5	<5

Parameter	SR-KS1	SR-KS2	SR-KS3	SR-KS4A	SR-KS4B	SR-KS6	SR-KS5
Fluoranthene (mg/kg)	<10	<10	<10	<10	<10	<10	<10
Fluorene (mg/kg)	<10	<10	<10	<10	<10	<10	<10
IACR (CCME) (mg/kg)	<0.15	<0.15	0.20 ± 0.028	<0.15	<0.15	<0.15	0.17
Indeno(1,2,3-c,d)pyrene (mg/kg)	<10	<10	<10	<10	<10	<10	<10
Naphthalene (mg/kg)	<10	<10	18.0 ± 5.657	<10	<10	<10	17.0
Perylene (ng/g)	<10	77.0	127.0 ± 24.042	41.0	49.0	53.0	101.0
Phenanthrene (mg/kg)	<10	20.0	35.0 ± 9.899	12.0	15.0	16.0	31.0
Pyrene (mg/kg)	<10	11.0	16.5 ± 3.536	<10	<10	<10	15.0
Quinoline (mg/kg)	<10	<10	<10	<10	<10	<10	<10

and subsequent settling in the sediment (Canadian Council of Ministers of the Environment 1999, McGrath et al. 2019). Because they can be found in high concentrations in sediments of lakes and rivers, they pose a toxicity threat to benthic organisms (Canadian Council of Ministers of the Environment 1999). PAHs can be classified as either low molecular weight (LMW) or high molecular weight (HMW), with the former being the more acutely toxic, and the latter being carcinogenic (Canadian Council of Ministers of the Environment 1999).

Concentrations of metals and PAHs from Slave River sediment samples were compared with CCME sediment quality guidelines for the protection of aquatic life (Canadian Council of Ministers of the Environment 2001a), which include interim freshwater sediment quality guidelines (ISQGs) and probable effect levels (PELs). Sediment quality guidelines were available for the metals arsenic, cadmium, chromium, copper, lead, mercury, and zinc, and the PAHs 2-methylnaphthalene, acenaphthene, acenaphthylene, anthracene, benz[a]anthracene, benzo[a]pyrene, chrysene, dibenz[a,h]anthracene, fluoranthene, fluorene, naphthalene, phenanthrene, and pyrene. In addition, benzo[a]pyrene Total Potency Equivalents and the Index of Additive Cancer Risk (IACR) were compared with guideline levels to ensure protection of humans and drinking water, respectively (Canadian Council of Ministers of the Environment 2010). Although this assessment is not specifically focused on drinking water safety, these indices provide additional measures of sediment contaminant levels.

Concentrations of most metals in sediments were below the guidelines for the protection of aquatic life (Table 6). Arsenic was the only metal to exceed the ISQG (Reach 3 and Reach 4B both exceeded the ISQG of 5.9 mg/kg; Table 6), but all levels remained below the PEL of 17.0 mg/kg. Arsenic levels in all reaches were similar to (and slightly lower than) those observed in previous sample years in most reaches

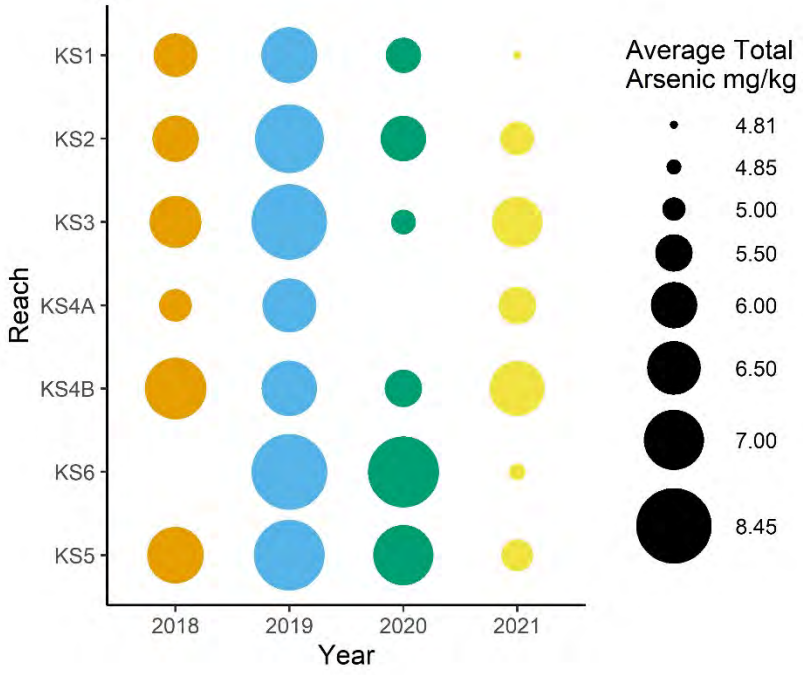

Figure 12. Bubble plot of total arsenic concentrations in sediment samples from the Slave River, with size-scaled bubbles representing the average concentration for each reach in each year. Reaches are arranged in order from upstream (top) to downstream (bottom). Variability in size from top to bottom indicates spatial variability among reaches, and variability from left to right indices of the degree of change over time in the Slave River. No bubble is shown for reaches that weren't sampled in a particular year.

(Figure 12), indicating that the exceedances did not reflect an increase from the previous years. These exceedances may relate to the presence of bitumen seeps from cliffs above Reach 3. Arsenic concentrations in sediment have exceeded the ISQG in at least one reach in each year of sampling, with the highest concentrations in 2019 (Figure 12). However, exceedances in all cases have been minor, and concentrations have not approached the PEL. Other metals in sediment samples from 2021 were below the ISQG and PEL or did not have guidelines.

Average concentrations for PAHs in sediments were generally low in Slave River reaches, and many PAHs were below detection limits (Table 6). However, concentrations of 2-methylnaphthalene were elevated above the ISQG in Reach 3 and Reach 6. 2-methylnaphthalene is an LMW-PAH, thus representing an acutely toxic species for benthic organisms (Canadian Council of Ministers of the Environment 1999). Concentrations of the PAH have exceeded the ISQG in several sample years; however, these exceedances may not represent levels that are high enough to do harm, as they are somewhat minor exceedances of the lower, interim guidelines.

Other measures of PAHs were generally low or below guidelines. For example, phenanthrene exceeded guidelines in 2020, but fell below the ISQG in 2021. Some HMW-PAHs (carcinogenic compounds) that were found to exceed ISQGs in previous years (e.g., chrysene) were below guidelines in 2021. The BaP Total Potency Equivalent, which is a measure of cancer risk to humans, was below detection limit in all reaches, and the IACR, which measures threats to drinking water, was below guideline levels in all

Table 7. Summary of biotic metrics for kick-site reaches sampled in the Slave River in 2021, including the mean ± standard deviation for BMI abundance and taxonomic richness metrics. EPT is the sum of Ephemeroptera, Plecoptera, and Trichoptera orders; Chironomidae is a family of Diptera; Diptera + Oligochaeta includes all true flies and segmented worms; and Mollusca includes bivalves (clams) and gastropods (snails). Reaches are ordered from upstream (KS1) to downstream (KS5).

Biotic Metric	SR- KS1	SR-KS2	SR-KS3	SR-KS4A	SR-KS4B	SR-KS6	SR-KS5
Total Abundance	2873 ± 681	2203 ± 1240	1738 ± 565	908 ± 499	1210 ± 873	1126 ± 425	1264 ± 676
EPT abundance	2548 ± 863	1780 ± 1054	1553 ± 522	694 ± 511	667 ± 620	935 ± 365	927 ± 587
Chironomidae abundance	19 ± 19	137 ± 131	88 ± 41	60 ± 44	27 ± 22	66 ± 26	42 ± 46
Diptera + Oligochaeta abundance	51 ± 55	168 ± 154	115 ± 36	67 ± 42	33 ± 23	78 ± 28	61 ± 53
Hydra abundance	255 ± 267	238 ± 186	50 ± 42	144 ± 74	503 ± 405	102 ± 88	273 ± 95
Percent EPT	87.4 ± 11.1	76.7 ± 11.1	89.0 ± 2.1	68.7 ± 21.3	51.3 ± 20.2	82.4 ± 4.9	70.3 ± 9.0
Percent Chironomidae	0.7 ± 0.6	8.4 ± 7.3	5.5 ± 2.4	6.4 ± 1.7	2.3 ± 1.0	6.4 ± 3.3	3.2 ± 2.5
Percent Diptera + Oligochaeta	1.8 ± 1.7	11.3 ± 11.2	6.8 ± 1.6	7.5 ± 1.7	2.9 ± 1.1	7.5 ± 3.4	4.7 ± 2.7
Percent *Hydra*	10.2 ± 10.9	11.1 ± 7.3	3.0 ± 2.4	23.7 ± 21.1	45.1 ± 20.7	9.0 ± 6.6	24.8 ± 9.3
Taxonomic Richness	10.4 ± 2.1	18.0 ± 5.8	16.8 ± 2.4	17.0 ± 3.6	15.6 ± 1.5	17.8 ± 1.1	15.6 ± 3.9
Richness of EPT	5.2 ± 0.4	6.6 ± 1.1	6.6 ± 0.5	5.0 ± 0.0	7.4 ± 0.9	7.2 ± 0.4	6.2 ± 0.4
Richness of Chironomidae	1.4 ± 1.1	5.2 ± 2.3	5.2 ± 2.0	7.7 ± 2.1	4.0 ± 1.2	6.0 ± 1.4	4.0 ± 2.1
Richness of Diptera + Oligochaeta	2.6 ± 2.3	8.4 ± 4.6	7.2 ± 2.2	10.0 ± 3.5	5.8 ± 0.8	7.8 ± 0.8	7.6 ± 3.6

reaches (Canadian Council of Ministers of the Environment 2010). Overall, sediment chemistry values in 2021 were generally similar to or less than in previous years, which did not indicate any potential concerns.

3.2.2. Spatial variation in benthic macroinvertebrates

3.2.2.1. Biotic metric variation

Biotic metrics were used to compare abundance, relative abundance, and taxonomic richness of key BMI taxonomic groups among sites and reaches in the Slave River. Whereas 2020 was an unusual year, with high water levels and a high total abundance of BMI at sites that was primarily driven by extremely high abundances of the freshwater cnidarian *Hydra*, 2021 marked a return to more usual water levels and more typical abundances of BMI taxa including *Hydra*. Total abundance ranged from 908 to 2973 individuals on average per reach, with moderate variability within reaches (Table 7). EPT abundance was high across all reaches, and EPT taxa made up > 50% on average of total BMI abundance in all reaches, with three reaches having an average relative abundance of EPT of > 80% (Table 7). In contrast, Chironomidae abundance remained low across reaches, and made up less than 10% of total abundance on average (as little as 0.7%, in Reach 1). Abundances of *Hydra* were low in most reaches, varying from 3% to 24.8% of the total abundance of BMI on average, with the exception of Reach 4B, where *Hydra*

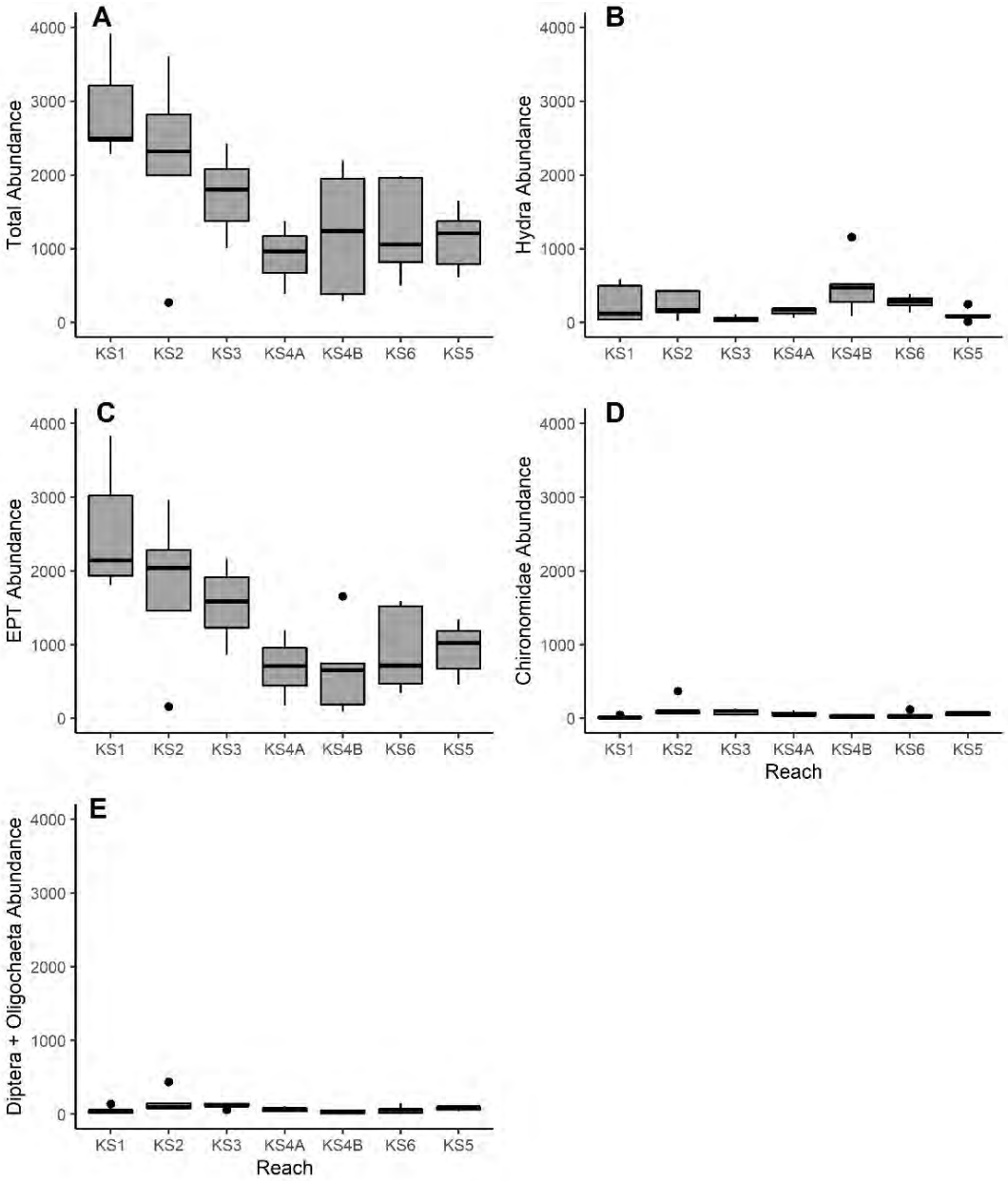

Figure 13. Box plots of abundance BMI metrics for the Slave River reaches sampled in 2021, including (A) total abundance, (B) abundance of Hydra*, (C) abundance of Ephemeroptera, Plecoptera, and Trichoptera (EPT), (D) abundance of Chironomidae (midges), (E) abundance of Diptera (true flies) + Oligochaeta (segmented worms). Box indicates the interquartile range, line through the box indicates the median, and whiskers indicate the range of data outside the lower and upper quartiles (1.5*interquartile range). Points indicate statistical outliers. Reaches are ordered from upstream (KS1) to downstream (KS5).*

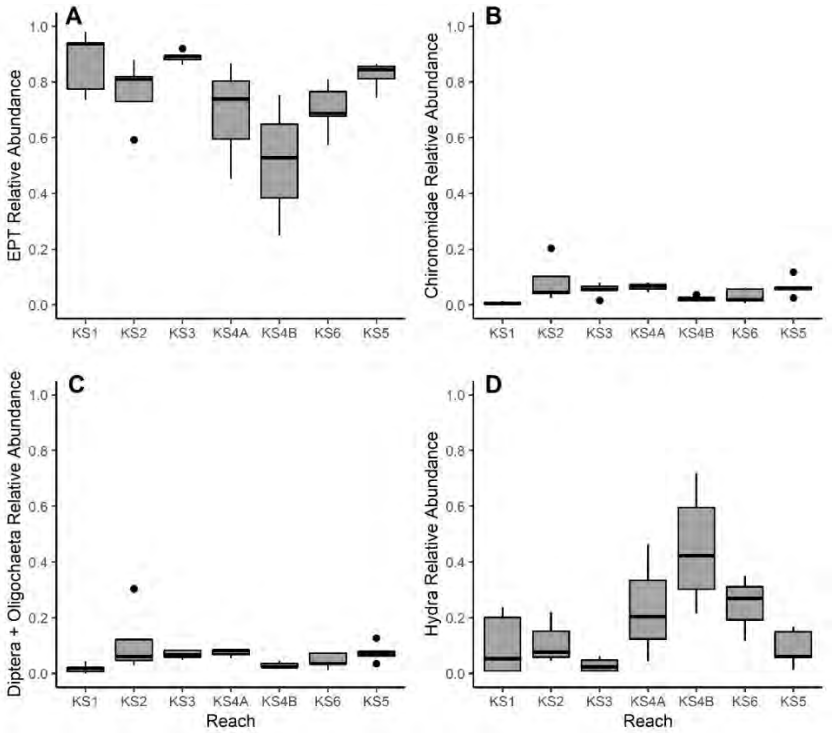

Figure 14. Box plots of relative abundance metrics for the Slave River reaches sampled in 2021, including (A) EPT, (B) Chironomidae (midges), (C) Diptera (true flies) + Oligochaeta (segmented worms), and (D) Hydra*. Box indicates the interquartile range, line through the box indicates the median, and whiskers indicate the range of data outside the lower and upper quartiles (1.5*interquartile range). Points indicate statistical outliers. Reaches are ordered from upstream (KS1) to downstream (KS5).*

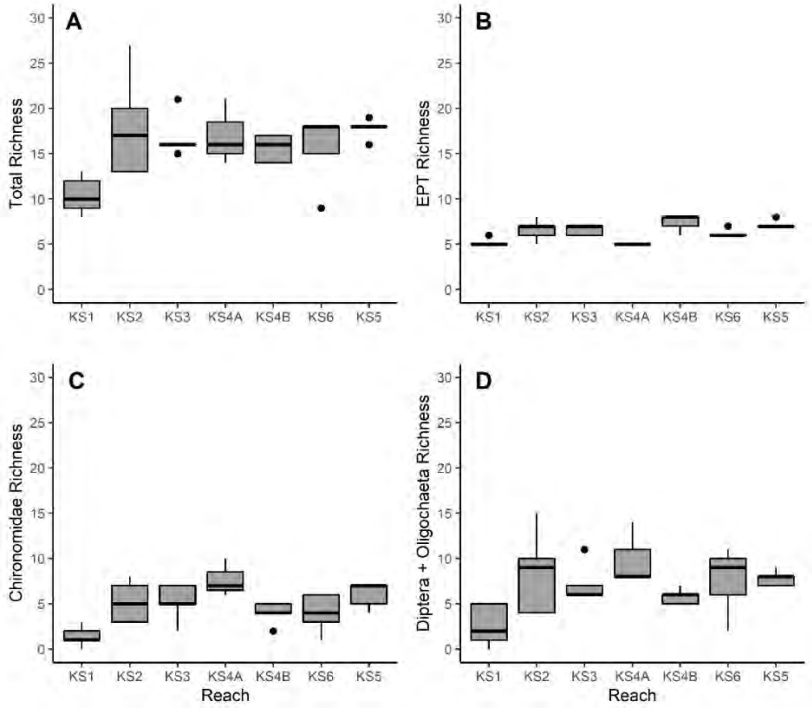

*Figure 15. Box plots of richness BMI metrics for the Slave River reaches sampled in 2021, including (A) total richness, and richness of (B) EPT, (C) Chironomidae (midges), and (D) Diptera (true flies) + Oligochaeta (segmented worms). Box indicates interquartile range, line indicates the median, and whiskers indicate the range of data outside the lower and upper quartiles (1.5*interquartile range). Points indicate statistical outliers. Reaches are ordered from upstream (KS1) to downstream (KS5).*

were more prevalent (relative abundance of 45.1%; Table 7). While *Hydra* was necessarily a focus of the 2020 assessment year to its dominance across most reaches, it was only found in high numbers in 2021 in a reach that has had a high relative abundance of *Hydra* across all sample years (Lento 2022).

Total abundance of BMI declined from Reach 1 through Reach 4A, and but remained similar on average across the downstream reaches (Figure 13). This pattern appeared to be entirely driven by the abundance of EPT taxa across all reaches. EPT taxa dominated assemblages across all reaches, though relative abundances of these taxa were more variable in Reach 4A and Reach 4B (Figure 14). Given the high mobility of many EPT taxa (both active mobility and passive, as part of the drift) and their ability to colonize new habitats, the predominance of this group may have partially reflected an ongoing shift in the benthic communities of these systems following the high water conditions in the previous year.

Taxonomic richness was lower in the Slave River in 2021 than in previous years, ranging from 10 to 18 taxa on average per reach (a decline from 2020, when richness ranged from 12 to 21 taxa on average) (Table 7; Figure 15). The lowest richness was in Reach 1, reflecting a lower taxonomic richness of both Chironomidae and EPT relative to other reaches (Figure 15). This suggests that despite their numerical abundance at Reach 1, EPT assemblages in that reach were predominantly composed of a small number of taxa. Total taxonomic richness and richness of each taxonomic group were similar on average across all other reaches (Figure 15).

3.2.2.2. Multivariate assessment of BMI assemblage composition

Multivariate analysis was used to characterize the biotic assemblage of the Slave River and evaluate similarities and differences in assemblage composition among reaches and sites. PCA was intended to assess correlations within and among reaches and identify the taxa driving compositional differences among sites, whereas PERMANOVA and homogeneity of multivariate dispersions assessed similarity in composition among and within reaches. BMI relative abundance data for all taxa were assessed at the family/subfamily level.

The PCA of BMI data identified two clear and orthogonal gradients among sites that were driven by a small number of dominant taxa. Sites in Reach 2, Reach 3, and Reach 5 were tightly clustered together and differed from sites in other reaches along the first axis of the PCA, which explained 61.6% of variability in assemblage structure (Figure 16A). At the other end of the gradient, sites in Reach 1 were positively associated with axis I and axis II (the latter of which explained 32.2% of variability in assemblage structure), and were orthogonal to sites in Reach 4B and Reach 6 (Figure 16A). Sites in Reach 4B and Reach 6 (as well as a site from Reach 4A) were primarily positively associated with *Hydra* (Figure 16B). In contrast, sites in Reach 1 were positively associated with the caddisfly Hydropsychidae, the mayflies Heptageniidae and Ephemerellidae, and the stonefly Perlodidae, all of which have adaptations for fast flows, and all of which were also found to be strongly associated with that reach in 2020. Most other taxa were associated with the cluster of remaining sites (predominantly Reach 2, Reach 3, and Reach 5), indicating that there was similarity among the remaining sites with few taxa driving particularly strong gradients in composition.

