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

ASSESSMENT OF RESULTS FROM THE FOUR YEAR OF SAMPLING

2020

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 2022

Executive Summary

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. The establishment of this long-term monitoring program 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. The initial focus of the 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. Within the Northwest Territories-Alberta transboundary river regions, there is relatively little information about the current composition and natural variation of benthic communities. Therefore, it is vital that routine monitoring be established to secure information about current conditions in these assemblages and to provide sufficient information to allow for future detection of change.

The assessment of baseline monitoring data from the transboundary monitoring program is focused on quantifying spatial variation in BMI assemblages within and among reaches, and estimating the normal range to characterize the degree of temporal variability that is to be expected given current conditions in a system. The boundaries of the normal range provide a measure of the level of change that would be deemed significant enough to be ecologically relevant, termed the critical effect size (CES). This approach provides an ideal method to characterize natural variability at the GNWT-GOA border in the absence of significant impacts and identify the magnitude of change in future conditions that would require additional monitoring/assessment and potentially management action. Initial establishment of CES to quantify within-year and temporal variability can be done with three or more years of 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. Assessment of temporal data allows for the development of location-specific CES that can be used in future years to detect deviations from normal range. Assessment of the third year of sampling data (2019) was focused on developing CES for biotic metrics (univariate measures of composition), and this report additionally assesses the possibility of developing normal range criteria for the full assemblage using multivariate methods.

River flow has been a significant source of variability in habitat conditions and a constraint on sampling efforts through the first four years of the GNWT and GOA large transboundary river BMI monitoring program, which have focused on sampling the Hay River and Slave River in northern Alberta/southern Northwest Territories. In the first two years of the program, sampling in the late summer was required for the Hay River to ensure there was sufficient flow to access shallow portions of the river, and sampling in early fall was selected for the Slave River to ensure water levels had receded enough to allow access to rocky shorelines. However, in 2019, high water levels in both rivers made sampling difficult, and in 2020, the continuation of high flows made it impossible to access the Hay River for sampling and severely limited the number of sites in the Slave River that could be sampled.

The objective of this report is to assess monitoring data from the Slave River from 2020 and re-evaluate initial estimates of normal range and CES with the addition of 2020 data, as well as to use data from both the Hay River and Slave River to explore the development of multivariate normal range and CES, and to evaluate flow variability and flow-ecology relationships in the rivers. The Slave River was sampled in October 2020. Water chemistry, sediment chemistry, physical habitat, and BMI kick samples were

analyzed to characterize spatial and temporal variability within the river, including quantification of the normal range and CES for a number of biotic metrics using data from 2017-2020. Multivariate approaches were used to explore potential assemblage-level measures of temporal variability for both rivers (2017-2019 for the Hay River, and 2017-2020 for the Slave River). In addition, hydro-ecological variables were evaluated for each river and flow-ecology relationships were assessed to identify community change thresholds in response to velocity. These additional measures will provide more tools to assess future change in these rivers and aid in the detection of any future impairment.

Methods

During the fourth year of sampling, the Slave River was sampled from October 5-7, 2020. Kick-sampling reaches of approximately 500 m in length were sampled in the river. Sampling took place in each reach on the river bank where rocky habitat was located. Five sites were selected within each reach, spaced evenly along the reach when habitat availability allowed, but only 18 of the 35 sites in the river could be accessed in 2020 (located in 6 of the 7 reaches in the Slave River). Sites had similar substrate composition, dominated by pebble, gravel, and cobble size classes. Sample collection followed the Canadian Aquatic Biomonitoring Network (CABIN) sampling protocol modified for large rivers, as described in the monitoring plan (see Lento 2018). Water chemistry, physical habitat descriptions, and sediment chemistry samples were collected at a subset of kick sites as supporting variables.

Water chemistry, sediment chemistry, and biotic metrics were summarized by reach, and chemical parameter means were compared with CCME guidelines. Multivariate analysis (Principal Component Analysis or PCA) was used to fully characterize the biotic assemblage and abiotic environment of each river, and PERMANOVA was used to test for differences in biotic composition between reaches. The relationship between the BMI data from kick samples and abiotic data was tested with Redundancy Analysis (RDA), with a subset of abiotic parameters selected for inclusion based on their importance in the abiotic PCA.

Temporal patterns in biota were explored through the comparison of biotic metrics among years at the site and reach level. The normal range of variability was assessed for within-year variability and temporal variability among sites and within reaches. Within-year variability was tested by developing single-year and multi-year CES limits and comparing 2020 data with those estimates of normal range. At the site scale, temporal variability was assessed by plotting 2017-2020 means ± standard error for each site and comparing with the grand mean (mean of means) ± 2 standard deviations (2SD) for the river. Within-reach variability was quantified by calculating the grand mean (mean of annual means) ± 2SD for each reach in each metric, as a measure of reach-specific normal range.

Multivariate temporal patterns were initially explored for both the Hay River and Slave River by creating a PCA ordination with all years of data included, and overlaying 95% normal probability ellipses for each year. Any overlap in ellipses indicated similarity in assemblages between years, while ellipses that did not overlap indicated that composition differed between years. At the site scale, temporal multivariate patterns were assessed by using Procrustes analysis to compare PCA ordinations between years. The sum of squared residuals for each pairwise comparison of years was plotted with Principal Coordinates Analysis (PCoA) to examine change trajectories over time, and site residuals for each pairwise comparison of years were used to develop multivariate normal range and CES.

Environmental flow components were calculated for the Hay and Slave Rivers using the full hydrologic record (1963-2020 and 1960-2020, respectively), a 30-year record (1990-2020), and the period of sampling (2017-2020) to assess flow variability during the sampling period in the context of the longer

record. Hydro-ecological variables were calculated and assessed to evaluate long-term patterns in flow, and identify possible flow-related groupings of years for CES. TITAN2 analysis was then used to evaluate community change thresholds in response to a velocity gradient (a site-scale proxy for variability in discharge) using data from all sampling years, and indicator taxa were identified.

Results and Discussion

Slave River

Spatial variation in physical/chemical habitat

Water samples were collected in each reach of the Slave River at one to three sites and analyzed for ions, nutrients, and physicals. Mean levels of water chemistry parameters were compared with Canadian guidelines, but there were no exceedances of long-term guidelines for any of the water chemistry parameters for which guidelines exist. Most water quality parameters had similar values to those observed in 2019, including alkalinity, conductivity, and nutrients. Some parameters like TDS and turbidity were lower in 2020, likely due to the less variable and less flashy flow conditions in 2020. Flow remained high throughout the spring and summer of 2020, and sediment transport was likely steady throughout this period. Estimates of mean TP in the Slave River were all lower than 0.100 mg/L, and reaches were classified as eutrophic based on the Canadian Guidance Framework. Water samples were also analyzed for metals, and both dissolved and total metal concentrations were generally found to be low. No dissolved metals exceeded long-term exposure water quality guidelines for the protection of aquatic life. Total aluminum concentrations exceeded long-term exposure water quality guidelines, but reach averages were much lower than the long-term median reported for the river, and were lower than concentrations observed in 2019.

The PCA ordination of water chemistry and physical habitat variables indicated a similar contribution of both types of variables to the environmental gradients along which reaches varied. Whereas low variability in water chemistry in 2019 resulted in stronger loadings of physical habitat variables on the first and second axes, axis loadings were more similar for chemical and physical habitat variables in 2020.

Sediment chemistry samples were collected in the Slave River and analyzed for metals and polycyclic aromatic hydrocarbons (PAHs). To determine whether levels of metals or PAHs were elevated beyond recommended levels in Slave River samples, mean values for each site were compared with CCME sediment quality guidelines for the protection of aquatic life, which include interim freshwater sediment quality guidelines (ISQGs) and probable effect levels (PELs). Concentrations of most metals in sediments were below the guidelines for the protection of aquatic life. Arsenic was the only metal to exceed the ISQG, but all levels remained below the PEL, and arsenic levels in all reaches were similar to (and slightly lower than) those observed in 2019, indicating that this did not reflect an increase from the previous year. Average concentrations for PAHs in sediments were generally low, and many PAHs were below detection limits. However, concentrations of 2-methylnaphthalene were elevated above the ISQG in four reaches, including the three downstream reaches. 2-methylnaphthalene represents an acutely toxic species for benthic organisms. In Reach 6, phenanthrene (another PAH with acute toxicity) was also found to exceed the ISQG. 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. Total PAHs measured in the Slave River samples were below recommended guidelines.