Based on a dissimilarity matrix of all sites, the PERMANOVA indicated that there were significant differences in assemblage composition among reaches in the Slave River ($F = 7.13$, $p = 0.001$). Pairwise PERMANOVA was used to identify which reaches had statistically significant differences in assemblages (significant at an FDR-corrected α-level, based on the rank of each p-value). The pairwise PERMANOVA

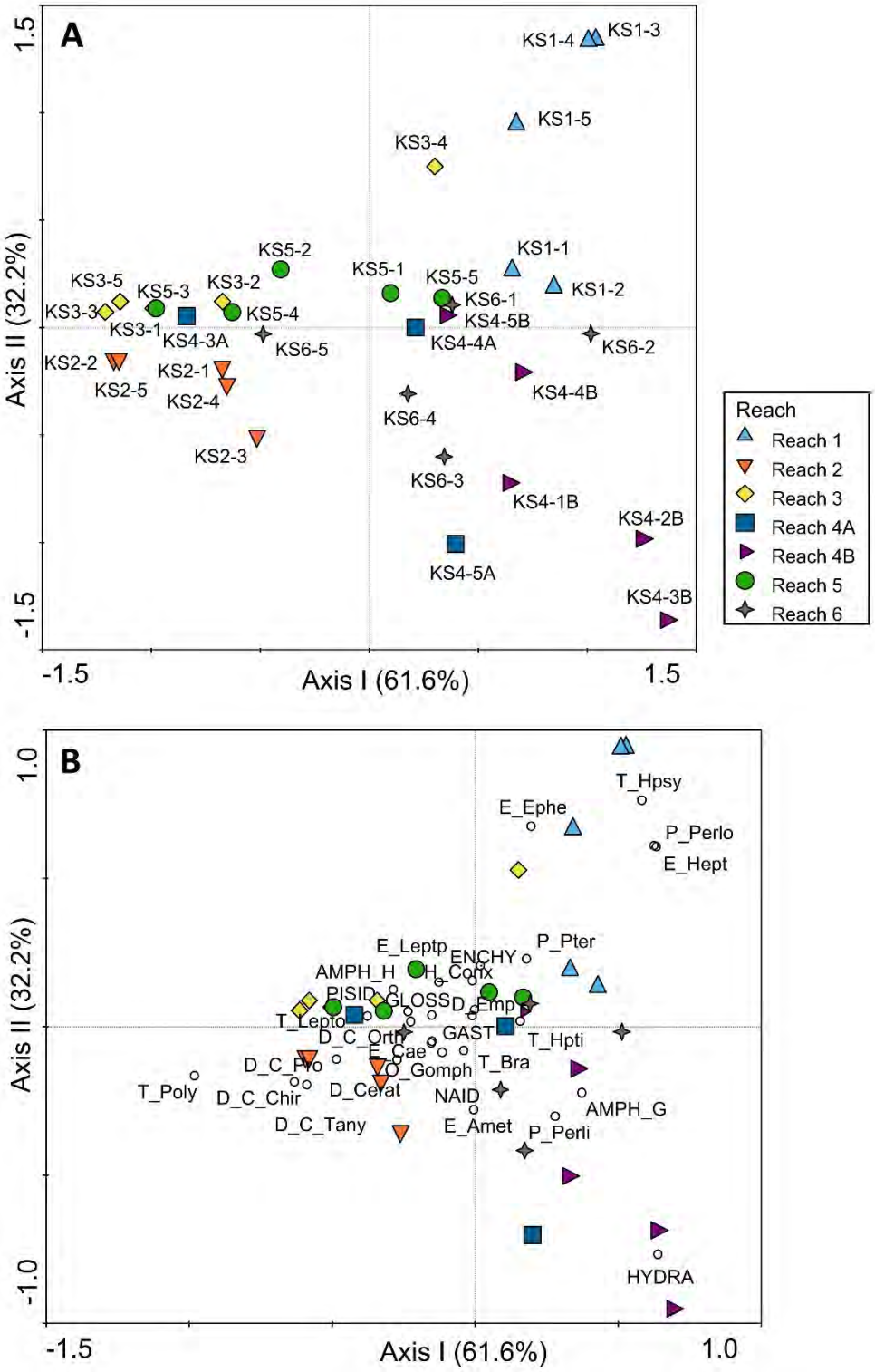

Figure 16. PCA ordination of BMI samples from kick samples in the Slave River in 2021, with sample points coloured by reach, sample points labelled in (A) and taxa points labelled in (B). Kick-sites in close proximity have similar assemblages, whereas samples on opposite ends of gradients have differences in their assemblages. Samples at right angles through the origin are uncorrelated. Kick-sites are located close to taxa with which they are positively correlated and opposite those with which they are negatively correlated. Taxonomic abbreviations are listed in the appendices.

Table 8. Results of pairwise PERMANOVA comparing assemblage dissimilarity among reaches of the Slave River, showing pairwise p*-values for each comparison. FDR-corrected α was calculated for each pairwise comparison based on* p*-value rank, and pairwise comparisons that were significant at the FDR-corrected level are indicated in bold.*

	KS1	KS2	KS3	KS4A	KS4B	KS5	KS6

KS1							
KS2	**0.009**						
KS3	**0.007**	0.094					
KS4A	**0.016**	0.105	0.162				
KS4B	**0.008**	**0.008**	**0.009**	0.133			
KS5	**0.01**	0.044	0.303	0.212	0.005		
KS6	**0.011**	**0.006**	**0.032**	0.248	0.11	0.118	

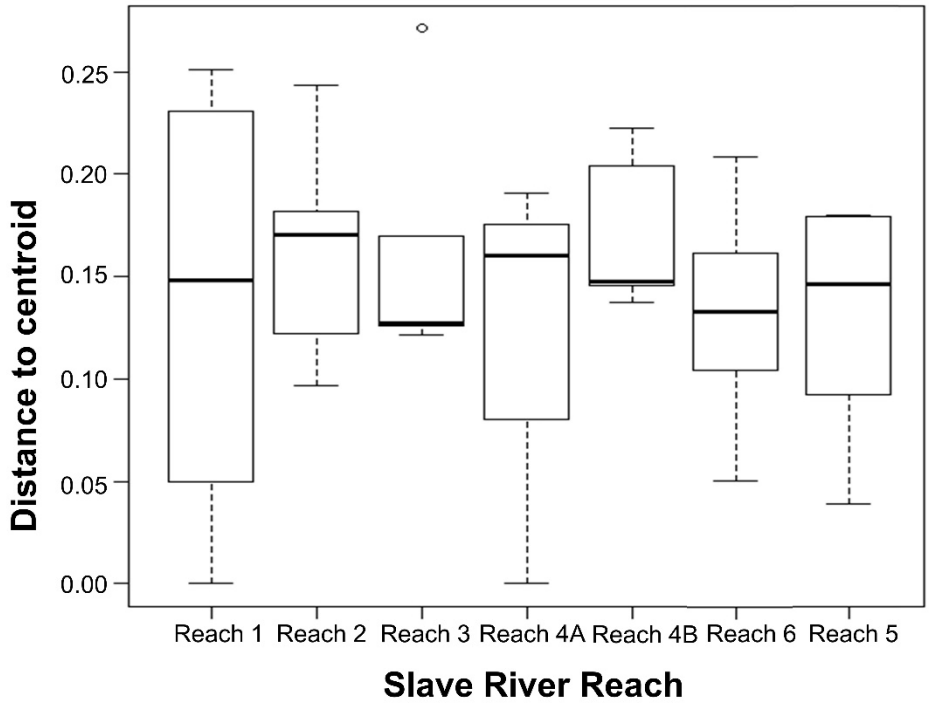

Figure 17. Results of homogeneity of multivariate dispersions analysis of Slave River BMI assemblages for reaches in 2021, showing the median distance to centroid for each reach (black bar), 25th and 75th percentiles (lower and upper bounds of box, respectively), minimum and maximum (whiskers), and outliers (points). Distance to centroid represents the spread of sites in multivariate space, where greater distance equals greater dissimilarity among sites. Low distance to centroid indicates similarity within reaches.

largely confirmed the patterns that were evident in the PCA, as it indicated that Reach 1 differed significantly from all other reaches, and that Reach 4B and Reach 6 differed significantly from Reach 1, Reach 2, and Reach 3 (Table 8). Though the lowest *p*-value was found for the comparison between Reach 4B and Reach 5, it was not significant at the FDR-corrected α level. These results provide further statistical support for the compositional differences of Reach 1, Reach 4B, and Reach 6 that were apparent in the PCA ordination.

Within-reach variability in assemblage composition was similar across all reaches of the Slave River (homogeneity of multivariate dispersions $F = 0.3867$, $p = 0.87$). The median distance to centroid was similarly low for all reaches (Figure 17). Distance to centroid was highly variable in Reach 1, whereas Reach 3 had low variability in the distance to centroid among sites, albeit with one statistical outlier. Sites in Reach 1 were dominated by a small number of abundant taxa, and differences in composition outside of that core group of taxa may have contributed to a greater distance to centroid for some sites.

3.2.3. Temporal characterization of BMI assemblages

3.2.3.1. Benthic macroinvertebrate composition

Examination of temporal patterns in BMI composition provides a means to understand shifts that have occurred in response to the variable flow conditions across the sampling period. Sampling in 2020

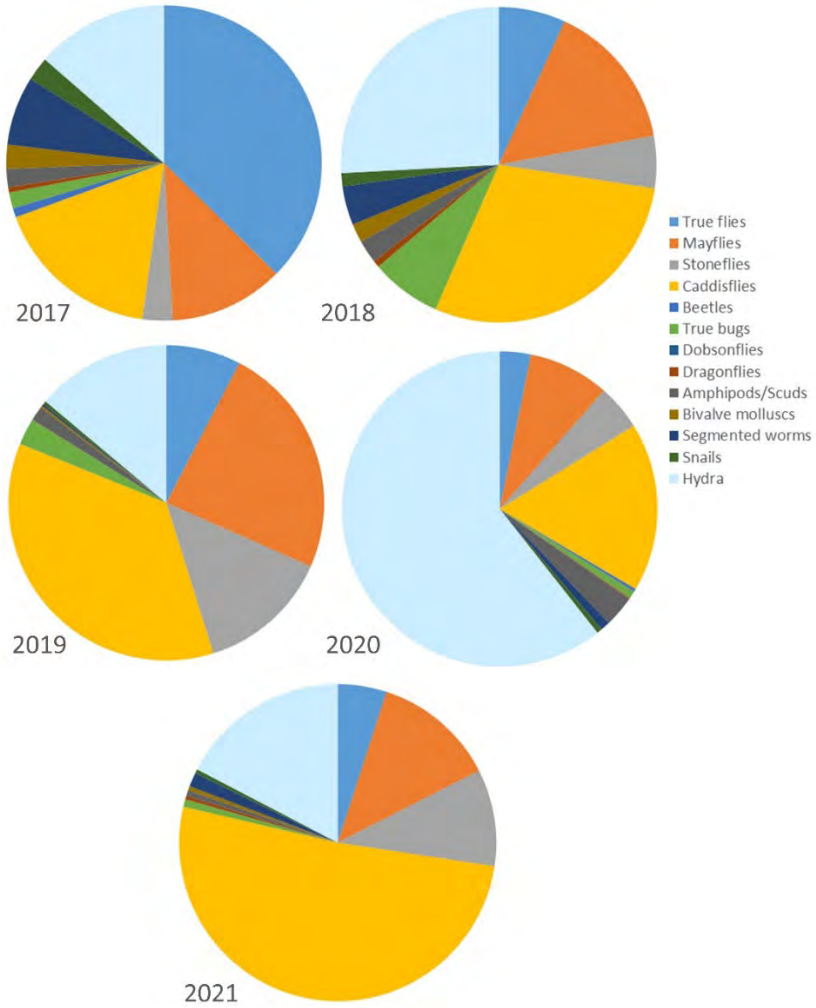

Figure 18. Average relative abundance of major taxonomic groups in Slave River kick samples collected in 2017, 2018, 2019, 2020, and 2021. Taxa are grouped as true flies (Diptera), mayflies (Ephemeroptera), stoneflies (Plecoptera), caddisflies (Trichoptera), beetles (Coleoptera), true bugs (Hemiptera), dobsonflies (Megaloptera), dragonflies (Odonata), amphipods/scuds (Amphipoda), bivalve molluscs (Bivalvia), segmented worms (Oligochaeta), snails (Gastropoda), and Hydra.

offered the opportunity to better understand the assemblage response to extreme high flow conditions, while data from 2021 can illustrate the recovery trajectory, as assemblages shift in response to more typical conditions. Compositional changes from 2017 to 2021 were summarized at the river level for the Slave River by assessing the average relative abundance of major taxonomic groups across all reaches in each year. Most striking is the increase in relative abundance of *Hydra* that was observed in 2020 (Figure 18). But with the addition of data for 2021, there is evidence of the general composition of the river returning to match that observed pre-2020. The relative abundances of true flies, mayflies, stoneflies, and *Hydra* in 2021 were more similar to those observed in 2018 and 2019, whereas the high abundance of caddisflies was similar to the pattern in 2019. Water levels in 2019 were higher than 2017-2018, and the antecedent conditions from 60 days prior to sampling were similar between 2019 and 2021.

Therefore, the similarity in caddisfly composition between 2019 and 2021 may have been due to similarly higher water levels in the period prior to sampling. Elevated abundance of caddisflies in 2021 may reflect a resiliency of this group under high flow conditions, particularly for those taxa attached to the bottom substrate with cases or nets. For example, net-spinning caddisflies (Hydropsychidae, which were abundant in Reach 1) have been shown to be resilient to high flows and to adjust the structure of their net to support the added stress of higher velocities (Loudon and Alstad 1992). These taxa have also been found to help stabilize gravel substrates under high flow conditions through the construction of their nets (Johnson et al. 2009), and it has been suggested that they are highly resilient to variability in flows and suspended sediment loads (Albertson and Daniels 2016). These and other caddisfly taxa may play a large role in the recovery of Slave River benthic assemblages following high flow conditions.

Bubble plots were used for visualization of broad-scale differences in composition among reaches and between years. Comparison of plots for the relative abundance of EPT, Chironomidae, and *Hydra* across the sampling period provides further evidence of the sharp change in composition observed in 2020 and the apparent recovery in 2021 to a composition that more closely resembled that in 2019 (Figure 19). The relative abundance of EPT varied among reaches in 2017 and 2018, but appeared consistently high among all reaches in 2019, likely reflecting the increased abundance of caddisflies in that year. Relative abundance of EPT declined in most reaches (with the exception of Reach 1) with the increase in *Hydra* in 2020, but returned to high levels across all reaches in 2021 (Figure 19A). Temporal patterns of relative abundance of Chironomidae primarily reflected the dominance of this group in 2017 followed by low abundances in all years since (Figure 19B). In 2018 and 2019, Chironomidae appeared to be most abundant in Reach 2, Reach 4A, and Reach 5. In 2021, these three reaches with the addition of Reach 3 had the highest relative abundances of this group, which may have contributed to the similarity of these reaches in the multivariate analysis. For *Hydra*, the highest relative abundance across all reaches was in 2020, although the plot indicated the consistently high relative abundance of this taxon at Reach 4B across all sampling years (Figure 19C). The consistently high relative abundance of *Hydra* in Reach 4B suggests that this reach contains ideal habitat for *Hydra*. This plot also highlighted the higher abundance of this taxon in Reach 4A, Reach 4B, and Reach 6 in 2021, which helped to drive differences among sites in the multivariate analysis.

Total taxonomic richness was highest in all reaches in 2017, and appears to have declined since then (Figure 20). In part, the decline in abundance reflected the decline in abundance and diversity of Chironomidae that occurred in 2018. Furthermore, the decline in richness in 2020 was suggested to be a result of the predominance of *Hydra* across sites, and a suggested impact of the high flows in that year. However, the bubble plot indicated that richness was again lower in 2021 (particularly in Reach 1). This

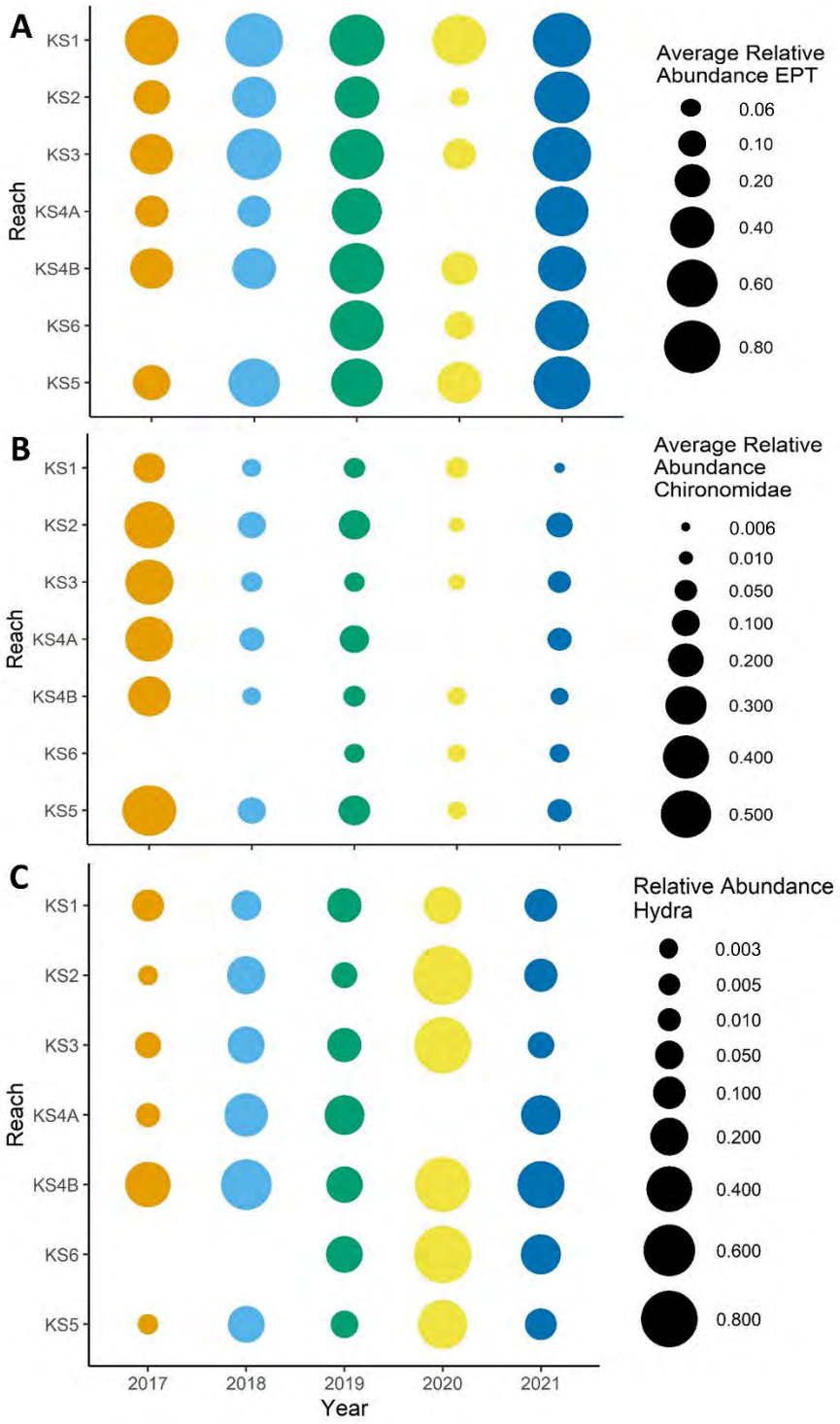

Figure 19. Bubble plots of BMI metrics for the Slave River, including relative abundance of (A) EPT, (B) Chironomidae (midges), and (C) Hydra*, with size-scaled bubbles representing the average relative abundance for each reach in each year. Reaches are arranged in order from upstream (top) to downstream (bottom). Variability in size from top to bottom indicates spatial variability among reaches, and variability from left to right indices of the degree of change over time in the Slave River. No bubble is shown for reaches that weren't sampled in a particular year.*

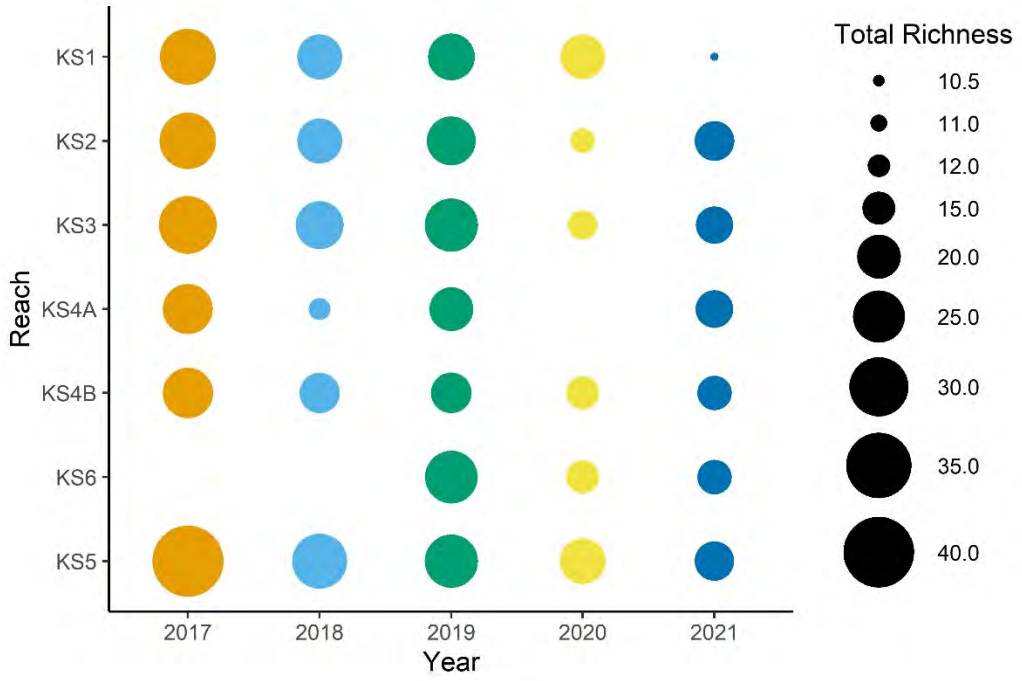

Figure 20. Bubble plot of total richness of BMI for the Slave River, with size-scaled bubbles representing the average richness for each reach in each year. Reaches are arranged in order from upstream (top) to downstream (bottom). Variability in size from top to bottom indicates spatial variability among reaches, and variability from left to right indices of the degree of change over time in the Slave River. No bubble is shown for reaches that weren't sampled in a particular year.

may reflect the fact that reaches are still in recovery following the high water levels in 2020 (which continued into winter and early spring of 2021).

The reach mean ± SE for each biotic metric in each year was plotted with data for all reaches overlain in single plots to further evaluate change over time in the Slave River (Figure 21). These plots again highlight the fact that 2020 was not a typical year, and BMI assemblages responded to the extreme high flow conditions. For example, total abundance increased dramatically in 2020 due to high abundances of *Hydra*, but declined again in 2021 to levels more typical of these reaches, albeit somewhat higher than previous years (Figure 21A). The increase in total abundance in 2021 was driven by increased abundance of EPT relative to previous years, particularly in Reach 1 and Reach 2 (Figure 21C). In contrast, abundances of *Hydra* and Chironomidae appeared similar to sample years prior to 2020 (Figure 21B, D).

Relative abundance metrics highlighted the return to pre-2020 levels for the relative abundance of EPT and relative abundance of *Hydra* in particular (Figure 22). EPT relative abundance in 2021 most closely resembled that observed in 2019, though relative abundance was higher in most reaches in 2021 (Figure 22A). Most reaches showed the same pattern over time for the relative abundance of Chironomidae, with a sharp decrease in 2018 and similar values from 2018-2021 (Figure 22C). Patterns over time in the relative abundance of Diptera + Oligochaeta were similar to those of Chironomidae, but most noticeable was a decline in Reach 2 in 2020 and 2021 relative to that observed in 2019, which suggests a loss of non-midge true flies and segmented worms in this reach since high flow conditions began (Figure 22D).