Spatial variation in BMI assemblages

Biotic metrics were used to compare abundance, relative abundance, and taxonomic richness of key organism groups among sites and reaches in the Slave River. Total abundance was high in Slave River samples in 2020, particularly compared to the low abundances observed in 2019. The highest total abundances were driven by extremely high abundance of the genus *Hydra*, a freshwater cnidarian that is related to sea anemone, jellyfish, and corals. *Hydra* were present at unusually high abundances across most reaches, and the genus accounted for 51.2 to 91.5% of individuals in samples on average in Reaches 2, 3, 4B, 6, and 5 (it only made up 19.3% of the sample on average in Reach 1). Thus, although abundances of Ephemeroptera, Plecoptera, and Trichoptera (EPT) and the midge Chironomidae were high in 2020, they made up a very small proportion of the samples in these reaches.

The genus *Hydra* is included as part of CABIN sample enumeration, but is not often the focus of research on BMI assemblages, because it is rarely the dominant taxon in kick samples. In international sampling and sorting protocols for BMI, *Hydra* are often excluded, as their numbers are considered to be underestimated when sampled with a kick net. As a result, little research exists on freshwater *Hydra* in the context of BMI community structure, and there is much about the importance of this group to benthic ecology that is unknown. Hydroids are clonal organisms that can form multiserial colonies that spread laterally and form benthic “animal forests”, becoming part of the benthic habitat and adding complexity to that habitat, altering flows and light penetration across the habitat, and providing food and shelter to other benthic organisms. They feed on zooplankton in the water column and can easily adapt to different environmental conditions, and respond to a variety of stressors through adaptation of their growth, reproduction, and behaviour. The extremely high abundance *Hydra* in Slave River samples in 2020 may have reflected their ability to adapt to the changing hydrologic conditions and increased flows in both 2019 and 2020. Moreover, a review of the data collected in the Slave River in previous years indicates that *Hydra* have consistently been common in the benthic samples, and that their abundance has changed with variation in flows.

The PCA identified an extremely strong gradient along the first axis that explained 94.3% of the variability in assemblage structure among sites. The first axis gradient separated sites in Reach 1 from all other sites due to a positive correlation of all other reaches with *Hydra*. Along the second PCA axis, which explained an additional 3.2% of the variation among sites, there was separation among the sites in Reach 1 that was primarily due to differences in the composition of EPT taxa. Sites in Reach 1 were associated with taxa that have adaptations for fast flows, including the caddisfly Hydropsychidae, the mayflies Ephemerellidae and Heptageniidae, and the stonefly Perlodidae. Similar to the PCA, the constrained RDA ordination primarily separated sites in Reach 1 from most sites in other reaches along the first axis, which explained 66.9% of unconstrained variation in assemblage structure (97.1% of constrained variation). Sites in Reach 1 were positively correlated with velocity, conductivity (and thus metals, which were positively correlated with conductivity), and % pebble.

Temporal variation in BMI assemblages

Compositional changes from 2017 to 2020 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. The relative abundance of *Hydra* was highest in 2018, when it accounted for more than one-quarter of the assemblage on average, and in 2020, when it accounted for more than half of the assemblage on average. The relative abundance of *Hydra* was similar and much lower in both 2017 and 2019, when antecedent median discharge was lower. Also apparent was the decline in relative abundance of Diptera (true flies) from 2017 to 2018, and the concurrent increase in relative abundance of Trichoptera (caddisflies) and of other mobile taxa such as Ephemeroptera (mayflies), Plecoptera (stoneflies), and

Hemiptera (true bugs). Average relative abundance of several taxonomic groups were similar in 2019 to what was found in 2018, particularly for true flies. With the dramatic increase in abundance of *Hydra* in 2020, relative abundances of all other taxonomic groups declined.