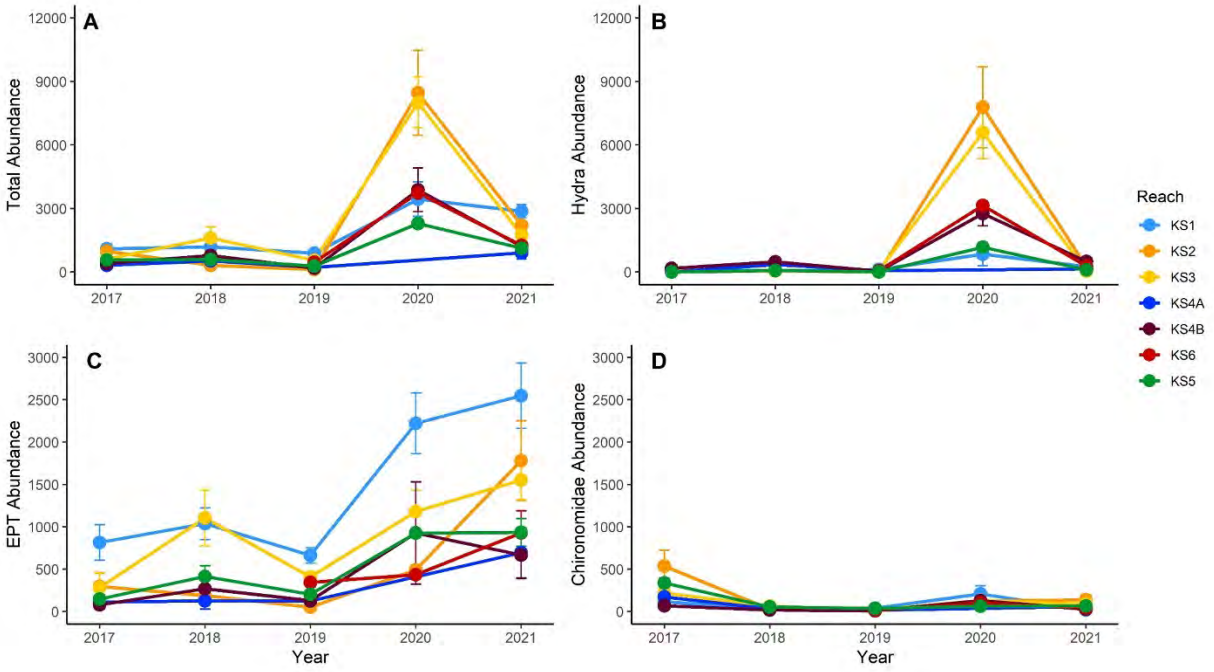

Figure 21. Line plots of changes over time in (A) total abundance, and abundance of (B) Hydra*, (C) Ephemeroptera, Plecoptera, Trichoptera (EPT), and (D) Chironomidae (midges) in the Slave River, showing the mean ± SE for each reach in 2017, 2018, 2019, 2020, and 2021.*

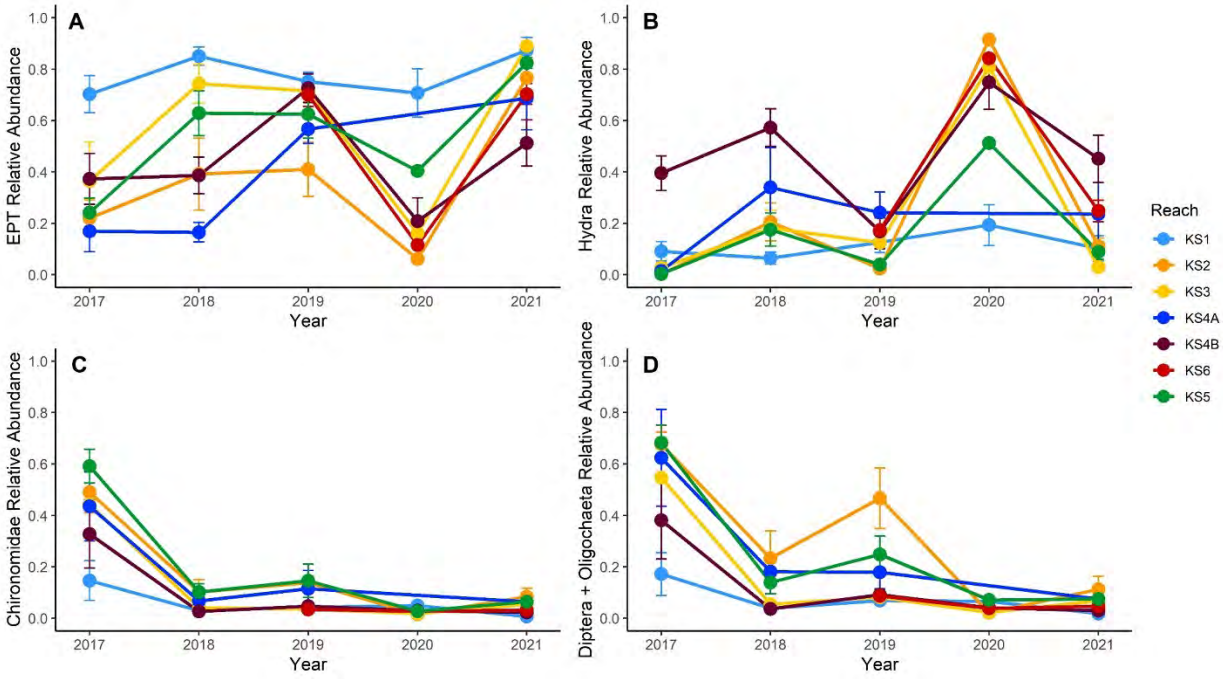

Figure 22. Line plots of changes over time in relative abundance of (A) Ephemeroptera, Plecoptera, and Trichoptera (EPT), (B) Hydra*, (C) Chironomidae (midges), and (D) Diptera + Oligochaeta (true flies and segmented worms) in the Slave River, showing the mean ± SE for each reach in 2017, 2018, 2019, 2020, and 2021.*

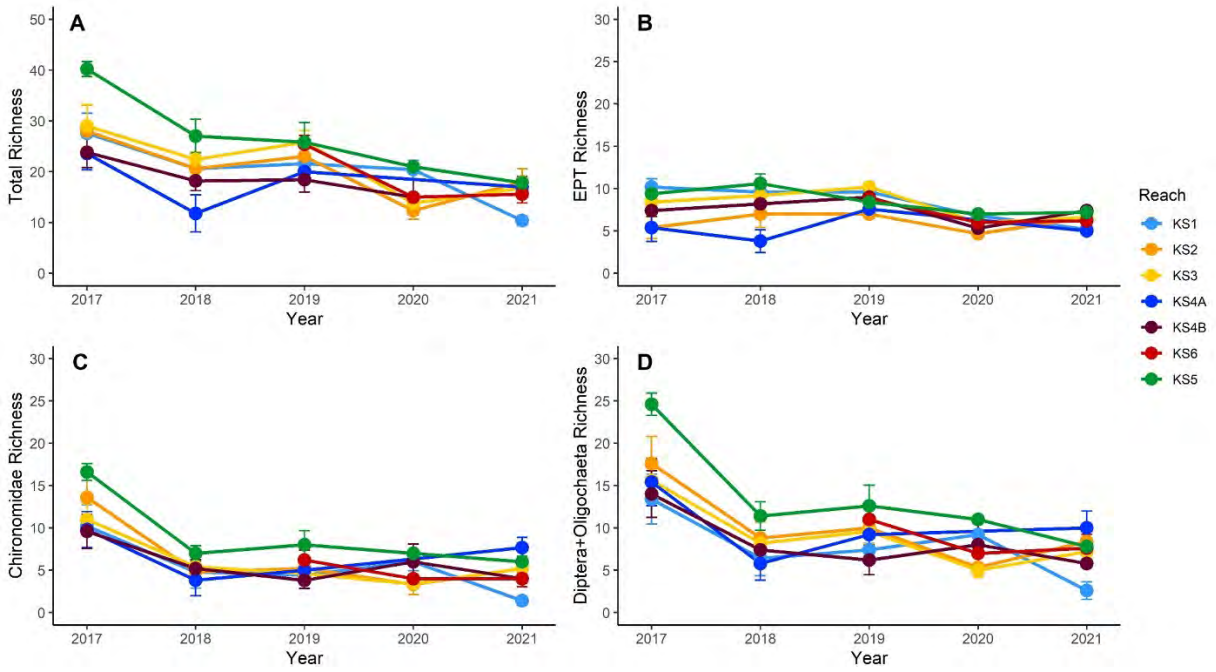

Figure 23. Line plots of changes over time in (A) total taxonomic richness and richness of (B) Ephemeroptera, Plecoptera, and Trichoptera (EPT), (C) Chironomidae (midges), and (D) Diptera + Oligochaeta (true flies and segmented worms) in the Slave River, showing the mean ± SE for each reach in 2017, 2018, 2019, 2020, and 2021.

Richness metrics showed similar trends over time in most reaches, with a decline from 2017-2018 and either stable or somewhat declining patterns from 2018-2021 (Figure 23). The decline in total richness appeared to be due to loss of Chironomidae taxa and Diptera + Oligochaeta taxa across most sites, and these taxa notably contributed to the decline in richness in Reach 1 that was observed in 2021 (Figure 23). This pattern suggests that the increased abundance of EPT taxa in Reach 1 was accompanied by a loss of diversity of non-EPT taxa in this reach. Figure 21 indicated that total abundance declined slightly on average in this reach, as there was both an increased in abundance of EPT and a decrease in abundance of Chironomidae. Total richness was also lower in 2021 for most other reaches, notably for Reach 5, which showed the most steady decline in richness since 2017. Together, these patterns point to a continued loss of diversity in these reaches, despite the return of water levels to more typical conditions following the freshet in 2021. Continued monitoring of this pattern is necessary to determine whether richness begins to increase again as more time passes since the extreme flow conditions of 2020, or whether this is indicative of a long-term trend in declining diversity in this river.

3.3. Normal range and CES for BMI metrics

Variation among samples was used to create an initial estimate of the normal range of variability and set preliminary CES boundaries to trigger additional monitoring or management action if test samples are impaired (i.e., if they fall outside the range of natural variability). The normal range is commonly defined as the range within which 95% of samples fall, equivalent to two standard deviations from the mean in a normal distribution (Munkittrick et al. 2009). While it is possible for samples to fall outside the CES,

there is a low probability (5% chance) of this happening if the sample is representative of the normal range. Thus, where sites have been exposed to anthropogenic impacts, samples outside of the CES may be an indication of impairment in a system. The normal range and CES are based on variability in the data, and changes in habitat conditions that result from natural variability (i.e., due to shifts in flow, timing of the spring freshet, water temperature, etc.) may affect the normal range from one year to the next. This idea is particularly relevant to the Slave River, where significant changes occurred between sampling years due to changes in the flow regime. With additional sampling, sites that were within the normal range in one year may fall outside the normal range in the next year if they are strongly affected by natural variability in the system. Monitoring of assemblages over several years and refining normal range estimates should therefore be used to get a better, more accurate estimate of the CES in a system that accounts for this natural variability. But where extreme conditions occur, such as the high flow in the Slave River in 2020, the noise that this introduces to estimates of normal range can be reduced by excluding particular years from the calculation of normal range.

Initial normal range estimates were developed by Lento (2021) using data from 2017 to 2019 and refined by Lento (2022) using data from 2017-2020. In this report, these normal range estimates are revisited using the period 2017-2021, and critically evaluated to identify years that may represent noise and that do not represent the typical variability in the system. Creating reliable CES estimates requires a strong set of baseline data with clear patterns over time, and these patterns can be difficult to detect if there is strong variability among years. Given how much flow conditions differed in 2020 and how variable the BMI assemblage was, it may be necessary to consider normal range estimates that exclude 2020.

Originally, quantification of the normal range of variability (within-year and temporal) and critical effect size was based on the BMI metrics total abundance, relative abundance of EPT, relative abundance of Chironomidae, relative abundance of Diptera + Oligchaeta, total taxonomic richness, richness of EPT, richness of Chironomidae, and richness of Diptera + Oligochaeta. However, the high abundances of *Hydra* in 2020 prompted the addition of abundance of *Hydra* and relative abundance of *Hydra* to the analysis. As well, because relative abundance of major taxonomic groups changed so much in 2020 because of *Hydra*, normal range and CES were also estimated for abundance of EPT and abundance of Chironomidae, to determine if these metrics would be less variable over time. Diptera + Oligochaeta metrics were omitted in some cases because they showed patterns that were extremely similar to Chironomidae metrics.

As flow and BMI communities were quite different across sampling years in the Slave River, the normal range calculated by Lento (2022) covered a large range for many metrics, particularly for those that were affected by the high abundance of *Hydra* in 2020. In this report, normal range was calculated using all years of data (2017-2021), but normal range using a subset of years was considered for some metrics, given that compositional data from 2021 indicated a shift to pre-2020 conditions. If the composition in 2020 represented a response to extreme conditions, then the composition in that year should be considered noise and the development of a normal range to describe typical conditions in the river should not include that response. Both sets of normal range plots are presented where applicable to provide the opportunity to refine CES without the influence of data from years that were not reflective of typical composition in the river.

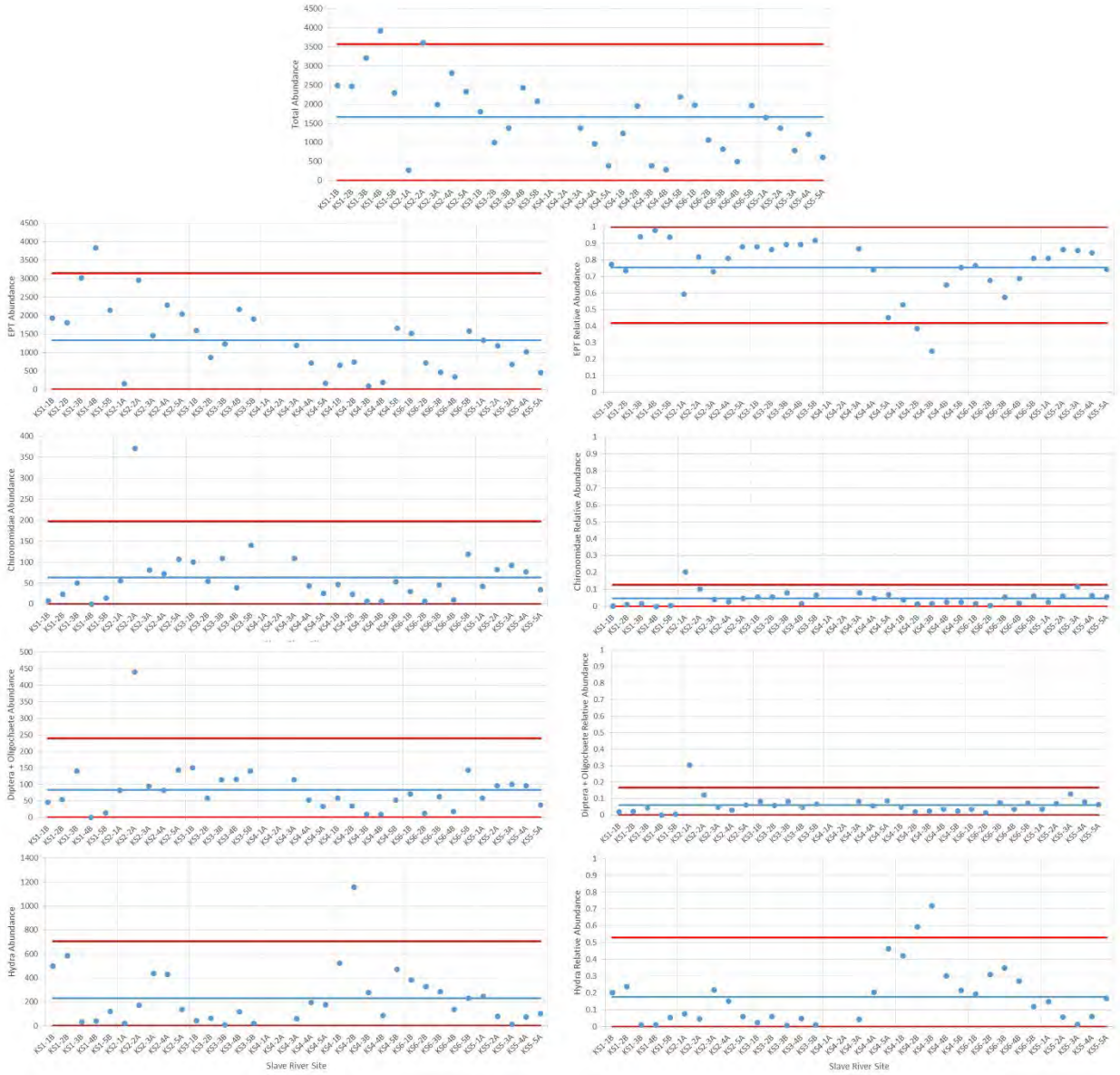

Figure 24. Abundance-based biotic metrics plotted for each site in the Slave River with the 2021 mean (blue line) and the upper and lower single-year critical effect size (CES; red lines), calculated as mean ± 2SD (calculated based on 2021 data). Each point represents a kick-site, moving from reach 1, site 1 (far left) to reach 5, site 5 (far right) on each plot, with gaps for those sites not sampled in 2021. Metrics include (top) total abundance, (left column) EPT abundance, Chironomidae abundance, Diptera + Oligochaeta abundance, Hydra *abundance, and (right column) EPT relative abundance, Chironomidae relative abundance, Diptera + Oligochaeta relative abundance, and* Hydra *relative abundance.*

Figure 25. Richness-based biotic metrics plotted for each site in the Slave River with the 2021 mean (blue line) and the upper and lower single-year critical effect size (CES; red lines), calculated as mean ± 2SD (calculated based on 2021 data). Each point represents a kick-site, moving from reach 1, site 1 (far left) to reach 5, site 5 (far right) on each plot, with gaps for those sites not sampled in 2021. Metrics include (left column) total richness, EPT richness, (right column) Chironomidae richness, and Diptera + Oligochaeta richness.

3.3.1. Within-year variability

For the analysis of within-year variability, mean values of each metric and standard deviations across all sites were used to calculate two sets of within-year CES boundaries: the single-year CES, calculated based on data from 2021, and the multi-year CES, calculated based on the mean and standard deviation of combined 2017-2021 data. Single-year and multi-year CES boundaries were compared with site data from 2021 to evaluate how much the samples collected in 2021 varied amongst each other and relative to all previous years.

Samples collected in the Slave River in 2021 generally fell within the CES boundaries developed using only 2021 data, though there were some exceptions, and boundaries were also large for some metrics due to variability among reaches (Figure 24, Figure 25). For example, total abundance was higher in the upstream reaches than it was in the downstream reaches (with the exception of KS2-1), and CES boundaries ranged from 0 to over 3500 (Figure 24). The pattern in total abundance strongly reflected the abundance of EPT, and site KS1-4 exceeded the upper CES boundary for both total abundance and EPT abundance. In contrast, two sites in Reach 4B fell below the lower CES boundary for the relative abundance of EPT (Figure 24). The normal range was narrower for Chironomidae and Diptera + Oligochaeta metrics, and all but one site fell within CES boundaries (site KS2-2 far exceeded the upper CES boundary for these metrics due to high abundance of Chironomidae relative to other sites). The normal range for *Hydra*, calculated based on 2021 data, was much lower than in 2020, and sites in Reach 4B exceeded the upper boundaries due to higher than average abundance of *Hydra* (Figure 24). Normal range boundaries were narrow for total richness and EPT richness, reflecting low variability in these metrics across sites in 2021 (Figure 25). There was more variability in the richness of Diptera + Oligochaeta, which differed more among sites.

Comparison of 2021 data with the multi-year CES developed based on data from 2017-2021 identified a number of sites that fell outside of the normal range for metrics. In particular, EPT abundance in 2021

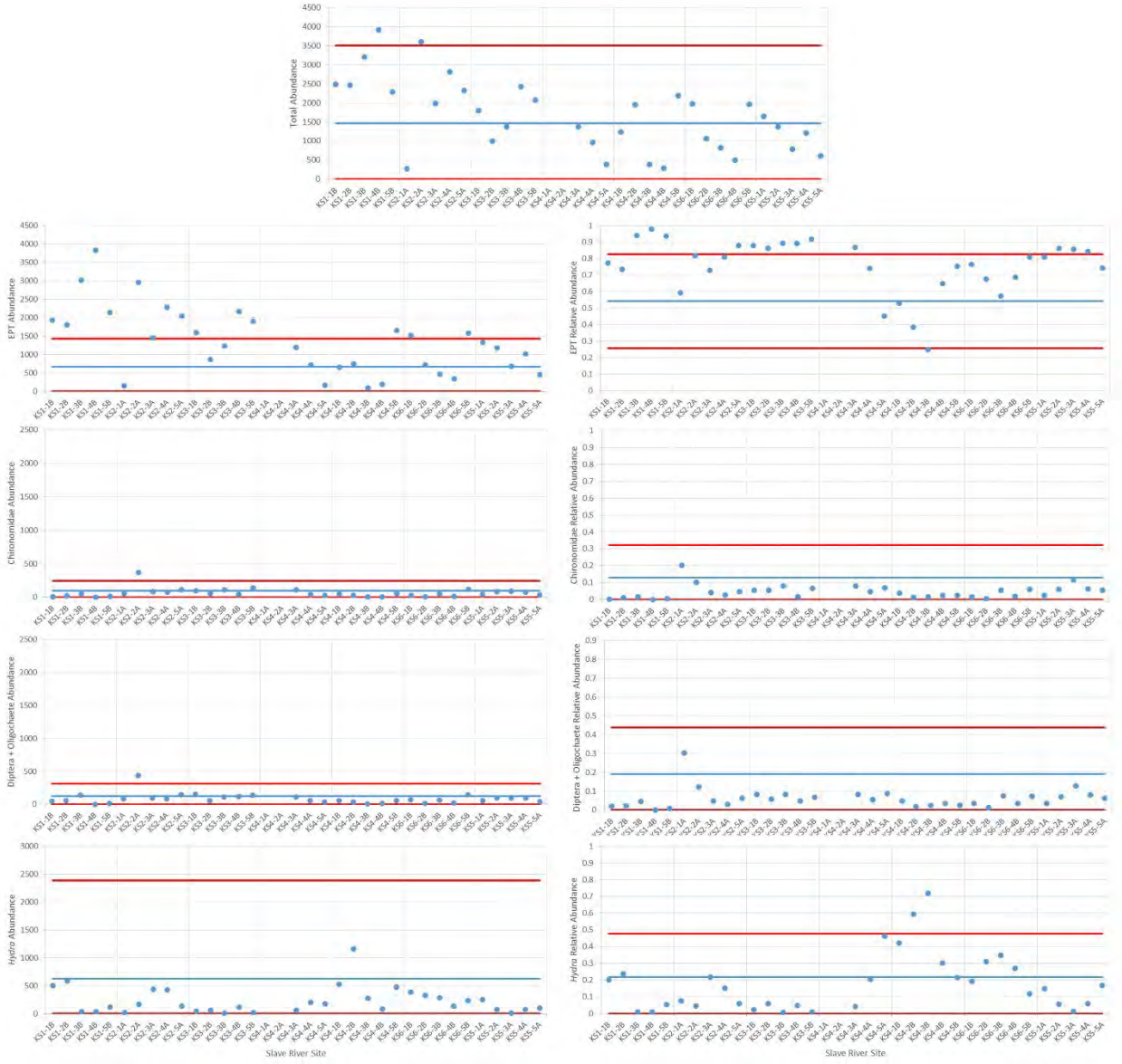

Figure 26. Abundance-based biotic metrics plotted for each site in the Slave River with 2021 data (blue points), a multi-year mean (mean of 2017-2021 data; blue line) and the upper and lower multi-year critical effect size (CES; red lines), calculated as mean ± 2SD (calculated based on 2017, 2018, 2019, 2020, and 2021 data). Each point represents a kick-site, moving from reach 1, site 1 (far left) to reach 6, site 5 (far right) on each plot. Metrics include (top) total abundance, (left column) EPT abundance, Chironomidae abundance, Diptera + Oligochaeta abundance, Hydra *abundance, and (right column) EPT relative abundance, Chironomidae relative abundance, Diptera + Oligochaeta relative abundance, and* Hydra *relative abundance.*

Figure 27. Richness-based biotic metrics plotted for each site in the Slave River with 2021 data (blue points), a multi-year mean (mean of 2017-2021 data; blue line) and the upper and lower multi-year critical effect size (CES; red lines), calculated as mean ± 2SD (calculated based on 2017, 2018, 2019, 2020, and 2021 data). Each point represents a kick-site, moving from reach 1, site 1 (far left) to reach 6, site 5 (far right) on each plot. Metrics include (left column) total richness, EPT richness, (right column) Chironomidae richness, and Diptera + Oligochaeta richness.

was above the upper multi-year CES boundary for 14 out of 33 sites (Figure 26), which highlighted how high EPT abundances in 2021 were compared to previous sample years. Exceedances for this metric were most evident in Reach 1, Reach 2, and Reach 3, but downstream reaches also had sites that exceeded the upper CES. A similar pattern was evident when relative abundance of EPT in 2021 was compared against the multi-year normal range, as 13 sites exceeded the upper multi-year CES boundary, notably including all sites in Reach 3 (Figure 26). Most sites fell within the normal range for Chironomidae and Diptera + Oligochaeta abundance metrics, with the exception of Site KS2-2, which had higher abundance of both sets of taxa. Relative abundance of *Hydra* was higher in Reach 4B than the upper multi-year CES boundary (Figure 26).

Taxonomic richness across all reaches was lower in 2021 than in previous years, and this was evident as a number of sites falling below the lower bounds of the multi-year CES for several richness metrics (Figure 27). Most notably, sites in Reach 1 were at or below the lower multi-year CES boundary for total richness, EPT richness, Chironomidae richness, and Diptera + Oligochaeta richness. Site KS6-2 also had low total richness and richness of Chironomidae and Diptera + Oligochaeta that fell outside the multi-year normal range (Figure 27). Total richness for most sites was below the multi-year average, further indicating that taxonomic richness had declined in 2021.

3.3.2. Among-year variability

3.3.2.1. Site-scale variability

Temporal variation at the site scale was assessed by comparing the 2017-2021 mean ± SE for each site with the normal range for the river, which was calculated as the grand mean for the river (mean of 2017, 2018, 2019, 2020, and 2021 means of all sites) ± 2SD (the standard deviation of annual means). The analysis visualizes temporal variability within sites relative to temporal variability across all sites, and represents one way in which future data may be compared with the expected normal range in this

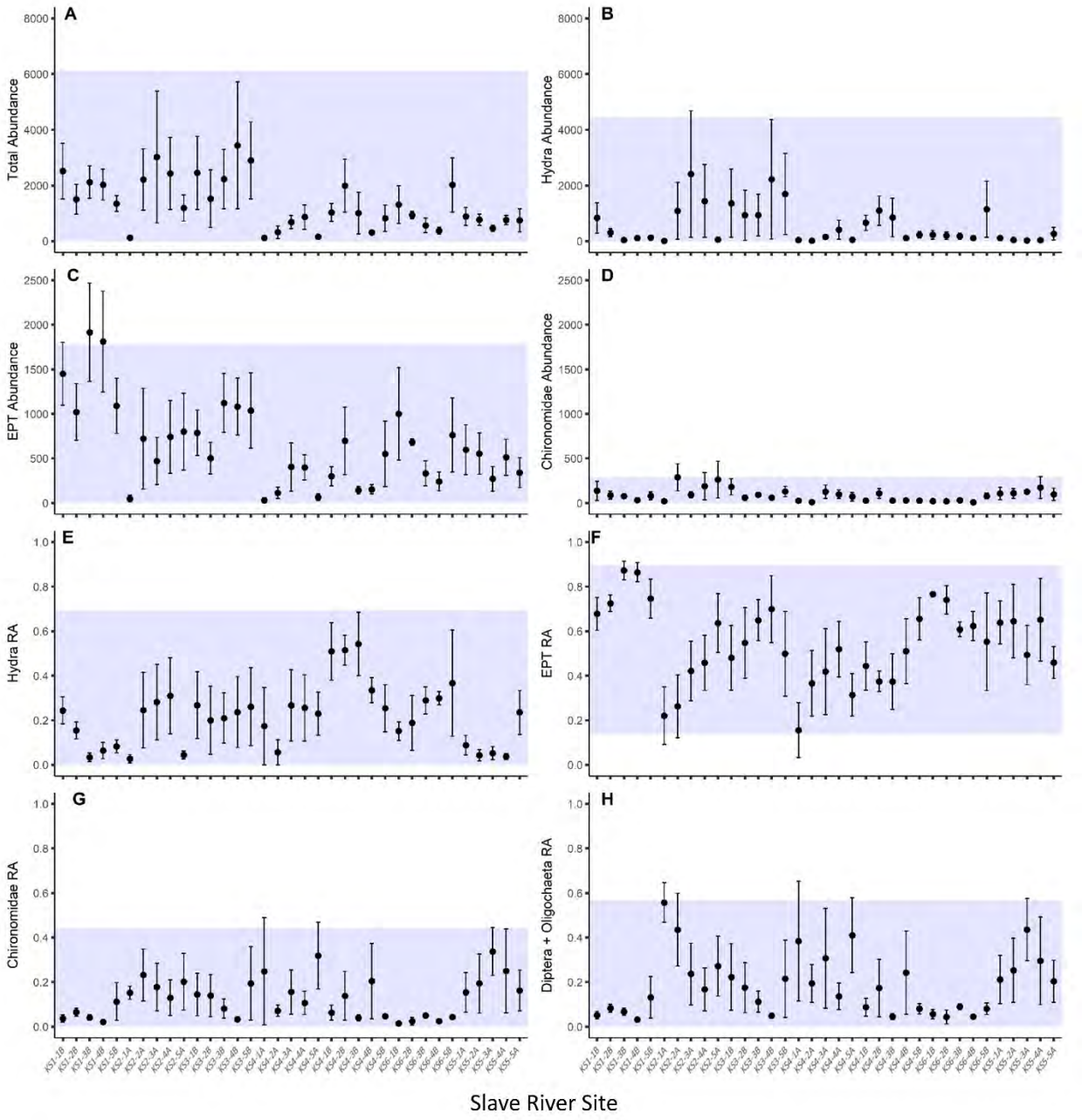

Figure 28. Site-scale temporal variability in abundance-based biotic metrics in the Slave River, showing mean ± SE for 2017-2021 for each site, with the grand mean (mean of annual means for the river) ± 2SD (normal range for the river) indicated by the shaded area, including (A) total abundance, (B) Hydra *abundance, (C) EPT abundance, (D) Chironomidae abundance, (E)* Hydra *relative abundance, (F) EPT relative abundance, (G) Chironomidae relative abundance, and (H) Diptera + Oligochaeta relative abundance. Sites are ordered from upstream (KS1-1B, left) to downstream (KS5-5A, right).*

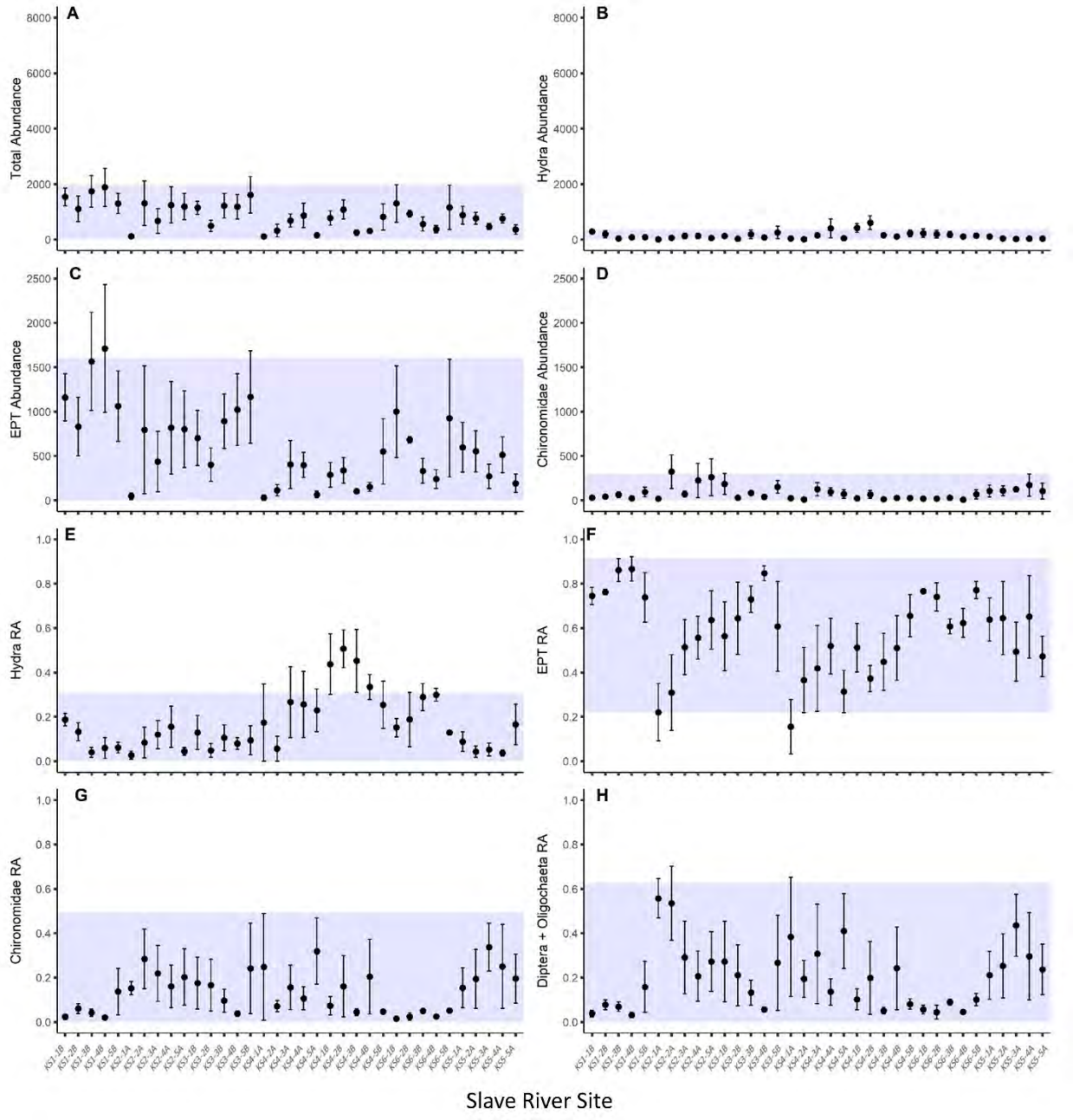

Figure 29. Site-scale temporal variability in abundance-based biotic metrics in the Slave River, showing mean ± SE for 2017-2021 (excluding 2020) for each site, with the grand mean (mean of annual means for the river, excluding 2020) ± 2SD (normal range for the river) indicated by the shaded area, including (A) total abundance, (B) Hydra *abundance, (C) EPT abundance, (D) Chironomidae abundance, (E)* Hydra *relative abundance, (F) EPT relative abundance, (G) Chironomidae relative abundance, and (H) Diptera + Oligochaeta relative abundance. Sites are ordered from upstream (KS1-1B, left) to downstream (KS5-5A, right).*

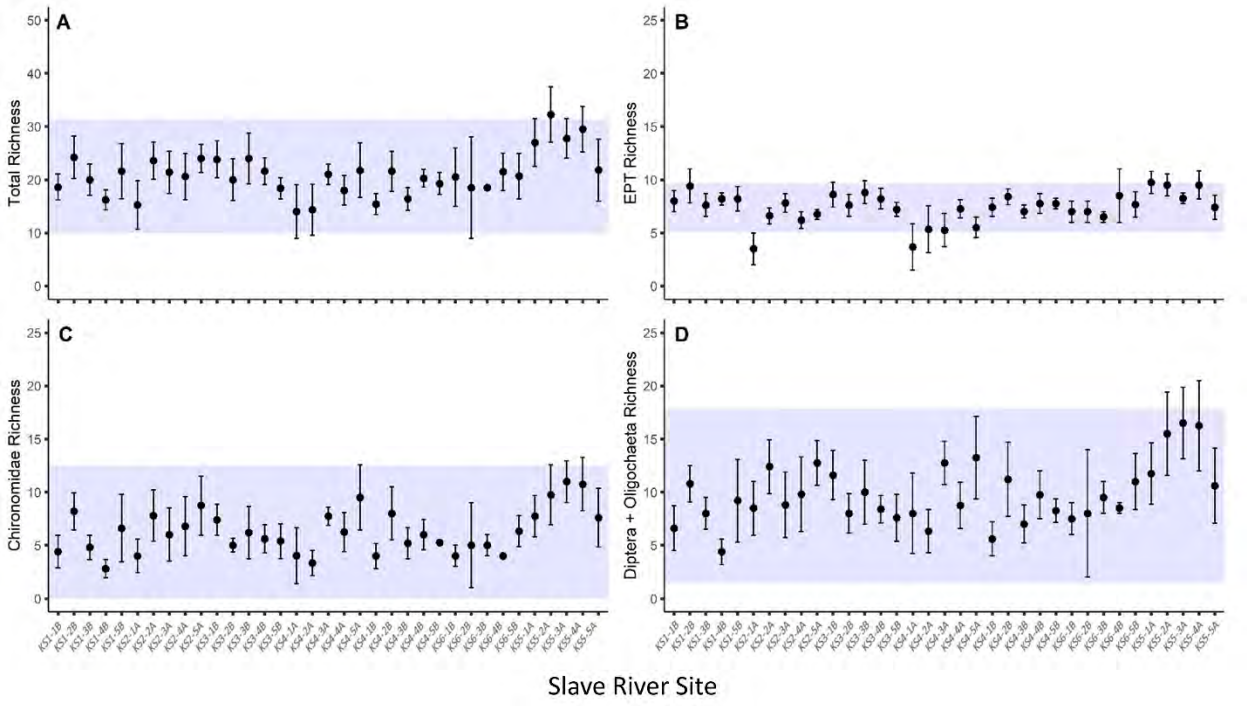

Figure 30. Site-scale temporal variability in richness-based biotic metrics in the Slave River, showing mean ± SE for 2017-2020 for each site, with the grand mean (mean of annual means for the river) ± 2SD (normal range for the river) indicated by the shaded area, including (A) total richness, (B) EPT richness, (C) Chironomidae richness, and (D) Diptera + Oligochaeta richness. Sites are ordered from upstream (KS1-1B, left) to downstream (KS5-5A, right).

system. This analysis was completed for abundance metrics, relative abundance metrics, and richness metrics. Because 2020 data were variable with respect to abundance metrics (due to high abundance of *Hydra*), the analysis was completed both with and without data from 2020 included, to explore the refinement of normal range limits with the removal of noise.

When all years (2017-2021) were included in the site-scale CES analysis, annual means across all sites varied widely for several abundance-based metrics, including total abundance, EPT abundance and relative abundance, *Hydra* abundance and relative abundance, and Diptera + Oligochaeta relative abundance (Figure 28). As a result of this variation, preliminary normal ranges were wide for each of these metrics, particularly for total abundance and *Hydra* abundance, driven primarily by Reach 2 and Reach 3. However, when data for 2020 were removed, and the influence of high abundance of *Hydra* thus removed from the normal range calculation, there was much less variability evident in total abundance and *Hydra* metrics, and the normal range for these metrics was much more narrow (Figure 29). The comparison between the plots with and without the data from 2020 speaks to the importance of refining CES boundaries when there is a known influence on assemblage structure in a particular year, as Figure 29A indicates that total abundance could be a useful metric for assessment based on the preliminary normal range, whereas CES boundaries for this metric in Figure 28A were largely too wide to be of use in detecting change. The narrower normal range for *Hydra* metrics and the observed association of *Hydra* with high water levels in 2020 suggests that this metric might be a useful indicator of flow-based impacts in this river.

The CES boundaries for Chironomidae abundance remained narrow with the addition of 2021 data, whether or not data from 2020 were included in the calculation. However, EPT abundance metrics were highly variable within and among sites and among years (both with and without 2020 data included; Figure 28C, F; Figure 29C, F). EPT are clearly important taxa in the Slave River, making up a large proportion of the BMI assemblage. The variability in abundance-based metrics for this taxonomic group suggests that it may not be a powerful diagnostic tool for small changes, given that the boundaries of the preliminary normal range were very wide. However, the normal range boundaries of this metric would allow for detection of either an increase or decrease (as the lower bound of the normal range did not extend to zero), which is useful for monitoring. It's possible that the normal range for these metrics will become more narrow as more data are added, but the preliminary CES boundaries, defined at the river scale, are very wide. It may be beneficial to focus on CES boundaries defined at the reach scale for these metrics (see section 3.3.1.2), as this may result in a narrower normal range with greater diagnostic power.

The preliminary estimate of the normal range of variability in site-scale analyses was more narrow for richness metrics, but did suffer from fairly wide inter-annual variability in site means across the entire river (Figure 30). The exception was EPT richness, which had an extremely narrow preliminary normal range based on the grand mean (Figure 30B), but two sites fell outside the CES boundaries because the normal range was so narrow. The lower CES boundaries for both total richness and EPT richness were greater than 0 (total richness of 10 taxa and EPT richness of 5 taxa), which does offer the opportunity to detect a loss of richness using the preliminary normal range for these metrics. In contrast, the lower CES boundaries for Chironomidae richness and Diptera + Oligochaeta richness were at or just above 0. Exclusion of data from 2020 when calculating the normal range did not have a large effect on richness metrics (results not shown), as the changes in 2020 were primarily seen in abundance metrics.

There were several metrics in the site-scale analysis that appear to have utility for detecting change. Total abundance (with 2020 data excluded), Chironomidae abundance, total richness, and EPT richness appeared to have the greatest initial potential for developing monitoring and management triggers. Total abundance and Chironomidae abundance would allow for detection of increases that are outside of the normal range (the lower CES boundary for both metrics was at zero, and therefore would not allow for detection of a decrease). Similarly, *Hydra* metrics could be considered as a way to detect change related to flow conditions, as a large increase in *Hydra* similar to those seen in 2020 may accompany future extreme high flow events. In contrast, total richness and EPT richness would allow for detection of values that fall either above the upper CES boundary (indicating richness above normal range) or below the lower CES boundary (indicating richness below the normal range). Although EPT relative abundance had a wide normal range, it is another metric that would allow for detection of changes in either direction. Total richness may be a particularly important metric to continue to monitor, given the apparent decline in total richness in several reaches from 2017-2021.

3.3.2.2. Reach-scale variability

Temporal variability at the reach scale was quantified by estimating the reach-specific normal range and developing preliminary CES boundaries based on variability among years. The preliminary estimates of the normal range for each reach were calculated based on the grand mean (mean of annual means for the reach from 2017, 2018, 2019, 2020, 2021) ± 2SD. Mean metric values ± SE for each reach in each year (averaged across sites, which are treated as replicates in this analysis) were compared with the

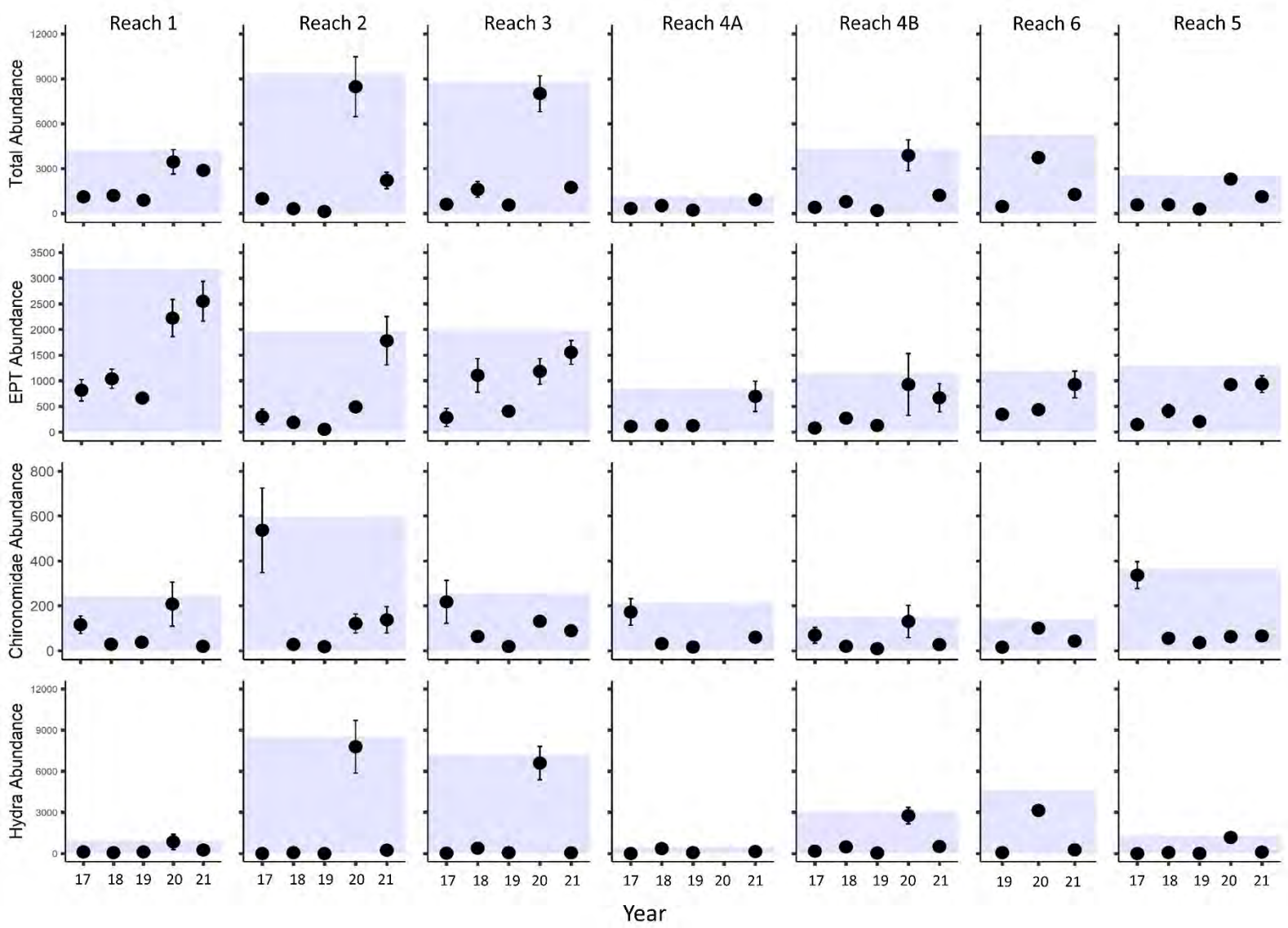

Figure 31. Reach-scale temporal assessment of normal range and critical effect size for abundance-based metrics in the Slave River, including total abundance and abundance of EPT, Chironomidae, and Hydra*. Points represent the mean ± SE across all sites in a reach, plotted for each year (2017-2021). Shaded area represents the normal range and CES boundaries for that reach, calculated based on the grand mean (mean of annual means for the reach) ± 2 SD. Each column shows data for a single reach.*

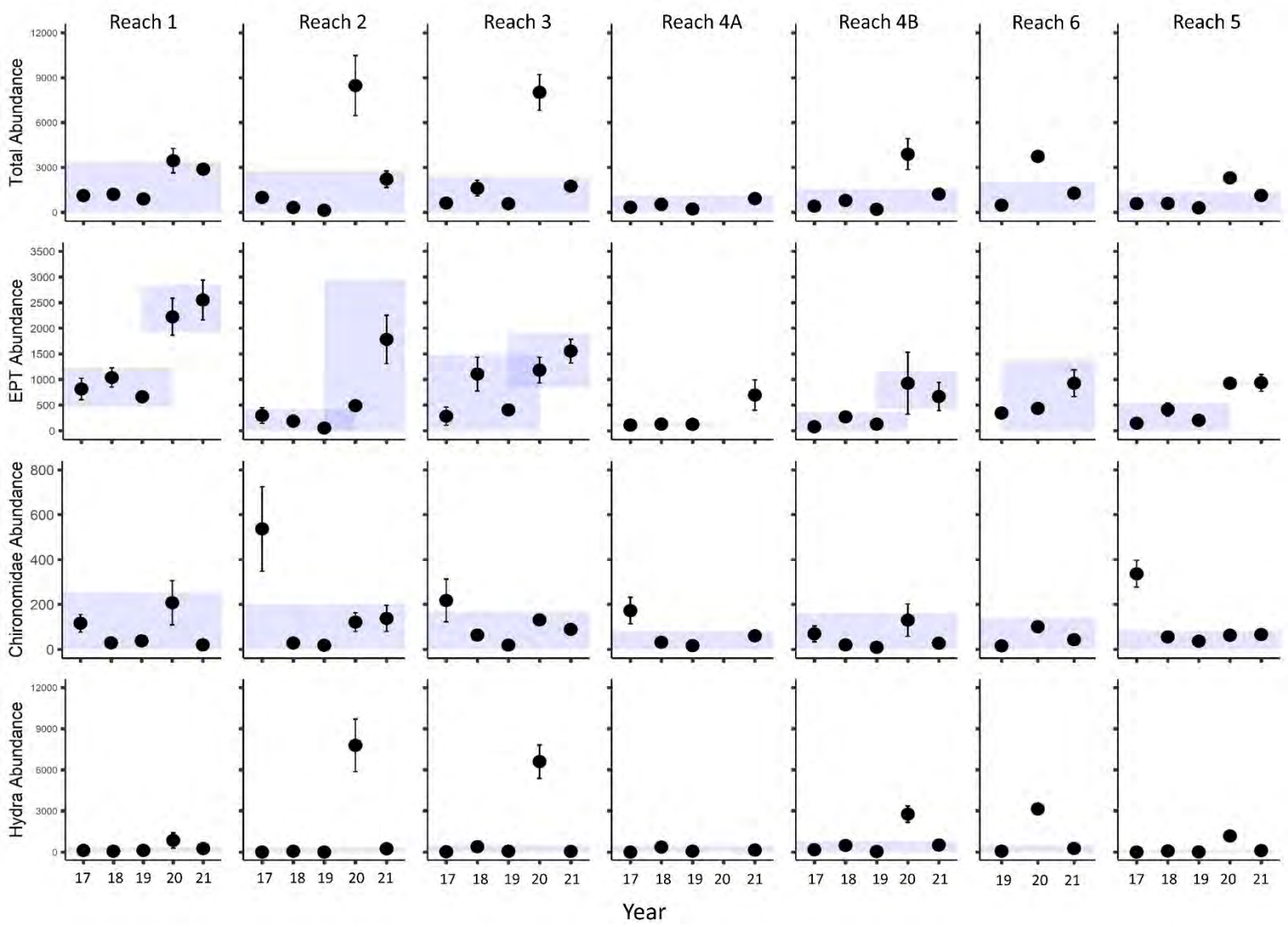

Figure 32. Reach-scale temporal assessment of normal range and critical effect size for abundance-based metrics in the Slave River, including total abundance and abundance of EPT, Chironomidae, and Hydra*, with normal range calculated based on a subset of years for each metric (total abundance: 2017-2021 excluding 2020; EPT: 2017-2019 and 2020-2021; Chironomidae: 2018-2021;* Hydra*: 2017-2021 excluding 2020). Points represent the mean ± SE across all sites in a reach, plotted for each year (2017-2021). Shaded area represents the normal range and CES boundaries for that reach, calculated based on the grand mean (mean of annual means for the reach) ± 2 SD. Each column shows data for a single reach.*

calculated normal range. This approach develops location-specific CES that can be used in continued monitoring at each reach to identify when samples are unusual or outside the range of expected variability for within the reach. Initial evaluation of these boundaries focused on their width, as a normal range may not be useful for detecting future impairment if it encompasses too wide a range of possible values. As with site-scale variability, the differences in assemblage composition observed in 2020 resulted in wider normal range boundaries for some abundance metrics than were observed previously for these reaches (Lento 2022). However, the width of the normal range for Chironomidae and Diptera + Oligochaeta metrics were more strongly affected by the high abundance of Chironomidae in 2017 (Lento 2022). In order to explore the potential for refinement of reach-specific normal range and preliminary CES boundaries, the analysis was run a second time for a selection of metrics with normal range calculated based on a subset of sampling years (as described below).