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. Mean abundance appeared to increase in all reaches in 2020 relative to earlier years, and 2017-2019 appeared to be relatively invariable compared to the change in 2020. The increased abundance was clearly driven by the increase in abundance of *Hydra*. The abundance of Chironomidae showed a decline from 2017 to 2018 and a slight increase in 2020. Relative abundance metrics showed more variability over time than abundance metrics. The relative abundance of EPT was particularly variable, with little consistency among reaches with the exception of a sharp decline in 2020 in most reaches. Richness metrics showed similar trends over time in most reaches. Total richness declined in 2018 and again in 2020, coinciding with the loss of Chironomidae in 2018 and the increased abundance of *Hydra* in 2020. The richness of EPT remained similar from 2017-2019 in most reaches, but declined in all reaches in 2020. Together, these patterns point to a loss of diversity in 2020 when water levels were high, and *Hydra* dominated most reaches.

Normal range and CES

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. 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.

Samples collected in the Slave River in 2020 generally fell within the CES boundaries developed using only 2020 data, though the boundaries were also large for some metrics due to variability among reaches. For example, total abundance was much higher in Reach 2 and Reach 3 than in the other reaches, and the CES therefore ranged from 0 to 12000. In contrast, the normal range was narrow for the relative abundance of both Chironomidae and Diptera+Oligochaeta, both of which were low across all reaches. Normal range boundaries were also more narrow for richness metrics, as variability in each of these metrics was much lower across all reaches. In particular, the normal range for EPT richness was low, varying between approximately 4 and 8 genera. Comparison of 2020 data with the multi-year CES developed based on data from 2017-2020 provided a better picture of how variable samples collected in 2020 were relative to previous years. There were a large number of deviations from the normal range when 2020 data were compared with the multi-year CES, particularly for total abundance, EPT abundance, and *Hydra* abundance. In contrast, both the abundance of Chironomidae and abundance of Diptera+Oligochaeta generally fell within the narrow bounds of the multi-year CES. This comparison of 2020 data with multi-year CES for Chironomidae and Diptera+Oligochaeta abundance metrics suggests there is possible utility in the use of these metrics for quantifying normal range in the river, despite the large differences in abundance that were observed from 2017-2018. Abundances of both groups have been relatively stable since then. Relative abundance metrics most clearly showed the effect of the high abundance of *Hydra* in 2020, with large deviations from normal range evident. The relative utility of relative abundance and absolute abundance, the latter of which was formerly considered to be the more variable, differed greatly in 2020 when there was such a strong change in composition and abundance.

Temporal variation at the site scale was assessed by comparing the 2017-2020 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, and 2020 means of all sites) ± 2SD. Annual means across all sites in the Slave River varied widely between 2017 and 2020 for most metrics, contributing to wide preliminary normal ranges for the river. Only the boundaries for Chironomidae abundance were narrow, suggesting some possible utility to this metric; however, as the lower bound was at an abundance of zero, it is important to note that this metric may only be useful for detecting increases in abundance. This is not surprising, given the low abundance of Chironomidae in the last three years of sampling. Relative abundance metrics were much more variable, both in terms of the width of the estimated normal range and in terms of site means ± SE, which often covered nearly the full range of possible values. The preliminary estimate of the normal range of variability in site-scale analyses was more narrow for richness metrics, but did suffer from wide inter-annual variability in site means across the entire river. The exception was EPT richness, which had an extremely narrow preliminary normal range based on the grand mean.