At the reach scale, there was a great deal of variability in the width of the estimated normal range for each abundance-based metric when all years were used to calculate the CES boundaries (Figure 31). For total abundance, this pattern reflected the large increase in abundance in 2020 due to the increase in *Hydra*, and the calculated normal range did not appear to reflect the patterns in abundance in other sampling years (Figure 31). Similarly, the *Hydra* abundance metric had a wide normal range in several reaches that resulted from increased abundance in 2020.

The normal range for EPT abundance was also wide for several reaches due to an increase in abundance in 2020 (particularly for Reach 1; Figure 31). However, in all reaches the highest abundance of EPT was found in 2021, indicating that conditions in 2021 remained ideal for some EPT taxa. In contrast, while Chironomidae abundance was elevated in some reaches in 2020 (namely Reach 1 and Reach 4B), the high abundance in 2017 appeared to contribute most strongly to a wide normal range for this metric. Since 2018, Chironomidae abundance has remained relatively more stable, and should be expected to have a more narrow normal range.

In order to account for the observed differences among years and try to refine the normal range and CES boundaries to better reflect typical composition in the Slave River reaches, reach means for each abundance metric were plotted with normal ranges based on subsets of years, with the chosen years differing depending on the metric. For total abundance and *Hydra* abundance, both of which were most affected by the 2020 data, normal range was calculated as the mean of means ± 2SD for all years excluding 2020. Exclusion of data for 2020 from normal range calculation resulted in a narrow normal range that appeared to better reflect the majority of sampling years for both total abundance and *Hydra* abundance (Figure 32). The narrower CES boundaries offer the opportunity to detect an increase in either metric in response to changes in environmental conditions or impairment. For example, if extreme flow conditions again lead to a large increase in abundance of *Hydra*, either metric should detect this. Total abundance could also be used to detect a large increase in the abundance of other taxa, which might follow impairment (for example, impacts that reduce competition for tolerant taxa and allow them to thrive).

As the normal range for Chironomidae appeared to be most affected by the high abundance of this group in 2017, a revised normal range was calculated as the mean of means ± 2SD for 2018-2021. Exclusion of data from 2017 resulted in a more narrow normal range for Chironomidae abundance for some reaches (notably Reach 2, Reach 3, Reach 4A, and Reach 5). The revised normal range appeared to be more reflective of the typical abundance of this group since 2018 (Figure 32).

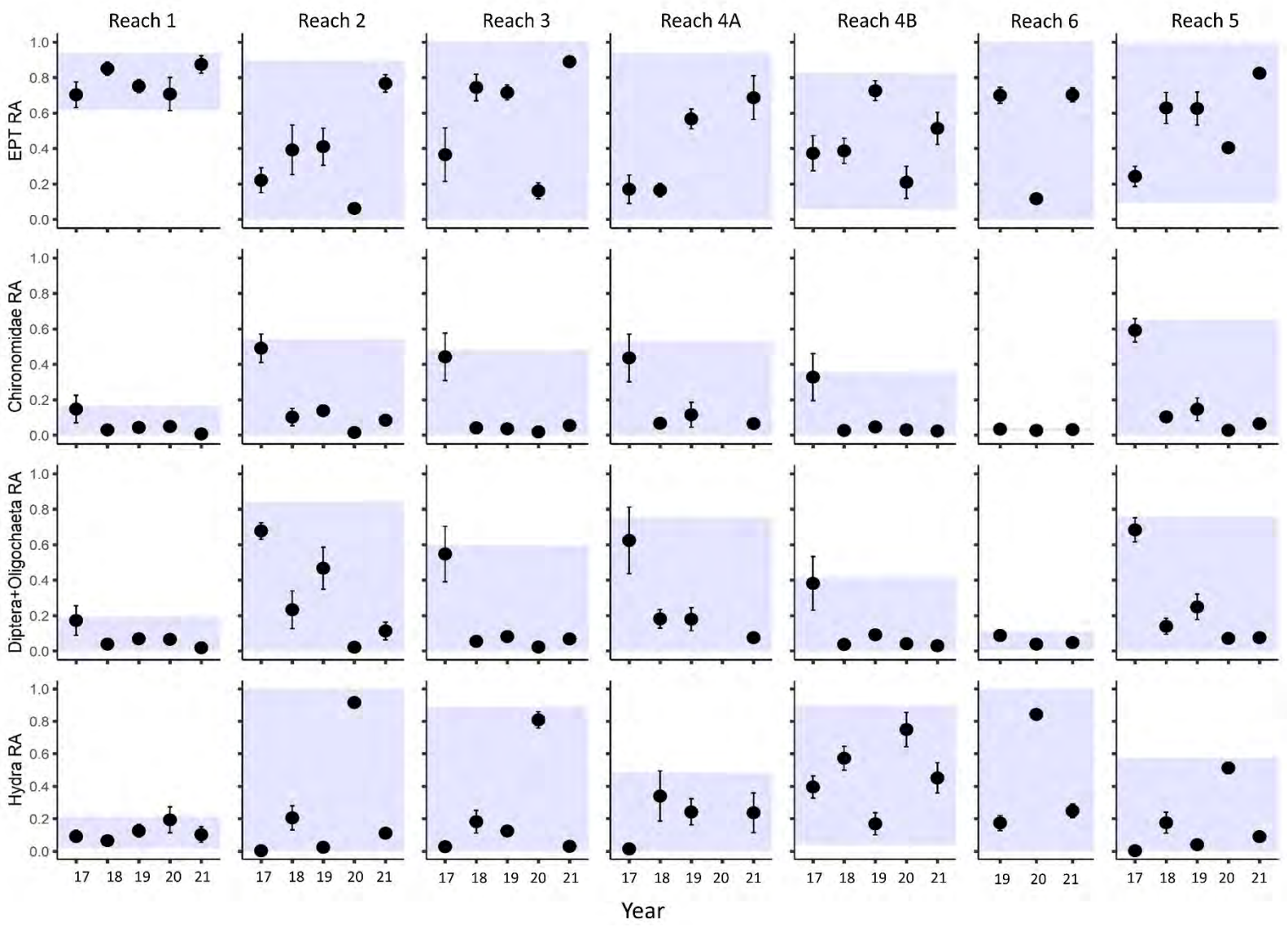
Reach 1
Reach 2
Reach 3
Reach 4A
Reach 4B
Reach 6
Reach 5
EPT RA
Chironomidae RA
Diptera+Oligochaeta RA
Hydra RA
0.0
0.2
0.4
0.6
0.8
1.0
17 18 19 20 21
19 20 21
Year

Figure 33. Reach-scale temporal assessment of normal range and critical effect size for relative abundance-based metrics in the Slave River, including relative abundance of EPT, Chironomidae, Diptera + Oligochaeta, and Hydra*. Points represent mean ± SE across all sites in a reach, plotted for each year (2017-2021). Shaded area represents the normal range and CES boundaries for the reach, calculated based on the grand mean (mean of annual means for the reach) ± 2 SD. Each column shows data for a single reach.*

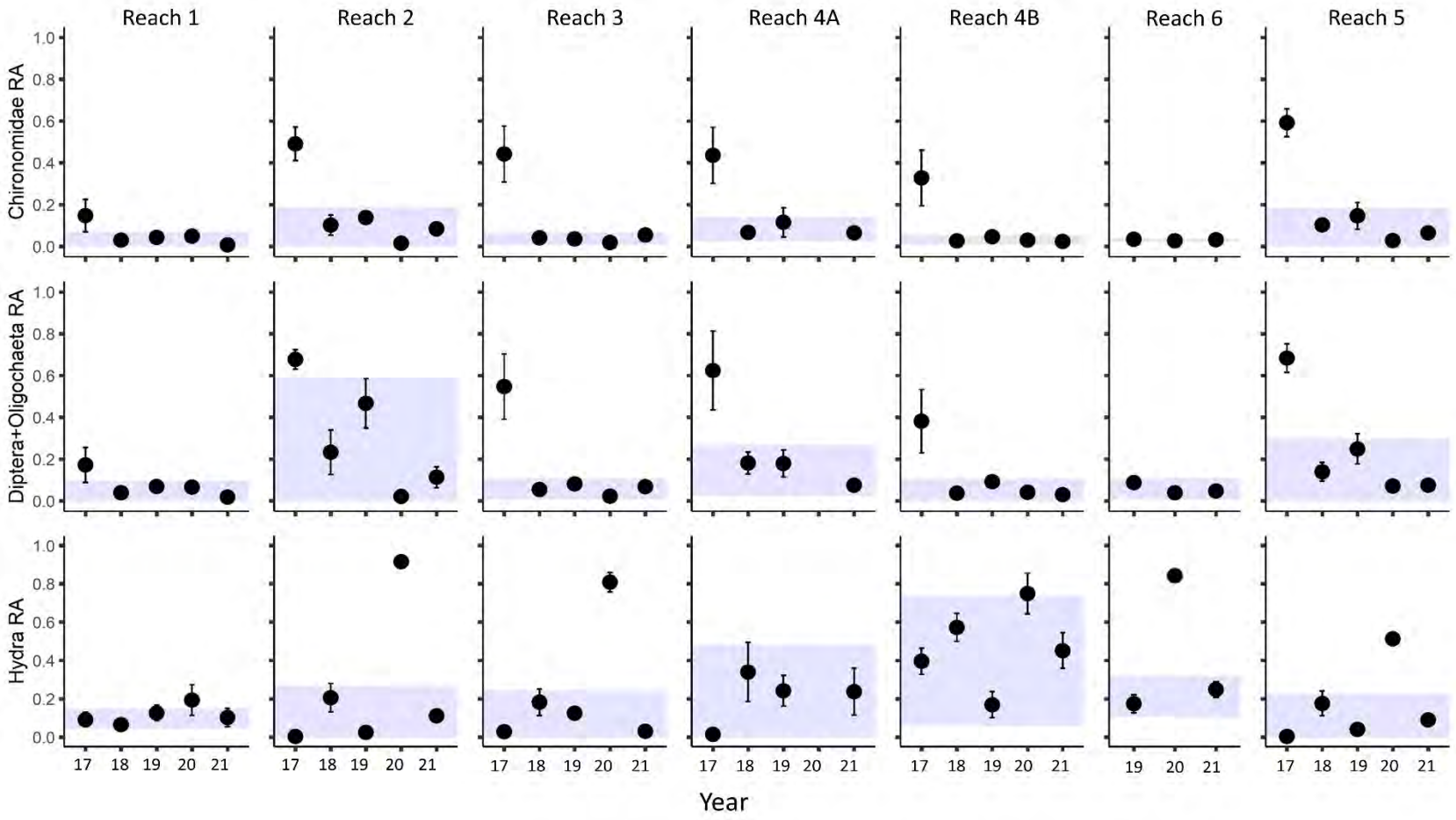

Figure 34. Reach-scale temporal assessment of normal range and critical effect size for relative abundance-based metrics in the Slave River, including relative abundance of EPT, Chironomidae, Diptera + Oligochaeta, and Hydra*, with normal range calculated based on a subset of years for each metric (Chironomidae RA: 2018-2021; Diptera + Oligochaeta RA: 2018-2021;* Hydra *RA: 2017-2021 excluding 2020). Points represent mean ± SE across all sites in a reach, plotted for each year (2017-2021). Shaded area represents the normal range and CES boundaries for the reach, calculated based on the grand mean (mean of annual means for the reach) ± 2 SD. Each column shows data for a single reach.*

Revision of normal range for EPT abundance was more exploratory. In all reaches, EPT abundance increased in 2021, and in most reaches it was also high in 2020. If the high abundance of the last two years of sampling reflects a response to high flows, then creating a normal range for the lower flow period (2017-2019) and a second normal range for the high flow period (2020-2021) provides an opportunity to classify these differences and refine the normal range based on environmental conditions. To explore this idea, mean EPT abundance was plotted with a separate normal range (mean of means ± 2SD) for each time period. In Reach 1, there was a clear distinction between the normal range derived for 2017-2019 and the much higher normal range for 2020-2021 (Figure 32). Though less extreme, creation of two sets of CES boundaries also helped distinguish these two time periods with differing mean abundance in Reach 4A, Reach 4B, and Reach 5. The remaining reaches had greater similarity in mean EPT abundance between the two time periods. For Reach 2 and Reach 3, this might have been a reflection of the lower EPT abundance in 2020 that appeared to result from the predominance of *Hydra* in these reaches (Figure 32). Although this approach appeared to be beneficial for the refinement of CES boundaries in several reaches, it is exploratory. Continued monitoring will be necessary to determine whether EPT abundance returns to levels seen before 2020, or whether it remains elevated.

Relative abundance metrics were much more variable across years, and for several metrics, the preliminary normal range encompassed nearly the full range of possible values (Figure 33). Normal range boundaries for relative abundance metrics were more narrow for in Reach 1, likely due to its relative invariability across years. However, for most reaches, the relative abundance metrics would have no ability to detect potential impact in future sampling years with the current set of data. For EPT relative abundance, the variability was due in part to high relative abundance of *Hydra* in 2020 (resulting in low relative abundance of EPT), but also due to high relative abundance of Chironomidae in 2017 (also leading to low relative abundance of EPT). For other metrics, however, it was possible to take a similar approach to that taken with abundance metrics, and calculate alternative CES boundaries by excluding years from the normal range calculation.

When data for 2017 were excluded from the calculation of the normal range, CES boundaries for Chironomidae relative abundance were narrower and more closely resembled the variability in the metric that's been evident since the second year of sampling (Figure 34). Similarly, excluding 2017 from normal range calculations created narrower CES boundaries for the relative abundance of Diptera + Oligochaeta for most reaches. The exception was Reach 2, which notably appeared to have high abundances of non-midge Diptera and/or Oligochaeta in 2018 and 2019 (Figure 34). Finally, excluding data from 2020 from the calculation of the normal range for the relative abundance of *Hydra* created more narrow CES boundaries for most reaches, but highlighted the generally higher relative abundance of this taxon at Reach 4A and Reach 4B, both of which had more variability in the revised normal range.

Preliminary estimates of the normal range of variability using all years of data were somewhat more narrow for richness-based metrics (Figure 35), reflecting weaker temporal variability in these metrics than was observed for relative abundance. In particular, EPT richness had extremely low variability within reaches and among years, contributing to narrow CES boundaries in all reaches (Figure 35). The richness of Chironomidae and Diptera + Oligochaeta had higher variability, due primarily to higher abundance and richness of these groups in 2017 (Figure 35). Exclusion of 2017 data from these metrics would likely result in narrower CES boundaries (results not shown).

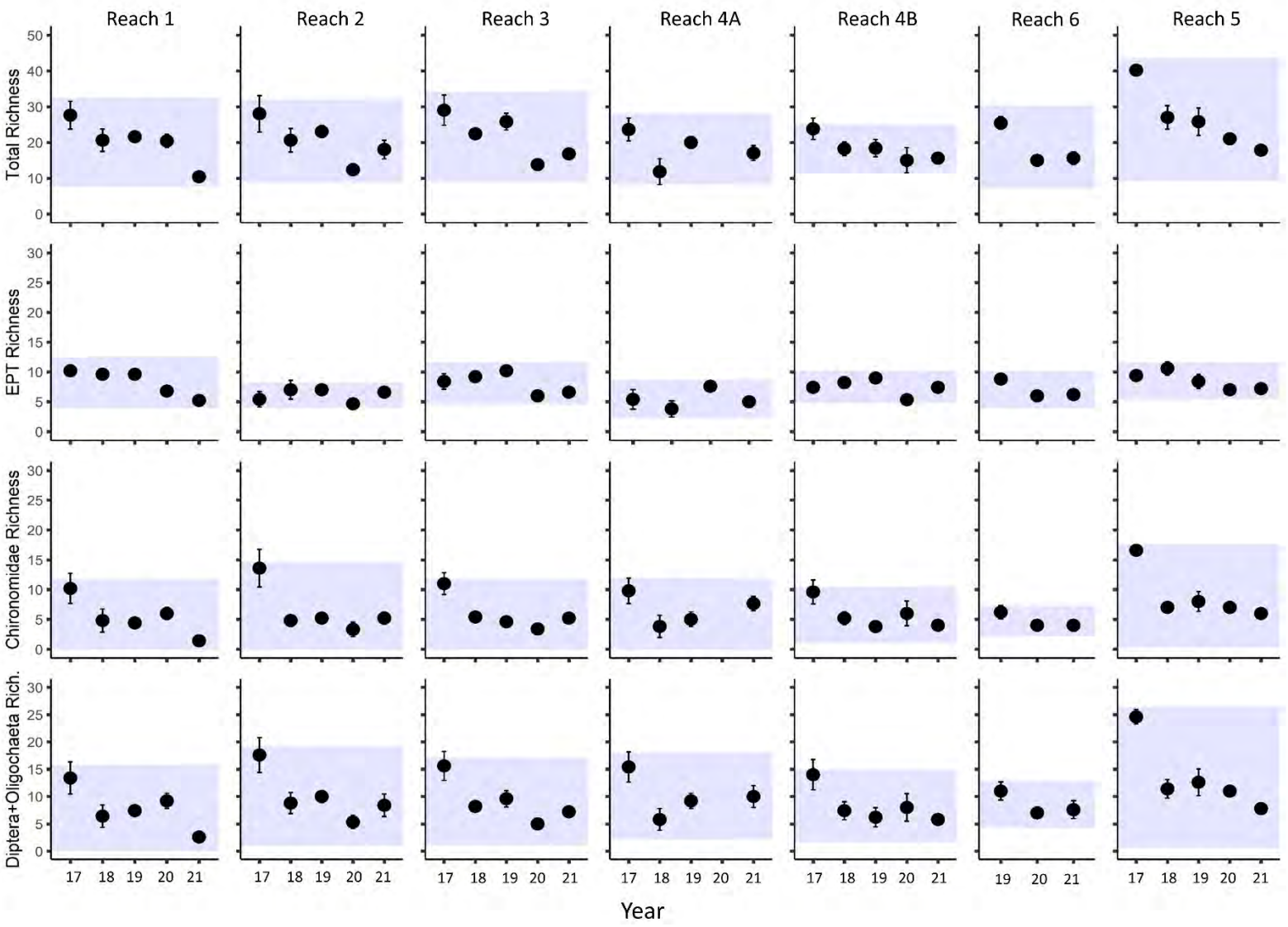

Figure 35. Reach-scale temporal assessment of normal range and critical effect size for richness-based metrics in the Slave River, including total richness and richness of EPT, Chironomidae, and Diptera + Oligochaeta. Points represent the mean ± SE across all sites in a reach, plotted for each year (2017-2021), and the shaded area represents the normal range and CES boundaries for that reach, calculated based on the grand mean (mean of annual means for the reach) ± 2 SD. Each column shows data for a single reach.

Total richness was among the most variable of richness metrics, and had a wide normal range for most metrics (Figure 35). This variability reflected the loss of Chironomidae richness in 2018 as well as continued loss of EPT, Chironomidae, and Diptera + Oligochaeta richness across sampling years in several reaches. Given that this appears to be an ongoing trend, it is important to retain the full set of sampling years in the calculation of normal range for this metric, and to continue to monitor change over time. If future years show a rebound in total taxonomic richness, then the normal range for this metric could be revisited and possibly refined with the exclusion of some years. For example, if Chironomidae richness remains low, then total richness values from 2017 may not be representative of these reaches. Furthermore, if total richness increases following the declines in 2020 and 2021, then these years could be considered to represent ongoing impact from high water levels in 2020, and total richness normal range could be refined with the exclusion of those years. However, more data will be needed to determine the range of natural variability in total richness.

With the refinement of CES boundaries through exclusion of years, there was evidence of several metrics in the reach-scale analysis that appear to have utility for detecting change. Total abundance (with 2020 data excluded), Chironomidae abundance and relative abundance (with 2017 data excluded), and Diptera + Oligochaeta relative abundance (with 2017 data excluded) would all allow for detection of increases that are outside the normal range. Similar to the site-scale analysis, *Hydra* metrics (with 2020 data excluded) could also be considered as a way to detect change related to flow conditions. EPT richness remained a strong metric for detecting either an increase or decrease that falls outside of the normal range of variability. And although total richness was found to be variable, it appeared to be an important metric to continue to monitor in order to track the ongoing decline across reaches.

3.4. Multivariate Normal Range and CES

Multivariate analysis of temporal variability in Slave River samples was used to assess compositional changes from 2017-2021 in the context of the full assemblage. The PCA of all sampling years (2017-2021) highlighted the differences between 2017 samples, which included abundant and diverse Chironomidae assemblages, and samples from all other years (Figure 36). The 95% probability ellipse for 2017 was large, and encompassed the 95% probability ellipses for all other years. Compared to 2017, every other year had much lower within-year variability among samples, indicated by a tight grouping of samples and smaller 95% normal probability ellipses. The ellipses for 2018-2021 overlapped, likely reflecting the lack of Chironomidae in these years. The overlap of ellipses from 2018-2021 with the ellipse for 2017 suggests partial similarity of 2018-2021 samples with a subset of samples from 2017 (Figure 36).