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. 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 calculated normal range. At the reach scale, there was a great deal of variability in terms of the width of the estimated normal range for each abundance-based metric, reflecting the large changes in abundance in 2020 for some reaches. With the exception of Reach 1, all other reaches had very narrow boundaries for EPT abundance, which suggested that this metric might work well to detect temporal changes at the reach scale. The preliminary normal range for Chironomidae abundance was narrow across most reaches as well, and could be considered as a useful metric as more data are collected and the normal range is refined. 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. Preliminary estimates of the normal range of variability were somewhat more narrow for richness-based metrics, 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. Total richness and Chironomidae richness also had fairly low variability in several reaches. Much of the variability observed in richness metrics was due to higher richness in most reaches in 2017, and additional years of data will likely help to refine these preliminary normal range estimates.

Multivariate normal range

Hay River

The PCA of Hay River samples from 2017-2019 indicated a strong overlap of samples from 2017 and 2018, and a greater spread and separation of samples from 2019. The 95% normal probability ellipses for 2017 and 2018 overlapped, indicating similar composition of BMI in kick samples in these two years, whereas the ellipse for 2019 only partially overlapped with the two earlier years, indicating a change in composition in 2019.

Procrustes analysis indicated that the degree of change in the relative position of sites was small and similar for all pairwise comparisons of years. An initial estimate of a multivariate normal range and CES boundaries was created based on the grand mean Procrustes residual (mean of mean residuals for each pairwise year comparison) ± 2 SD, and mean ± SE Procrustes residuals were plotted for each site to identify site-scale temporal variability relative to the normal range. Because the degree of dissimilarity

in pairwise comparisons was similar across years, the normal range was narrow, and the mean residual for 11 of the 25 sites fell outside of CES boundaries (above or below CES). Of these sites, 6 had a mean residual above the upper CES, indicating greater temporal change in BMI composition than expected based on the normal range. Though some sites were more variable than others (larger SE), most sites in Hay River had low mean Procrustes residuals that were within or below the normal range of variability, indicating generally limited site-scale temporal multivariate change.

Slave River

Multivariate analysis of temporal variability in Slave River samples yielded stronger patterns that reflected the large compositional shifts that were described by temporal analysis of metrics. The PCA of all sampling years (2017-2020) highlighted the differences between 2017 samples, which included abundant and diverse Chironomidae assemblages, and samples from 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, overlapping 95% normal probability ellipses.

Procrustes results indicated greater temporal variability in the position of sites in multivariate space than was observed for the Hay River. Such variability would occur if BMI composition did not change in the same way in all sites from one year to the next, or if there were temporal changes in composition in some sites but not in others. The greatest shift in the relative position of sites in multivariate space occurred between 2019 and 2020, and that the relative position of sites in multivariate space in 2020 was more similar to what was observed in 2017. This pattern was reflected in the PCoA, which showed a large change in position from 2017 to 2018, followed by shift along the first axis in 2019, and a larger shift in 2020 that brought the trajectory closer to 2017 than in previous years.

The normal range based on Procrustes residuals was calculated separately for 2017-2019 data and for 2017-2020 data. For the period 2017-2019, when more sites were sampled, the normal range was narrow and similar to that observed for the Hay River. However, in contrast to the Hay River, the mean residuals for few sites fell within the normal range, error bars were wide for most sites, and the means for 11 sites were above the upper CES (indicating greater change across years than expected based on the normal range). The high variability at the site scale is not surprising, given the strong compositional changes that were observed across years. Assessment of initial normal range for 2017-2020 was limited to only 17 sites that were sampled across all four years. The normal range for this subset of sites over all four years of sampling was much wider, which indicated greater temporal variability in the spatial arrangement of sites in multivariate space in the reduced ordinations. However, 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, when nearly twice as many sites were included in the ordinations.

Ecological response to flow conditions

Hay River

The long-term hydrograph for the Hay River showed a history of periodic small floods and large floods, with the periods of highest flow (large floods, exceeding the peak flows of 2020) occurring prior to 1990. In the context of these earlier extreme high flows (discharge greater than 1000 m^3/s), the discharge in 2020 was categorized as a small flood, similar to other high-flow events in the last 30 years. However, the small flood event in 2020 represented a greater shift in flow conditions because peak flows were relatively low in the years prior. The change in peak flows from 2019 to 2020 was greater than that

observed in most previous years, with the exception of the large flood in 2013. Furthermore, peak flows in 2017 and 2018 were relatively low compared to previous years, and the spring freshets in 2018 and 2019 were preceded by extreme low flows. The increase in flows was also faster than in previous years, as the rise rate plotted from 1990-2020 indicated that 2020 had the fastest rise rate in the 30-year record.

The ecological response to flow was assessed by using TITAN2 analysis to identify changes in composition across the gradient of flow velocity in all samples. TITAN2 identifies taxa that are associated with fast flow velocity (positive responders) and those associated with slower flow velocity (negative responders) and identifies the point along the velocity gradient at which there is the greatest change in composition (community response threshold). In the TITAN2 analysis of Hay River BMI, there were 6 taxa identified as pure and reliable negative responders to flow (i.e., consistently associated with slower velocities) and 9 taxa identified as pure and reliable positive responders to flow (i.e., consistently associated with faster velocities), out of 41 possible taxa. There was some overlap of the density probability plots for negative and positive responders, but community response peaks were clearly different. The results provide information about the velocity at which there is a change in the assemblage, as well as the taxa that are associated with slow or fast flow conditions in the river.

Slave River

The hydrograph for the full record of Slave River data (1960-2020) indicated more variable flows with higher peak discharge in the earlier part of the data record, and the high flows in 2020 were characterized as a small flood in the context of the larger record. However, peak discharge in 2020 was the highest observed since 1997 in the Slave River. Though the magnitude of peak flows in 2020 was not unusual for the long-term hydrograph, such high flows have not been observed in the river for over two decades. Though the assessment indicated that flows have been variable across the period of sampling, it did not provide insight into the strong changes in assemblage composition observed between 2017 and 2018 in Slave River samples. The hydrograph did differ somewhat in 2017, with a plateau in the fall rather than the typical decline in flow, but peak flow was similar between 2017 and 2018, and flows were fairly regular in the river to that point. Variability in flow in 2018 may have contributed to the shift in assemblage composition, and the hydrograph in 2019 was similarly variable, but more years of data would be required to assess whether flow variability contributed to the change. Partitioning data based on flow magnitude alone would suggest that 2017 and 2018 should be used to create a normal range and CES for typical flow conditions in the river, but these two years differed greatly in assemblage composition. Furthermore, the TITAN2 analysis did not indicate a clear community change threshold related to velocity. It is possible, therefore, that flow-based CES will not be possible for this river without additional data and additional examples of assemblages that are typical for low and high flow years.

Recommendations and Conclusions

Despite the strong influence of *Hydra* on the 2020 data, there was some evidence of consistent patterns across the 2017-2020 period. Of the metrics tested for the Slave River, Chironomidae abundance and EPT richness appeared to have the greatest initial potential for developing monitoring and management triggers. Both metrics were relatively invariable across years (Chironomidae more so if 2017 was excluded), with a narrow normal range and CES boundaries. Abundance metrics (e.g., EPT abundance, Chironomidae abundance) and richness metrics generally appeared to be more effective for developing normal range criteria than relative abundance metrics, as the latter were highly influenced by *Hydra* in 2020.

Some of the variability in normal range estimates is due to inter-annual differences in assemblages, which could result from differences in flow conditions, temperature, or other environmental drivers from one year to the next, and this variability can be accounted for by developing different normal range criteria for different environmental conditions. The multi-year PCA with 95% normal probability ellipses provides a potential approach to identify the years that should be grouped together in the development of more precise biotic metric normal range and CES. For example, the PCA for Hay River sites indicated a strong similarity of 2017 and 2018, and indicated that composition differed in 2019. This suggests that the first two years of sampling could be grouped to create CES, whereas 2019 could be considered separately, once data are collected under similar conditions. The PCA for the Slave River indicated that 2018-2020 were most similar, while composition differed in 2017. While grouping 2018-2020 and excluding 2017 might not lead to more narrow normal range estimates for all metrics (for example, those that were affected by high abundance of *Hydra* in 2020), there was a strong similarity among data from 2018-2020 for a selection of biotic metrics that included Chironomidae and Diptera + Oligochaeta abundance, relative abundance, and richness, as well as other taxonomic richness metrics. The relationship between multivariate probability ellipses and biotic metric CES should continue to be examined as more data are collected, as the utility and composition of such groupings will likely evolve as more years of data are added.