The normal probability ellipse for 2021 was the smallest, which indicated the strongest similarity among sites and likely reflected the lower taxonomic richness in that year. The small size of this ellipse may have been due to the assemblage changes that resulted from the elevated water levels in 2020 and the subsequent decline in water levels. If assemblages across all sites were in a state of recovery in 2021, this may have contributed to low richness and a greater similarity among sites.

Normal probability ellipses declined in size from 2017 to 2021, which may have been a reflection of the loss of chironomids (and richness due to chironomids) in 2017 and the lower richness in 2021. However, it's also possible that greater similarity in samples among reaches has resulted from the maturation of the program, with routine sampling contributing to greater precision in sampling. Assemblage patterns

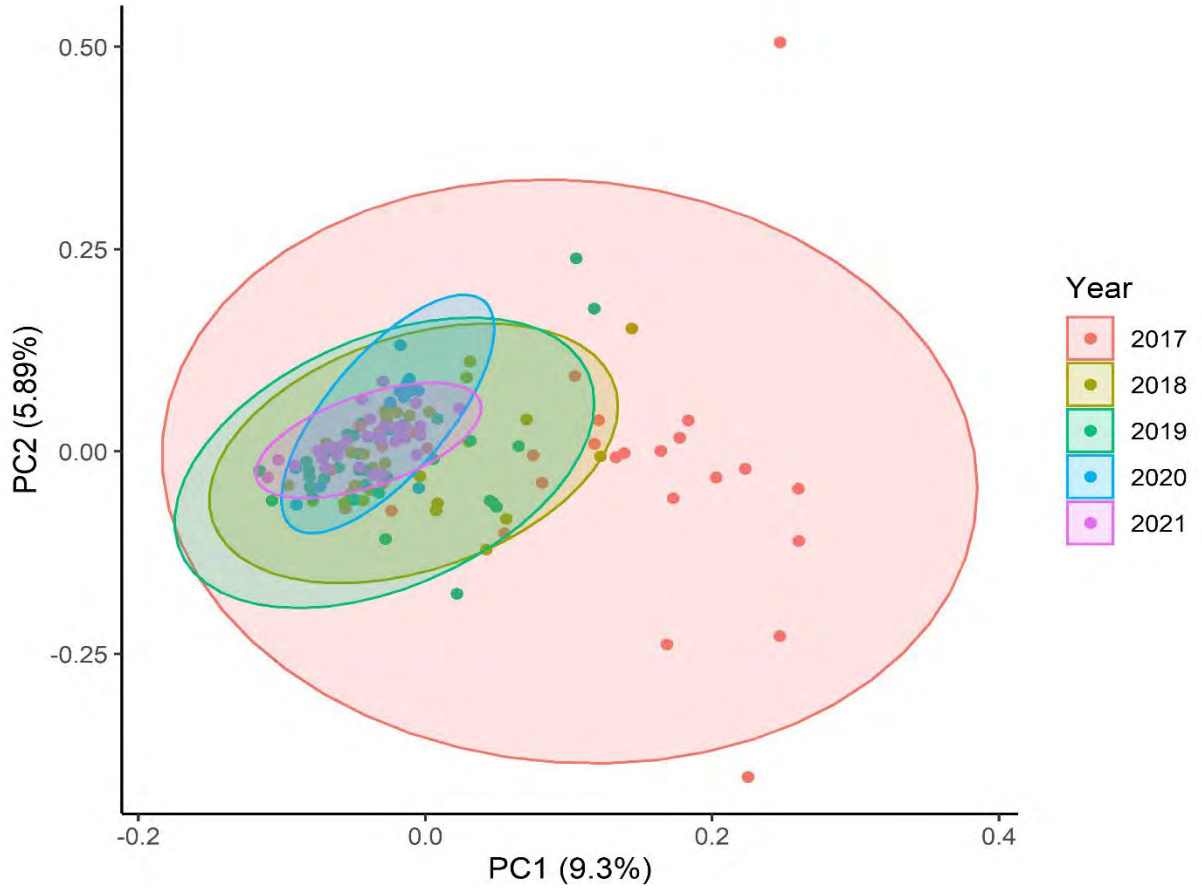

Figure 36 PCA ordination of Slave River samples from 2017 (red), 2018 (yellow), 2019 (green), 2020 (blue), and 2021 (purple) with 95% normal probability ellipses for each year. Normal probability ellipses indicate the area of multivariate space in which there's a 95% probability of a sample falling if it is representative of the population of samples collected in that year. Overlap of ellipses indicates similar composition in two years of sampling.

in 2022 will hopefully indicate whether this trend is ongoing. The use of normal probability ellipses in this context, with all years of sampling plotted simultaneously, provides a measure of the difference in taxonomic composition in the river among years, irrespective of site-specific variation. It highlights the strong overlap among years as well as the clear differences between 2017 (the most taxonomically diverse year) and all other sample years.

Procrustes analysis was used to compare individual ordinations among years and determine whether the placement of sites changed relative to one another between years. Shifts in assemblage composition of sites between years may result in high Procrustes residuals if a site was found to be more similar to a different group of sites in a latter year. For example, if all sites in Reach 1 plotted near each other in 2018 (indicating similar BMI composition) but one or two sites from Reach 1 plotted closer to a different reach in 2019 (indicating a shift in BMI composition relative to the rest of Reach 1), this pattern would be evidenced by high residuals in the comparison of 2018 and 2019. Although not directly a measure of change in composition, high residuals indicate compositional changes indirectly by providing a measure of the change in sites relative to one another.

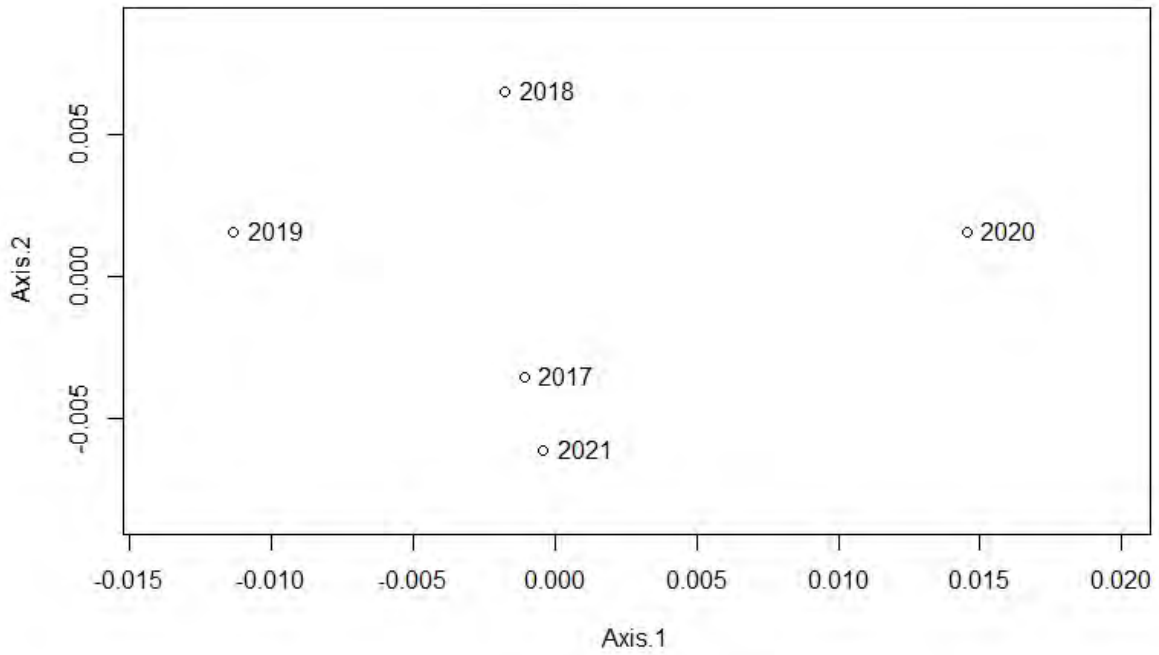

Figure 37 PCoA ordination of a dissimilarity matrix of m_{12}^2 values from pairwise comparisons of years in Procrustes analysis of Slave River samples. Each m_{12}^2 was divided by the number of sites in the pairwise comparison to account for differences due to sample size. Distance between years on the ordination biplot is representative of dissimilarity of samples between years.

Procrustes analysis identified site-scale changes among years in the Slave River. The sum of squared Procrustes residuals (m_{12}^2) ranged from 0.29 to 0.48 across all pairwise comparisons of years. The two most similar years were 2017 and 2021, whereas 2019 and 2021 were the two most dissimilar years with respect to the spatial arrangement of sites in ordination space. 2019 and 2020 were different and were also found to differ from all other years (at α =0.05). In contrast, ordinations in 2017, 2018, and 2021 were significantly more similar to each other than could be obtained by chance (at α =0.05). These results indicate that 2017, 2018, and 2021 all showed a similar arrangement of sites in multivariate space based on assemblage composition. The similarity of 2021 to the first two sample years and dissimilarity of 2019 and 2020 with all years is particularly interesting given the known impacts of high flows in 2020, as it suggests that assemblage composition in 2021 was returning to the pre-high-flow state.

These patterns were reflected in the PCoA, which showed a clear recovery trajectory from 2017 to 2021 (Figure 37). The PCoA was based on a matrix of Procrustes residuals, and showed a change from 2017 to 2018 along the second axis, a shift along the first and second axis in 2019 (orthogonal to 2017 and 2018), a strong shift along the first axis from 2019 to 2020 indicating the strong dissimilarity of these two years, and finally another orthogonal shift along the first and second axes to find 2021 located near 2017 (Figure 37). The strong similarity between 2017, 2018, and 2021 is interesting when considered in the context of the other analyses of composition among years particularly the temporal analysis of biotic metrics. But it is important to keep in mind that the results presented in Figure 37 are a reflection of the similarity among sites based on assemblage composition, not directly a reflection of assemblage composition. Regardless of the individual taxa contributing to the patterns, the similarity and

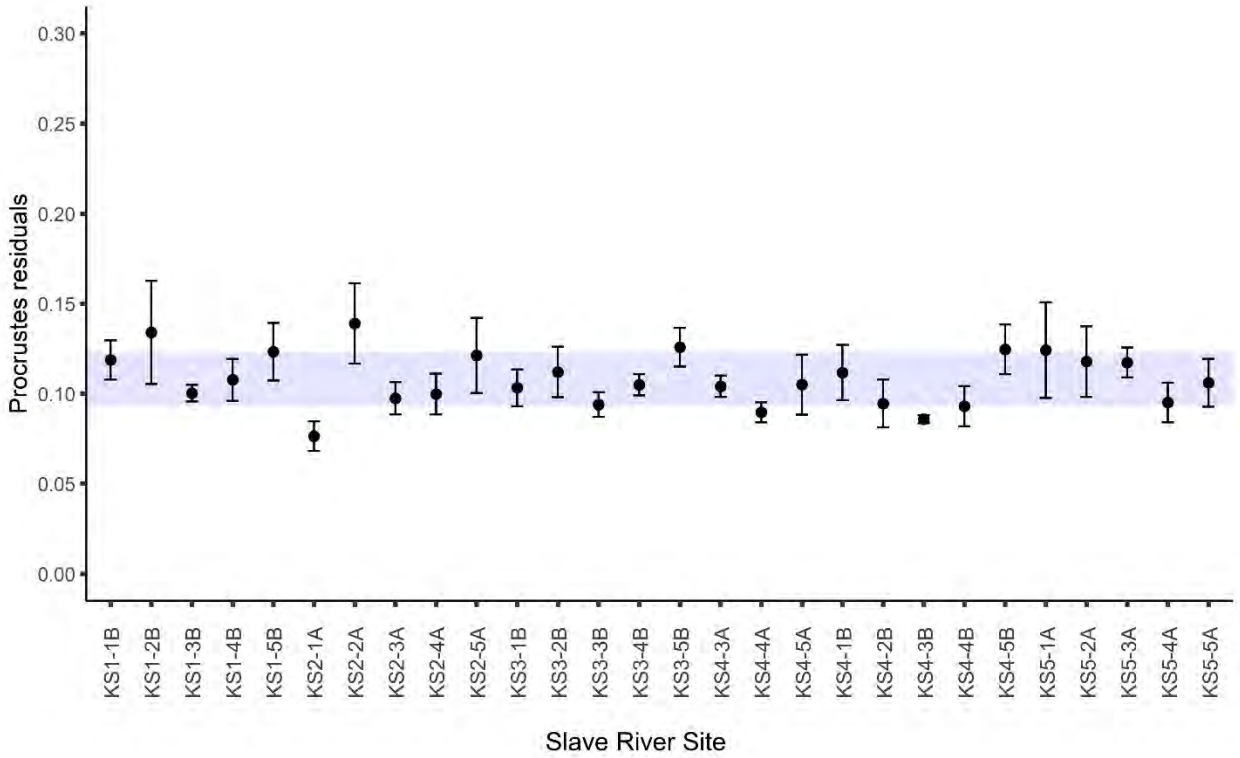

*Figure 38 Multivariate normal range and CES boundaries for the Slave River, with the mean ± SE Procrustes residual (2017, 2018, 2019, 2021) plotted for each site and the grand mean (mean of means for each pairwise year comparison) ± 2SD Procrustes residual indicated by the grey shaded area. Only sites sampled in all four years are included (*n *= 28).*

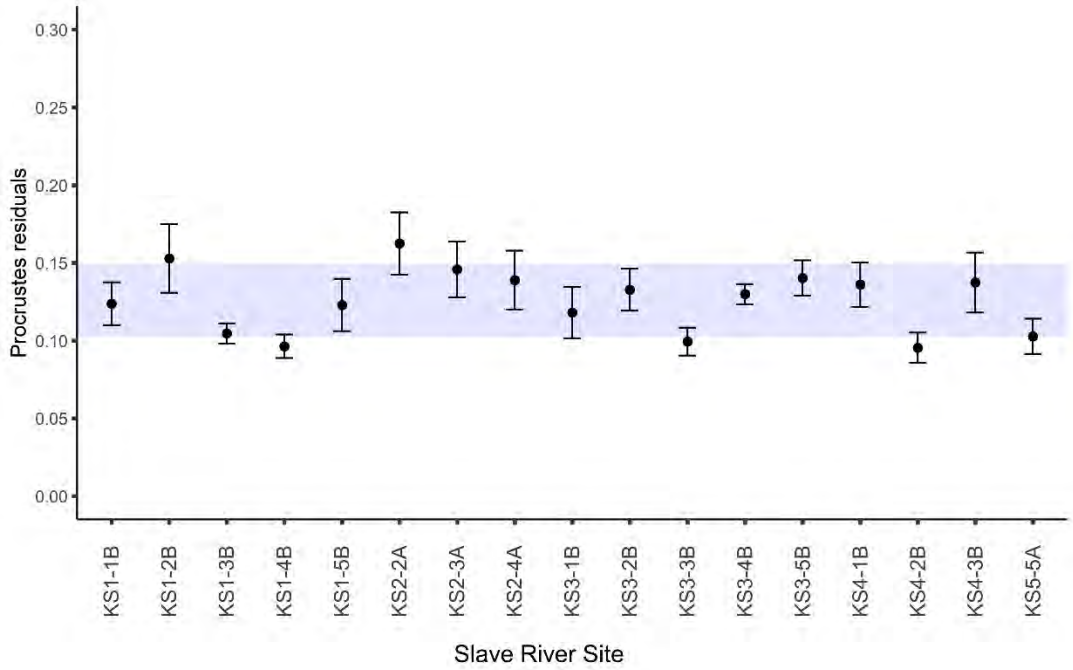

*Figure 39 Multivariate normal range and CES boundaries for the Slave River, with the mean ± SE Procrustes residual (2017-2021) plotted for each site and the grand mean (mean of means for each pairwise year comparison) ± 2SD Procrustes residual indicated by the grey shaded area. Only sites sampled in all five years are included (*n *= 17).*

dissimilarity among sites both within and among reaches in 2021 was comparable to that observed in the first year of sampling.

Preliminary estimates of a multivariate normal range and CES boundaries were created for the Slave River based on data from the period 2017-2021 excluding 2020 (28 sites) and 2017-2021 (17 sites). The normal range was calculated as the grand mean Procrustes residual (mean of mean residuals for each pairwise year comparison) ± 2 SD for each time period, and mean ± SE Procrustes residuals were plotted for each site to identify site-scale temporal variability relative to the normal range. For the period 2017-2021 (excluding 2020), when more sites were sampled, the normal range was narrow (Figure 38). Error bars were wide for several sites in Reach 1, Reach 2, and Reach 5, and some sites fell outside the normal range (either above the upper CES, representing a greater mean residual in pairwise comparisons, or below the lower CES boundary, indicating a lower mean residual). The mean residuals for six of 28 sites were above the upper CES boundary, whereas the mean residuals for three sites were below the lower CES boundary. Sites in Reach 3 had the most narrow error bars, indicating relatively little variability among all years (Figure 38).

Assessment of initial normal range for 2017-2021 was limited to only 17 sites that were sampled across all five years. The normal range for this subset of sites over all five years of sampling was similar in size to that of the larger group of sites, but it was shifted upwards, indicating higher residuals on average and greater temporal variability in the spatial arrangement of sites in multivariate space in the reduced ordinations (Figure 39). Sites in Reach 2 had higher mean residuals and were more variable when 2020 was included. This may have reflected the large shift in composition in Reach 2 in 2020 (where *Hydra* was quite dominant); however, it may also have reflected differences in the spatial arrangement of sites with the reduced subset of 17 sites. The utility of these results is unclear due to the limited sampling in 2020, resulting in a smaller subset of sites for the analysis. The reduced number of sites sampled in 2020 adds a confounding factor to the analysis, and results cannot therefore be directly compared with those from 2017-2019 + 2021, when many more sites were included in the ordinations. However, continuing to sample the full set of sites in the Slave River will allow for further refining of estimates of the normal range of Procrustes residuals with data from 2020 excluded.

3.5. Test of Generalized Procrustes Analysis

Generalized Procrustes Analysis (GPA) was explored as a potential additional method to quantify the normal range of multivariate variation among sites and to allow new data to be tested against an ordination that summarizes temporal variability in assemblage composition. GPA can potentially extend the concept of multivariate normal range explored in section 3.4 by creating a consensus ordination, or ordination that represents the average of multiple years of data. This consensus ordination would represent a reference ordination with which new data could be compared. If the ordination of newly collected data is not significantly different from the consensus ordination, then it suggests that the spatial arrangement of sites in multivariate space based on assemblage structure has not changed significantly relative to the temporal average. Furthermore, data from additional years of sampling can be added to the consensus ordination to further refine the reference point for comparison.

In order to test this method, GPA was run with data from 2017-2019 (excluding 2020 because of the small number of sites sampled and because of known differences in that year due to high water levels) to create a consensus ordination. The consensus ordination was then compared with an ordination of

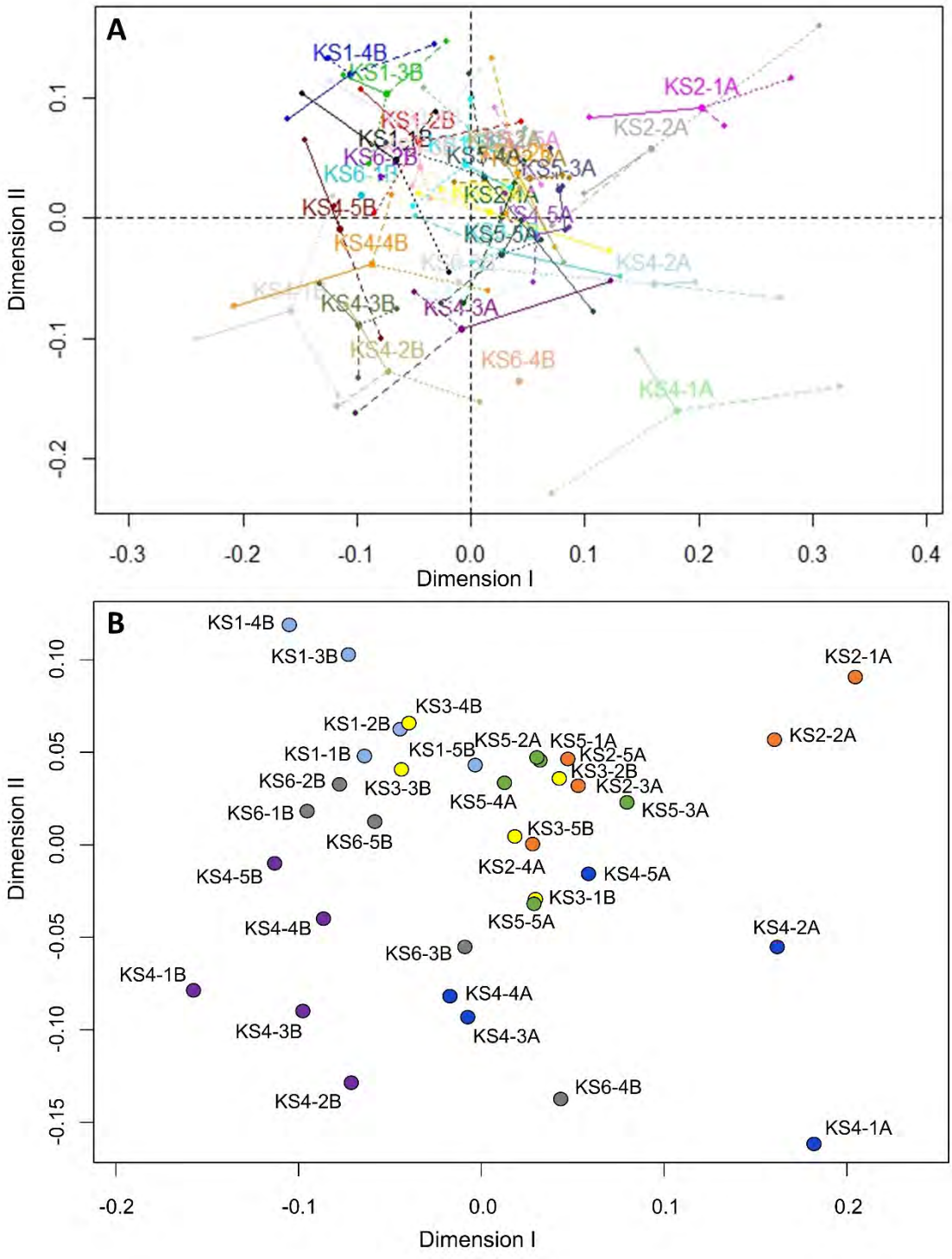

Figure 40. Results of GPA for Slave River data from 2017-2019, showing (A) the position of sites in the consensus ordination with residual vectors extending to each site's position in the original ordination from each of the three years (solid, dashed, and dotted line patterns represent the 2017, 2018, and 2019, respectively), and (B) the final consensus ordination, representing the average ordination of 2017-2019. Points in (B) are coloured by reach.

data from 2021 using Procrustes analysis, to determine whether there was a significant difference between the consensus ordination and new data.

The results from the GPA included a plot of residual vectors, with points indicating the position of each site in the consensus ordination and residual vectors showing each site's position in the three input ordinations (Figure 40A). Residual vectors were long for many sites, and in many cases, there were long residual vectors for all three original ordinations. For example, KS4-1A in the lower right quadrant of Figure 40A was the average of three points that were spatially separated, and thus had 3 long residual vectors. In the case of Reach 6, location in the consensus matrix was due to only a single ordination, as this reach was not sampled in 2017 or 2018. GPA is able to handle missing data when constructing consensus ordinations, but the position of sites in Reach 6 should be interpreted with caution, as they only represent a single year.

The consensus ordination generally grouped sites from the same reach together, which suggested a general similarity within reaches across years (Figure 40B). The spatial arrangement of sites reflected dissimilarity of Reach 4A and 4B from the remaining reaches along the second axis. The first axis separated Reach 4A from Reach 4B, and separated Reach 1 from Reach 2, while other reaches were located near the origin.

The consensus ordination for 2017-2019 was compared with an ordination of data from 2021, and their similarity was tested with Procrustes analysis. Although the sum of squared residuals was higher for the comparison than was observed in pairwise year comparisons ($m_{12}^2 = 0.54$), the two ordinations were significantly more similar than could occur by chance ($p = 0.001$), which indicated that the spatial arrangement of sites in multivariate space was similar between 2021 and the temporal average of 2017-2019. This result is to be expected, as 2021 was found to be similar to 2017 and 2018 in earlier Procrustes analysis. A plot of site residuals from the analysis indicated that residuals were long for many sites, but that sites generally shifted out from the origin, rather than shifting relative to surrounding sites (Figure 41). Sites in reaches KS1, KS3, and KS5 were the most similar in 2021 compared to the consensus matrix (i.e., they generally had fairly short vectors; Figure 41). In contrast, some sites in KS2, KS4B, and KS6 exhibited long vectors that indicated larger differences relative to the consensus matrix. In particular, sites KS2-2, KS4-2A, and KS6-4 had Procrustes residuals > 0.2, and showed stronger dissimilarity with the consensus matrix. The results highlight the potential utility of this test, because although the total Procrustes residual and some site residuals were fairly high for this comparison, the ordinations were still found to be statistically significantly similar.

The consensus ordination was also compared with the ordination of data collected in 2020 to gauge the level of dissimilarity of sites in 2020 from the average of 2017-2019. This analysis was completed with sites that weren't sampled in 2020 removed from the consensus matrix (but maintaining the position of remaining sites in the consensus ordination based on the average for all sites). Procrustes analysis results showed that the ordination of 2020 data was different from the consensus ordination ($m_{12}^2 = 0.60$; $p = 0.444$), and therefore the spatial arrangement of the 18 sites sampled in 2020 was different from their position in the consensus matrix. The largest site residuals were in reaches KS2 and KS4B (Figure 42), including residuals > 0.20 in KS4-2B and KS4-3B and residuals = 0.30 in KS2-2 and KS4-1B that contributed the most to the difference between ordinations. Given the degree to which assemblages in KS2 and KS4B changed in 2020 due to highly elevated abundances of *Hydra*, it is not surprising that these reaches stood out as differing from the consensus matrix in 2020.

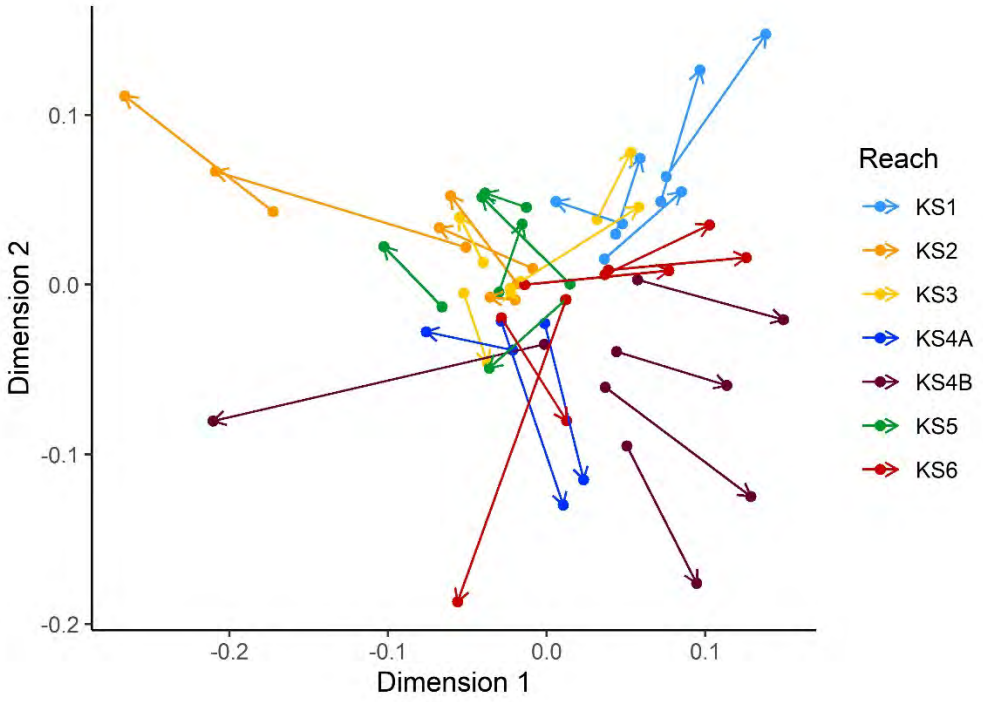

Figure 41. Slave River site residuals from Procrustes analysis of a consensus ordination from GPA (2017-2019) compared with an ordination of data from 2021, with vectors showing the movement of sites from the consensus ordination to the 2021 ordination following rotation and scaling. Points and vectors are coloured by Slave River reach.

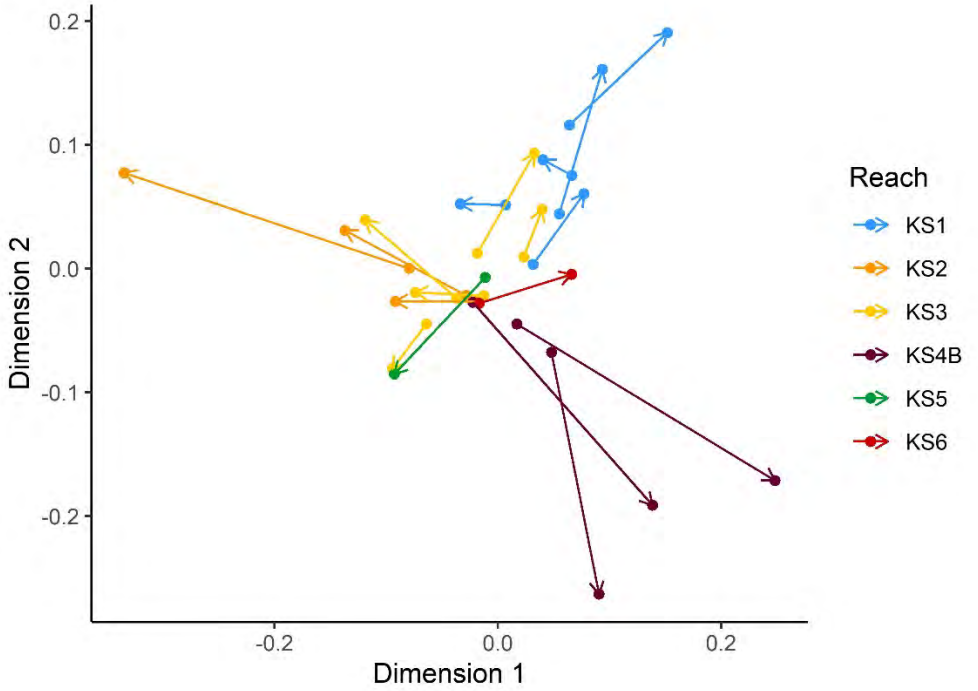

Figure 42. Slave River site residuals from Procrustes analysis of a consensus ordination from GPA (2017-2019) compared with an ordination of data from 2020 (18 sites), with vectors showing the movement of sites from the consensus ordination to the 2020 ordination following rotation and scaling. Points and vectors are coloured by Slave River reach.

A test consensus matrix with only the 18 sites sampled in 2020 was also created and compared with the 2020 ordination, to ensure differences between 2020 and the consensus matrix were not due to retaining the spatial configuration based on all sites. The 2020 ordination was still found to differ from this reduced consensus matrix (m_{12}^2 = 0.58; p = 0.267), and differences were strongly driven by the same sites in reaches KS2 and KS4B. These results confirmed that the detected differences were not an artifact of building the consensus matrix with the full set of sample sites and reaches.

Because the ordination for 2021 was found to be statistically similar to the consensus ordination for 2017-2019, the data for all four years were combined to create a new consensus ordination using GPA (Figure 43). The updated consensus ordination with data for four years did not differ strongly from the consensus ordination for 2017-2019, though Reach 2 was more tightly clustered with Reach 3 and Reach 5 in the revised ordination (Figure 43B). The plot of site residuals from the GPA showed relatively short residuals for the added 2021 data for several sites (Figure 43A), which was consistent with the similarity between the 2021 ordination and the 2017-2019 consensus ordination. In general, the addition of 2021 data to the consensus ordination appeared to refine estimates of temporal variability by strengthening the data used to create those estimates. In future sampling years, new data can be compared with this revised consensus ordination in order to detect temporal differences in the spatial arrangement of sites.

4. Recommendations and Conclusions

This report summarizes the first five years of sampling data in the Slave River, and introduces a more adaptive approach to the development of normal range and CES boundaries by considering the effect of extreme conditions. Though sampling in 2020 was challenging and was necessarily limited to a subset of reaches and sites, the benefits of collecting and characterizing data before, during, and after the high flow event of 2020 cannot be overstated. Through the use of biotic metrics, it was possible to see some initial variability in response to higher flows in 2019, a large shift in response to the extreme high flows in 2020, and a return to pre-2020 conditions in 2021. Furthermore, evaluation of assemblage composition with multivariate analysis showed evidence of a recovery trajectory, as the spatial arrangement of sites based on assemblage structure in 2021 was found to be most similar to that observed in 2017. These results indicate the potential response of BMI assemblages to extreme high flows in this system, but they also highlight the resiliency of these assemblages following a return to more typical habitat conditions.

The collection of additional data facilitates the development of more accurate and precise normal range estimates, as it becomes easier to separate the true pattern from the noise. With five years of data collected, and with knowledge of the extreme flow event in 2020, it was possible in this report to begin to adapt normal range estimates through the exclusion of years that did not appear representative of the typical composition of the river. When CES boundaries were calculated for the river as a whole, this was done through exclusion of data from 2020 from the calculation of the normal range. But at the reach scale, it was possible to refine these calculations further. For total abundance, the exclusion of data for 2020 resulted in a more precise normal range estimate, but for abundance metrics that included Chironomidae, the exclusion of data from 2017 (when Chironomidae were very abundant) provided the greatest increase in precision. Lento (2022) recommended that this approach be considered if the abundance of Chironomidae remained low, as it has been from 2018-2021. Although the reason for the much higher abundance of Chironomidae in 2017 remains unclear, abundances have

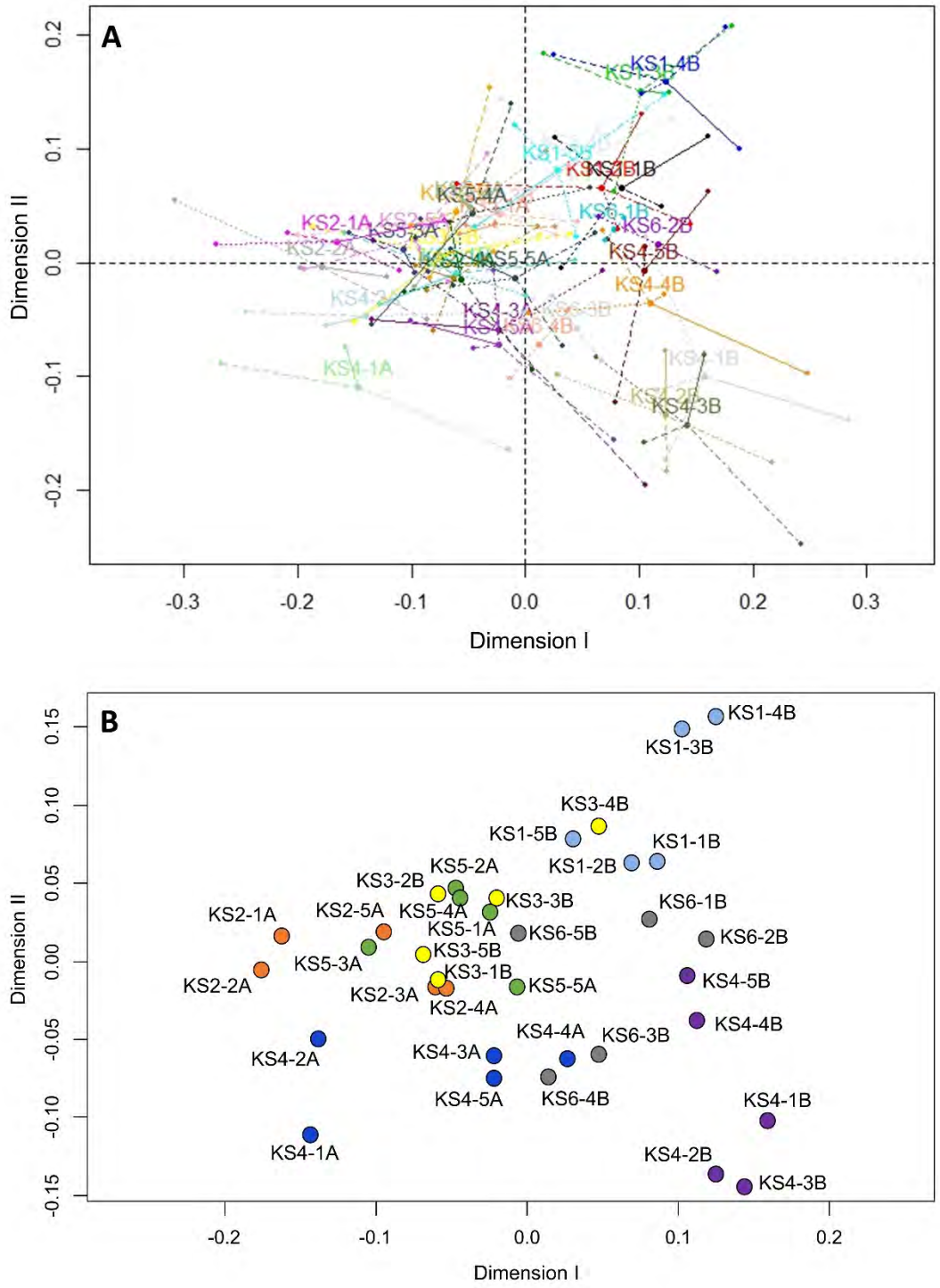

Figure 43. Results of GPA for Slave River data from 2017-2019 and 2021, showing (A) the position of sites in the consensus ordination with residual vectors extending to each site's position in the original ordination from each of the four years (solid, dashed, dotted, and dot-dash lines correspond to 2017, 2018, 2019, and 2021, respectively), and (B) the final consensus ordination, representing the average ordination of 2017, 2018, 2019, and 2021. Points in (B) are coloured by reach.

been sufficiently low in the four years since to warrant the creation of more precise CES boundaries. Following the changes to the calculation of the normal range, both total abundance and Chironomidae abundance appeared to be useful metrics for detecting future change. However, these refined normal range estimates and CES boundaries should be reassessed in future years to determine if the observed patterns in abundance are maintained.

Reach-scale variability in EPT abundance suggested that it might be beneficial to consider separate normal range criteria for 2020-2021 (when water levels were high or receding) and 2017-2019 (when water levels were lower). Development of separate normal range criteria for each time period provided an effective means to characterize temporal variability for some reaches, though the mechanistic support for this pattern remains unknown. It is possible that resilience to high flows, or the high mobility of some EPT taxa and their ability to colonize new habitats may have contributed to high abundances in 2020-2021, but additional data will be necessary to determine whether this pattern holds. The refined normal range estimates for EPT abundance are exploratory, and should be reassessed in future years.

Temporal variability in some taxonomic richness metrics remained low, highlighting the utility of these metrics for detecting future changes to the Slave River. In particular, EPT richness covered a narrow normal range and was identified as having high diagnostic potential. However, the assessment of normal range for richness metrics also highlighted the loss in total richness from 2017-2021 in most Slave River reaches. Multivariate analysis also indicated the decrease in taxonomic richness over time. In part, the decline in richness may have been a reflection of the loss of Chironomidae in 2018 and the impacts of high water levels in 2020, but it remains a pattern of potential concern that should continue to be monitored.

The addition of more data in 2021 facilitated further exploration of the use of multivariate analysis to detect temporal changes in assemblage composition. Pairwise Procrustes analysis offered insights into the change trajectory in the river from 2017 to 2021, and was also used to further refine normal range estimates and CES boundaries based on Procrustes residuals for each site. The normal range for Procrustes residuals for 2017-2019 and 2021 was narrow, and most sites fell within or below the CES boundaries. Further development of these boundaries with the addition of more data should facilitate the use of this approach to detect when one or more sites changes to an unusual degree relative to other sites (and thus has a high Procrustes residual) in future years.

A new approach that was explored was the use of GPA to develop a consensus ordination that summarized temporal variability among BMI ordinations. The consensus ordination acts as a reference point, capturing the variability across sample years, and allowing for comparison with newly collected data. The approach was tested by developing a consensus ordination using data from 2017-2019 and comparing it with an ordination from 2021, and the results found statistically significant similarity between the two ordinations, consistent with what was expected based on earlier analyses. A consensus ordination of 2017-2019 and 2021 combined can now be used in future years as a reference ordination to be compared with new data, to identify any significant changes to site configurations. Though not a direct test of assemblage structure, preliminary analysis suggests that it may be an effective technique to detect assemblage-level changes that lead to a shift in one or more sites relative to the others. As more data are added to the consensus ordination, it should become a more accurate and precise reflection of typical relationships among sites and reaches in the Slave River.

5. References

Albertson, L.K., and Daniels, M.D. 2016. Resilience of aquatic net-spinning caddisfly silk structures to common global stressors. Freshwater Biol. **61**(5): 670-679. doi: 10.1111/fwb.12737.

Anderson, M.J. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). Wiley StatsRef: Statistics Reference Online: 1-15. doi: 10.1002/9781118445112.stat07841.

Anderson, M.J., Ellingsen, K.E., and McArdle, B.H. 2006. Multivariate dispersion as a measure of beta diversity. Ecol. Lett. **9**: 683-693.

Arciszewski, T.J., and Munkittrick, K.R. 2015. Development of an adaptive monitoring framework for long-term programs: An example using indicators of fish health. Integrated environmental assessment and management **11**(4): 701-718.

Arciszewski, T.J., Munkittrick, K.R., Scrimgeour, G.J., Dubé, M.G., Wrona, F.J., and Hazewinkel, R.R. 2017. Using adaptive processes and adverse outcome pathways to develop meaningful, robust, and actionable environmental monitoring programs. Integrated environmental assessment and management **13**(5): 877-891.

Bailey, R.C., Norris, R.H., and Reynoldson, T.B. 2004. Bioassessment of freshwater ecosystems using the reference condition approach. Kluwer Academic Publishers, Boston, Massachusetts.

Barbour, M.T., Gerritsen, J., Snyder, B.D., and Stribling, J.B. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish. Second edition. Technical Report EPA 841-B-99-002, US Environmental Protection Agency, Office of Water, Washington, DC. 337 p.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Stat. Soc. B Met. **57**(1): 289-300.

Bonada, N., Prat, N., Resh, V.H., and Statzner, B. 2006. Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. Annu. Rev. Entomol. **51**: 495-523.

Bowman, M.F., and Somers, K.M. 2006. Evaluating a novel Test Site Analysis (TSA) bioassessment approach. J. N. Am. Benthol. Soc. **25**(3): 712-727.

Bunn, S.E., and Arthington, A.H. 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. Environmental management **30**(4): 492-507.

Buss, D.F., Carlisle, D.M., Chon, T.-S., Culp, J., Harding, J.S., Keizer-Vlek, H.E., Robinson, W.A., Strachan, S., Thirion, C., and Hughes, R.M. 2015. Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. Environ. Monit. Assess. **187**(1): 4132.

Canadian Council of Ministers of the Environment. 1999. Canadian sediment quality guidelines for the protection of environmental and human health: Polycyclic Aromatic Hydrocarbons. *In* Canadian Environmental Quality Guidelines, 1999. Canadian Council of Ministers of the Environment, Winnipeg, MB, Canada.

Canadian Council of Ministers of the Environment. 2001a. Canadian sediment quality guidelines for the protection of aquatic life. *In* Canadian Environmental Quality Guidelines, 1999. Canadian Council of Minsters of the Environment, Winnipeg, MB, Canada.

Canadian Council of Ministers of the Environment. 2001b. Canadian water quality guidelines for the protection of aquatic life. *In* Canadian Environmental Quality Guidelines, 1999. Canadian Council of Ministers of the Environment, Winnipeg.

Canadian Council of Ministers of the Environment. 2010. Canadian soil quality guidelines for the protection of environmental and human health: Polycyclic Aromatic Hydrocarbons. *In* Canadian Environmental Quality Guidelines, 1999. Canadian Council of Ministers of the Environment, Winnipeg, MB, Canada.

Culp, J.M., Lento, J., Curry, R.A., Luiker, E., and Halliwell, D. 2019. Arctic biodiversity of stream macroinvertebrates declines in response to latitudinal change in the abiotic template. Freshwater Science **38**(3): 465-479. doi: 10.1086/704887.

Culp, J.M., Lento, J., Goedkoop, W., Power, M., Rautio, M., Christoffersen, K.S., Guðbergsson, G., Lau, D., Liljaniemi, P., Sandøy, S., and Svoboda, M. 2012a. Developing a circumpolar monitoring framework for Arctic freshwater biodiversity. Biodiversity **13**(3-4): 215-227.

Culp, J.M., Goedkoop, W., Lento, J., Christoffersen, K.S., Frenzel, S., Guðbergsson, G., Liljaniemi, P., Sandøy, S., Svoboda, M., Brittain, J., Hammar, J., Jacobsen, D., Jones, B., Juillet, C., Kahlert, M., Kidd, K., Luiker, E., Olafsson, J., Power, M., Rautio, M., Ritcey, A., Striegle, R., Svenning, M., Sweetman, J., and Whitman, M. 2012b. The Arctic Freshwater Biodiversity Monitoring Plan. CAFF Monitoring Series Report Nr. 7, CAFF International Secretariat.p.

Dagg, J. 2016. Vulnerability Assessment of the Slave River and Delta. summary report for the Community Workshop convened in Fort Smith, January 24-26, 2012, The Pembina Institute. 56 p.

Dubé, M.G. 2003. Cumulative effect assessment in Canada: a regional framework for aquatic ecosystems. Environmental Impact Assessment Review **23**(6): 723-745.

Dubé, M.G., Duinker, P., Greig, L., Carver, M., Servos, M., McMaster, M., Noble, B., Schreier, H., Jackson, L., and Munkittrick, K.R. 2013. A framework for assessing cumulative effects in watersheds: an introduction to Canadian case studies. Integrated environmental assessment and management **9**(3): 363-369.

Environment and Climate Change Canada. 2019. Federal Environmental Quality Guidelines: Iron. Environmenta nd Climate Change Canada, Ottawa, ON. 9 p.

Environment Canada. (ed.) 2011. Integrated monitoring plan for the oil sands: expanded geographic extent for water quality and quantity, aquatic biodiversity and effects, and acid sensitive lake component. EN14-49/2011E-PDF, Environment Canada, Gatineau, QC, Canada.

Environment Canada. 2012. Canadian Aquatic Biomonitoring Network Field Manual - Wadeable Streams. Dartmouth, NS, Canada, Government of Canada Publications. 57 p.

Environment Canada. 2014. Canadian Aquatic Biomonitoring Network Laboratory Methods: processing, taxonomy, and quality control of benthic macroinvertebrate samples. Dartmouth, NS, Canada, Government of Canada Publications. 36 p.

Flotemersch, J., Stribling, J., Hughes, R., Reynolds, L., Paul, M., and Wolter, C. 2011. Site length for biological assessment of boatable rivers. River Res. Applic. **27**(4): 520-535.

Glozier, N.E., Donald, D.B., Crosley, R.W., and Halliwell, D. 2009. Wood Buffalo National Park water quality: status and trends from 1989-2006 in three major rivers: Athabasca, Peace and Slave. Environment Canada, Environment and Climate Change Canada. 104 p.

Golder Associates. 2010. Aquatic Ecosystem Health - Report on State of the Knowledge. Final Report prepared for the Government of the Northwest Territories. Report number 09-1328-0036. 93 pp.

Gower, J.C. 1975. Generalized procrustes analysis. Psychometrika **40**(1): 33-51.

Hirst, C.N., and Jackson, D.A. 2007. Reconstructing community relationships: the impact of sampling error, ordination approach, and gradient length. Diversity Distrib. **13**: 361-371.

Jackson, D.A. 1995. PROTEST: A PROcrustean Randomization TEST of community environment concordance. Ecoscience **2**(3): 297-303.

Jackson, J.K., and Füreder, L. 2006. Long-term studies of freshwater macroinvertebrates: a review of the frequency, duration and ecological significance. Freshwater Biol. **51**(3): 591-603. doi: 10.1111/j.1365-2427.2006.01503.x.

Johnson, M.F., Reid, I., Rice, S.P., and Wood, P.J. 2009. Stabilization of fine gravels by net-spinning caddisfly larvae. Earth Surface Processes and Landforms **34**(3): 413-423. doi: 10.1002/esp.1750.

Kilgour, B.W., Somers, K.M., Barrett, T.J., Munkittrick, K.R., and Francis, A.P. 2017. Testing against "normal" with environmental data. Integrated environmental assessment and management **13**(1): 188-197.

Le, S., Josse, J., and Husson, F. 2008. FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software **25**(1): 1-18.

Lento, J. 2017. Review and development of recommendations on monitoring and assessment protocols for benthic macroinvertebrate communities in transboundary rivers. Report prepared for the Government of the Northwest Territories. Prepared for the Government of the Northwest Territories. 23 p.

Lento, J. 2018. Benthic macroinvertebrate monitoring plan for the transboundary rivers of the Northwest Territories. Report prepared for the Government of the Northwest Territories. 20 pp.

Lento, J. 2020. Benthic macroinvertebrate monitoring plan for large transboundary rivers in the Alberta-Northwest Territories region: Assessment of results from the second year of sampling (2018). Report prepared for the Alberta-Northwest Territories Bilateral Management Committee, Government of Alberta, and Government of the Northwest Territories, Government of Northwest Territories. 103 p.

Lento, J. 2021. Benthic macroinvertebrate monitoring plan for large transboundary rivers in the Alberta-Northwest Territories region: Assessment of results from the third year of sampling (2019). Report prepared for the Alberta-Northwest Territories Bilateral Management Committee, Government of Alberta, and Government of the Northwest Territories, Government of Northwest Territories. 104 p.

Lento, J. 2022. Benthic macroinvertebrate monitoring plan for large transboundary rivers in the Alberta-Northwest Territories region: Assessment of results from the fourth year of sampling (2020). Report prepared for the Alberta-Northwest Territories Bilateral Management Committee, Government of Alberta, and Government of the Northwest Territories, Government of Northwest Territories. 98 p.

Lento, J., Dillon, P.J., Somers, K.M., and Reid, R.A. 2008. Changes in littoral benthic macroinvertebrate communities in relation to water chemistry in 17 Precambrian Shield lakes. Can. J. Fish. Aquat. Sci. **65**: 906-918.

Lento, J., Monk, W.A., Culp, J.M., Curry, R.A., Cote, D., and Luiker, E. 2013. Responses of low Arctic stream benthic macroinvertebrate communities to environmental drivers at nested spatial scales. Arctic, Antarctic, and Alpine Research **45**(4): 538-551.

Lento, J., Laske, S., Lavoie, I., Bogan, D., Brua, R.B., Campeau, S., Chin, K., Culp, J.M., Levenstein, B., Power, M., Saulnier-Talbot, É., Shaftel, R., Swanson, H., Whitman, M., and Zimmerman, C.E. 2022a. Diversity of diatoms, benthic macroinvertebrates, and fish varies in response to different environmental correlates in Arctic rivers across North America. Freshwater Biol. **67**(1): 95-115. doi: 10.1111/fwb.13600.

Lento, J., Culp, J.M., Levenstein, B., Aroviita, J., Baturina, M.A., Bogan, D., Brittain, J.E., Chin, K., Christoffersen, K.S., Docherty, C., Friberg, N., Ingimarsson, F., Jacobsen, D., Lau, D.C.P., Loskutova, O.A., Milner, A.M., Mykrä, H., Novichkova, A.A., Ólafsson, J.S., Schartau, A.K., Shaftel, R., and Goedkoop, W. 2022b. Temperature and spatial connectivity drive patterns in freshwater macroinvertebrate diversity across the Arctic. Freshwater Biol. **67**(1): 159-175. doi: 10.1111/fwb.13805.

Lento, J., Goedkoop, W., Culp, J., Christoffersen, K., Fefilova, E., Guðbergsson, G., Liljaniemi, P., Ólafsson, J.S., Sandøy, S., Zimmerman, C., Christensen, T., Chambers, P., Heino, J., Hellsten, S., Kahlert, M., Keck, F., Laske, S., Lau, D.C.P., Lavoie, I., Levenstein, B., Mariash, H., Rühland, K., Saulnier-Talbot, E., Schartau, A.K., and Svenning, M. 2019. State of the Arctic Freshwater Biodiversity Report. Conservation of Arctic Flora and Fauna (CAFF) International Secretariat,,. 124 p.

Loudon, C., and Alstad, D.N. 1992. Architectural plasticity in net construction by individual caddisfly larvae (Trichoptera: Hydropsychidae). Can. J. Zool. **70**(6): 1166-1172. doi: 10.1139/z92-163.

MacDonald Environmental Sciences Ltd. 1995. Expert's Workshop on the Development of Ecosystem Maintenance Indicators for the Transboundary River Systems within the Mackenzie River Basin: Workshop Summary Report. Government of the Northwest Territories. 56 p.

Martinez Arbizu, P. 2020. pairwiseAdonis: Pairwise multilevel comparison using adonis. (R package version 0.4).

Matteucci, S.D., and Pla, L. 2010. Procrustes analysis as a tool for land management. Ecological Indicators **10**(2): 516-526. doi: 10.1016/j.ecolind.2009.09.005.

McArdle, B.H., and Anderson, M.J. 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. Ecology **82**(1): 290-297. doi: https://doi.org/10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2.

McGrath, J.A., Joshua, N., Bess, A.S., and Parkerton, T.F. 2019. Review of Polycyclic Aromatic Hydrocarbons (PAHs) Sediment Quality Guidelines for the Protection of Benthic Life. Integrated Environmental Assessment and Management **15**(4): 505-518. doi: 10.1002/ieam.4142.

Milner, A.M., Loza Vega, E.M., Matthews, T.J., Conn, S.C., and Windsor, F.M. 2023. Long-term changes in macroinvertebrate communities across high-latitude streams. Global Change Biol. **00**: 1-12. doi: https://doi.org/10.1111/gcb.16648.

Monk, W.A., Wood, P.J., Hannah, D.M., and Wilson, D.A. 2008. Macroinvertebrate community response to inter-annual and regional river flow regime dynamics. River Res. Applic. **24**(7): 988-1001.

Monk, W.A., Compson, Z.G., Armanini, D.G., Orlofske, J.M., Curry, C.J., Peters, D.L., Crocker, J.B., and Baird, D.J. 2018. Flow velocity–ecology thresholds in Canadian rivers: A comparison of trait and taxonomy-based approaches. Freshwater Biol. **63**(8): 891-905.

Munkittrick, K.R., and Arciszewski, T.J. 2017. Using normal ranges for interpreting results of monitoring and tiering to guide future work: A case study of increasing polycyclic aromatic compounds in lake sediments from the Cold Lake oil sands (Alberta, Canada) described in Korosi et al.(2016). Environ. Pollut. **231**: 1215-1222.

Munkittrick, K.R., Arens, C.J., Lowell, R.B., and Kaminski, G.P. 2009. A review of potential methods of determining critical effect size for designing environmental monitoring programs. Environmental Toxicology and Chemistry **28**(7): 1361-1371.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., and Wagner, H. 2020. Vegan: Community Ecology Package. (R package version 2.5-7). Available from https://CRAN.R-project.org/package=vegan.

Paterson, M., Lawrence, M., and Sekerak, A. 1991. Benthic Invertebrates and Biomonitoring in the Slave River, N.W.T.: A Pilot Study. Report for the Slave River Environmental Quality Monitoring Program. 164 p.

Paterson, M., Lawrence, M., and Sekerak, A. 1992. Benthic Invertebrates and Biomonitoring in the Slave River, N.W.T.: 1991 Survey. Report for the Slave River Environmental Quality Monitoring Program. 59 p.

Pembina Institute. 2016. State of the Knowledge of the Slave River and Slave River Delta. Prepared for the Slave River and Delta Partnership. 124 p.

Peters, D.L., Monk, W.A., and Baird, D.J. 2014. Cold-regions Hydrological Indicators of Change (CHIC) for ecological flow needs assessment. Hydrological Sciences Journal **59**(3-4): 502-516.

R Development Core Team 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.R-project.org. ISBN 3-900051-07-0

Resh, V.H. 2008. Which group is best? Attributes of different biological assemblages used in freshwater biomonitoring programs. Environ. Monit. Assess. **138**: 131-138.

Reynoldson, T.B., Norris, R.H., Resh, V.H., Day, K.E., and Rosenberg, D.M. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. J. N. Am. Benthol. Soc. **16**(4): 833-852.

Sanderson, J., Czarnecki, A., and Faria, D. 2012. Water and Suspended Sediment Quality of the Transboundary Reach of the Slave River, Northwest Territories. Aboriginal Affairs and Northern Development Canada. 415 p.

Somers, K.M., Kilgour, B.W., Munkittrick, K.R., and Arciszewski, T.J. 2018. An adaptive environmental effects monitoring framework for assessing the influences of liquid effluents on benthos, water and sediments in aquatic receiving environments. Integrated environmental assessment and management.

Stantec Consulting Ltd. 2016. State of Aquatic Knowledge for the Hay River Basin. Report prepared for the Government of the Northwest Territories, Department of Environment and Natural Resources, Government of the Northwest Territories. 285 p.

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., and Norris, R.H. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecol. Appl. **16**(4): 1267-1276.

ter Braak, C.J.F., and Šmilauer, P. 2002. Reference Manual and User's Guide to CANOCO for Windows (version 4.5). Center for Biometry, Wageningen.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. Available from https://ggplot2.tidyverse.org. ISBN 978-3-319-24277-4

6. Appendices

Table 9 Names and coordinates of all kick-sampling sites in the Hay River and Slave River. Sites sampled in 2021 are indicated with a sampling date.

River	Reach	Site	Latitude	Longitude	Date	River	Reach	Site	Latitude	Longitude	Date	Notes
HAY RIVER	REACH 1	HR-KS1-1A	59.93403	-116.95028		SLAVE RIVER	REACH 1	SR-KS1-1B	59.40805	-111.46321	9/8/2021	
		HR-KS1-2A	59.93591	-116.95175				SR-KS1-2B	59.40805	-111.46321	9/8/2021	
		HR-KS1-3A	59.93211	-116.95237				SR-KS1-3B	59.40846	-111.46196	9/8/2021	
		HR-KS1-4A	59.93135	-116.95506				SR-KS1-4B	59.40879	-111.46082	9/8/2021	
		HR-KS1-5A	59.93124	-116.95613				SR-KS1-5B	59.40913	-111.45985	9/8/2021	
	REACH 2	HR-KS2-1A	59.94548	-116.95565			REACH 2	SR-KS2-1A	59.42689	-111.46155	9/8/2021	
		HR-KS2-2A	59.94617	-116.95618				SR-KS2-2A	59.42709	-111.46199	9/8/2021	
		HR-KS2-3A	59.94654	-116.95647				SR-KS2-3A	59.42761	-111.46294	9/8/2021	
		HR-KS2-4A	59.94703	-116.95702				SR-KS2-4A	59.42799	-111.46361	9/8/2021	
		HR-KS2-5A	59.94759	-116.95744				SR-KS2-5A	59.42858	-111.46458	9/8/2021	
	REACH 3	HR-KS3-1A	59.98767	-116.93236			REACH 3	SR-KS3-1B	59.53395	-111.45934	9/9/2021	
		HR-KS3-2A	59.98827	-116.93060				SR-KS3-2B	59.53372	-111.45978	9/9/2021	
		HR-KS3-3A	59.98845	-116.93037				SR-KS3-3B	59.53502	-111.45774	9/9/2021	
		HR-KS3-4A	59.99023	-116.93049				SR-KS3-4B	59.53538	-111.45703	9/9/2021	
		HR-KS3-5A	59.99182	-116.93127				SR-KS3-5B	59.53562	-111.45651	9/9/2021	
	REACH 4	HR-KS4-1A	60.00158	-116.97036			REACH 4A	SR-KS4-1A	59.58906	-111.41968	9/9/2021	No kick sample
		HR-KS4-2A	60.00205	-116.97145				SR-KS4-2A	59.58947	-111.4196	9/9/2021	No kick sample
		HR-KS4-3A	60.00261	-116.97126				SR-KS4-3A	59.59122	-111.41951	9/9/2021	
		HR-KS4-4A	60.00308	-116.97089				SR-KS4-4A	59.59178	-111.41949	9/9/2021	
		HR-KS4-5A	60.00319	-116.97009				SR-KS4-5A	59.59225	-111.41946	9/9/2021	
	REACH 5	HR-KS5-1B	60.01064	-116.92032			REACH 4B	SR-KS4-1B	59.58887	-111.42283	9/9/2021	
		HR-KS5-2B	60.01096	-116.92088				SR-KS4-2B	59.58975	-111.42273	9/9/2021	
		HR-KS5-3B	60.01125	-116.92177				SR-KS4-3B	59.58995	-111.42268	9/9/2021	
		HR-KS5-4B	60.01138	-116.92274				SR-KS4-4B	59.5909	-111.42261	9/9/2021	
		HR-KS5-5B	60.01163	-116.92348				SR-KS4-5B	59.59139	-111.42264	9/9/2021	
	REACH 6	HR-KS6-1B	60.02772	-116.92342			REACH 6	SR-KS6-1B	60.02772	-116.92342	9/10/2021	
		HR-KS6-2B	60.02779	-116.92217				SR-KS6-2B	60.02779	-116.92217	9/10/2021	
		HR-KS6-3B	60.02785	-116.92155				SR-KS6-3B	60.02785	-116.92155	9/10/2021	
		HR-KS6-4B	60.02787	-116.92075				SR-KS6-4B	60.02787	-116.92075	9/10/2021	
		HR-KS6-5B	60.02802	-116.91985				SR-KS6-5B	60.02802	-116.91985	9/10/2021	
							REACH 5	SR-KS5-1A	59.71284	-111.50644	9/10/2021	
								SR-KS5-2A	59.71304	-111.50646	9/10/2021	
								SR-KS5-3A	59.71823	-111.50577	9/10/2021	
								SR-KS5-4A	59.71853	-111.50594	9/10/2021	
								SR-KS5-5A	59.67804	-111.48615	9/10/2021	

Table 10. Summary of metal water chemistry parameters sampled in the Slave River in 2021 at six sample reaches, indicating site mean ± standard deviation for Reach SR-KS3 where duplicate samples were collected, and the single sample value for all other reaches. For Reach SR-KS3, the detection limit (DL) is presented for parameters for which both samples were below DL, and when only one sample was below DL (only for total mercury and dissolved selenium), the value above DL is presented. Values in bold were greater than CCME long-term exposure guidelines for the protection of aquatic life (Canadian Council of Ministers of the Environment 2001b). Reaches are ordered from upstream (KS1) to downstream (KS5).

Parameter	SR-KS1	SR-KS2	SR-KS3	SR-KS4A	SR-KS4B	SR-KS6	SR-KS5
Aluminum Diss. (µg/L)	1.70	2.00	2.20 ± 0.00	2.50	2.30	2.40	2.30
Aluminum Total (µg/L)	**710.0**	**622.0**	**584.5 ± 50.20**	**607.0**	**770.0**	**831.0**	**414.0**
Antimony Diss. (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Antimony Total (µg/L)	0.10	0.10	<0.1	<0.1	<0.1	<0.1	<0.1
Arsenic Diss. (µg/L)	0.40	0.40	0.40 ± 0.00	0.40	0.40	0.40	0.30
Arsenic Total (µg/L)	1.00	0.90	0.85 ± 0.07	0.90	1.00	1.00	0.70
Barium Diss. (µg/L)	37.10	36.50	35.35 ± 1.34	35.20	36.30	33.80	33.00
Barium Total (µg/L)	53.20	49.80	47.55 ± 1.06	47.30	54.10	54.20	44.00
Beryllium Diss. (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Beryllium Total (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Bismuth Diss. (µg/L)	<0.2	<0.2	<0.2	<0.2	<0.2	<0.2	<0.2
Bismuth Total (µg/L)	<0.2	<0.2	<0.2	<0.2	<0.2	<0.2	<0.2
Boron Diss. (µg/L)	16.60	15.80	15.95 ± 0.92	15.80	16.40	15.00	14.60
Boron Total (µg/L)	19.40	18.10	17.95 ± 0.78	18.00	18.70	17.80	16.20
Cadmium Diss. (µg/L)	<0.04	<0.04	<0.04	<0.04	<0.04	<0.04	<0.04
Cadmium Total (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Cesium Diss. (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Cesium Total (µg/L)	0.20	0.20	0.15 ± 0.07	0.10	0.20	0.20	0.10
Chromium Diss. (µg/L)	0.10	<0.1	<0.1	<0.1	<0.1	<0.1	0.30
Chromium Total (µg/L)	1.00	0.90	0.85 ± 0.07	0.90	1.20	1.30	0.70
Cobalt Diss. (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Cobalt Total (µg/L)	0.60	0.50	0.40 ± 0.00	0.40	0.60	0.60	0.30
Copper Diss. (µg/L)	0.70	0.70	0.75 ± 0.07	0.70	1.20	0.70	0.60
Copper Total (µg/L)	1.80	1.70	1.65 ± 0.07	1.60	1.90	2.00	1.50
Iron Diss. (ug/L)	15.0	14.0	13.0 ± 0.00	14.0	14.0	13.0	17.0
Iron Total (µg/L)	**1340.0**	**1170.0**	**1055.0 ± 21.21**	**1060.0**	**1410.0**	**1450.0**	**784.0**
Lead Diss. (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Lead Total (µg/L)	0.70	0.70	0.60 ± 0.00	0.60	0.80	0.80	0.40
Lithium Diss. (µg/L)	4.80	4.50	4.45 ± 0.21	4.30	4.30	4.20	3.90
Lithium Total (µg/L)	5.70	5.20	5.10 ± 0.14	5.00	5.40	5.40	4.70
Manganese Diss. (µg/L)	0.20	0.20	0.25 ± 0.07	0.20	0.20	0.20	0.30
Manganese Total (µg/L)	55.20	46.60	43.30 ± 1.56	42.30	53.10	50.30	32.10
Mercury Diss. (UL) (ng/L)	0.30	<0.2	<0.2	0.30	<0.2	<0.2	<0.2
Mercury Total (UL) (ng/L)	2.70	2.30	2.00 ± 0.00	2.00	3.30	2.30	2.10

Parameter	SR-KS1	SR-KS2	SR-KS3	SR-KS4A	SR-KS4B	SR-KS6	SR-KS5
Mercury Diss. (µg/L)	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Mercury Total (µg/L)	0.020	0.010	0.010	0.010	<0.01	<0.01	<0.01
Molybdenum Diss. (µg/L)	0.60	0.60	0.55 ± 0.07	0.50	0.50	0.50	0.50
Molybdenum Total (µg/L)	0.60	0.60	0.55 ± 0.07	0.60	0.60	0.60	0.50
Nickel Diss. (µg/L)	0.80	0.80	0.75 ± 0.07	0.80	0.80	0.70	1.00
Nickel Total (µg/L)	2.10	1.90	1.75 ± 0.07	1.80	2.20	2.30	1.50
Rubidium Diss. (µg/L)	0.80	0.80	0.80 ± 0.00	0.80	0.80	0.70	0.70
Rubidium Total (µg/L)	2.60	2.30	2.20 ± 0.14	2.30	2.70	2.70	1.80
Selenium Diss. (µg/L)	0.30	0.40	0.40	<0.3	0.40	0.30	<0.3
Selenium Total (µg/L)	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5	<0.5
Silver Diss. (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Silver Total (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Strontium Diss. (µg/L)	125.0	120.0	117.0 ± 5.66	115.0	120.0	110.0	107.0
Strontium Total (µg/L)	131.0	122.0	121.5 ± 3.54	119.0	127.0	120.0	113.0
Thallium Diss. (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Thallium Total (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Tin Diss. (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Tin Total (µg/L)	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
Titanium Diss. (µg/L)	0.20	0.10	0.10 ± 0.00	0.20	0.20	0.10	0.20
Titanium Total (µg/L)	15.40	13.00	12.10 ± 1.41	13.70	17.10	16.00	10.00
Uranium Diss. (µg/L)	0.30	0.30	0.25 ± 0.07	0.20	0.20	0.20	0.20
Uranium Total (µg/L)	0.30	0.30	0.30 ± 0.00	0.30	0.30	0.30	0.30
Vanadium Diss. (µg/L)	0.20	0.20	0.20 ± 0.00	0.20	0.20	0.20	0.20
Vanadium Total (µg/L)	2.30	2.10	2.00 ± 0.14	2.10	2.50	2.70	1.50
Zinc Diss. (µg/L)	<0.4	<0.4	<0.4	<0.4	<0.4	<0.4	<0.4
Zinc Total (µg/L)	<5	<5	<5	<5	5.10	5.70	<5

Table 11 BMI names and abbreviations used in PCA ordinations.

Order/Group	Family	Subfamily	Code
Amphipoda			AMPH
Bivalvia	Pisidiidae		PISID
Coleoptera	Elmidae		C_Elm
Diptera	Ceratopogonidae		D_Cerat
Diptera	Chironomidae	Chironominae	D_C_Chir
Diptera	Chironomidae	Diamesinae	D_C_Dia
Diptera	Chironomidae	Orthocladiinae	D_C_Orth
Diptera	Chironomidae	Prodiamesinae	D_C_Pro
Diptera	Chironomidae	Tanypodinae	D_C_Tany
Diptera	Diptera Pupa		D_Pupa
Diptera	Empididae		D_Emp
Diptera	Simuliidae		D_Simu
Diptera	Tabanidae		D_Tab
Diptera	Tipulidae		D_Tipu
Ephemeroptera	Acanthametropodidae		E_Acan
Ephemeroptera	Ameletidae		E_Amel
Ephemeroptera	Ametropodidae		E_Amet
Ephemeroptera	Baetidae		E_Bae
Ephemeroptera	Caenidae		E_Cae
Ephemeroptera	Ephemerellidae		E_Ephe
Ephemeroptera	Ephemeridae		E_Eph
Ephemeroptera	Heptageniidae		E_Hept
Ephemeroptera	Isonychiidae		E_Iso
Ephemeroptera	Leptophlebiidae		E_Lept
Ephemeroptera	Metretopodidae		E_Met
Gastropoda			GAST
Hemiptera	Corixidae		H_Corix
Hirudinea	Glossiphoniidae		GLOSS
Odonata	Aeshnidae		O_Aesh
Odonata	Gomphidae		O_Gomph
Oligochaeta	Enchytraeidae		ENCHY
Oligochaeta	Lumbriculidae		LUMB
Oligochaeta	Naididae		NAID
Plecoptera	Capniidae		P_Cap
Plecoptera	Chloroperlidae		P_Chl
Plecoptera	Perlidae		P_Perli
Plecoptera	Perlodidae		P_Perlo
Plecoptera	Pteronarcyidae		P_Pter
Trichoptera	Brachycentridae		T_Bra
Trichoptera	Hydropsychidae		T_Hpsy
Trichoptera	Hydroptilidae		T_Hpti
Trichoptera	Lepidostomatidae		T_Lepi
Trichoptera	Leptoceridae		T_Lepto
Trichoptera	Limnephilidae		T_Limn
Trichoptera	Philopotamidae		T_Phil
Trichoptera	Polycentropodidae		T_Poly
Trichoptera	Trichoptera Pupa		T_Pupa

Figure 44. Pictures of sample locations, including (A) upstream view from Hay River Reach 1, (B) downstream view from Hay River Reach 1, (C) upstream view from Slave River Reach 1, and (D) downstream view from Slave River Reach 1. Photos taken in 2017.