Procrustes residuals should also continue to be explored as more data are collected. With additional years of data, it will be possible to continue to assess the degree of temporal stability in the composition of sites relative to one another, and potentially identify any sites that change composition substantially relative to previous years. This provides an additional tool to detect possible impairment based on the full assemblage, rather than a particular biotic metric.

Hydrologic characterization of the Hay River and Slave River emphasized that although the high flows observed in 2020 were similar to high discharge events in past years, other conditions related to the timing and magnitude of flows prior to 2020 contributed to the extreme conditions. For example, the high rate and magnitude of increase in discharge in the Hay River, which followed a period of lower flows in 2019, appeared to be unusual. Development of flow-specific normal range and CES for the Hay River could therefore include 2017 and 2018 as typical non-flood years (consistent with the grouping suggested by the multivariate analysis). The inclusion of data from 2019 may not be advisable, given the delayed timing of peak flows and its likely effect on assemblage composition. For the Slave River, 2020 represented the highest peak flow since the late 1990s. But the hydrograph did not highlight strong differences between 2017 and 2018 that might have contributed to differences in BMI assemblage composition. Furthermore, partitioning data based on flow magnitude alone would suggest that 2017 and 2018 should be used to create a normal range and CES for typical conditions in the river, despite the fact that the BMI assemblage in 2017 appeared to differ from all other sampling years. Assessment of flow-ecology relationships also failed to identify strong associations with velocity for the Slave River, though this may simply indicate that site-scale spot velocity measurements are a poor proxy for inter-annual variability in discharge. It is possible, therefore, that flow-based CES will not be possible for the Slave River without additional data and additional examples of assemblages that are typical for low and high flow years.

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. 2019). 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. 2022). With climate change, it is predicted that biodiversity in northern regions will begin to resemble more closely 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, 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 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.

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.

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.

1.4. Importance of Flow Variability

River flow has been a significant source of variability in habitat conditions and a constraint on sampling efforts through the first four 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 the first two years of the program, earlier sampling was required for the Hay River to ensure there was sufficient flow to access shallow portions of the river, and later sampling was selected for the Slave River to ensure water levels had receded enough to allow access to rocky shorelines. However, in 2019, high water levels in both rivers made sampling difficult, and in 2020, the continuation of high flows made it impossible to access the Hay River for sampling and severely limited the number of sites in the Slave River that could be sampled.

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.

Shifts in BMI assemblage composition in response to changes in flows might appear indicative of impacts if there is no quantification of the biotic response to natural variation in the flow regime of a system. This involves estimating variability in hydro-ecological variables that are relevant to BMI assemblages, quantifying community change points in response to flow velocity, identifying taxa that are more associated with low or high flows, and evaluating whether it is possible to develop normal range criteria and CES that are specific to either low flow or high flow conditions in a system. Such information will support the continued assessment of assemblage structure while controlling for changes that may be due to flow variability.

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 2020, it was not possible to sample the Hay River, and only a subset of the sites in the Slave River were accessible for sampling. Therefore, this report includes an assessment of spatial patterns in only the Slave River sampling data from October 2020 and temporal patterns in Slave River data from 2017 to 2020. 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 to updating spatial and temporal assessments of the Slave River, this report is focused on developing additional measures of normal range for both the Hay and Slave rivers, using data from 2017-2019 for the Hay River and data from 2017-2020 for the Slave River. First, there is an assessment of temporal variability using multivariate methods, as a means of developing multivariate measures of normal range and CES for the rivers. Second, there is a focus on assessing variability in hydro-ecological variables for the rivers and estimating ecological change thresholds in the Hay and Slave rivers related to flow. Given the variability in BMI assemblage structure across the first four years of sampling and the apparent relation to river flows, this assessment is aimed at determining whether separate estimates of

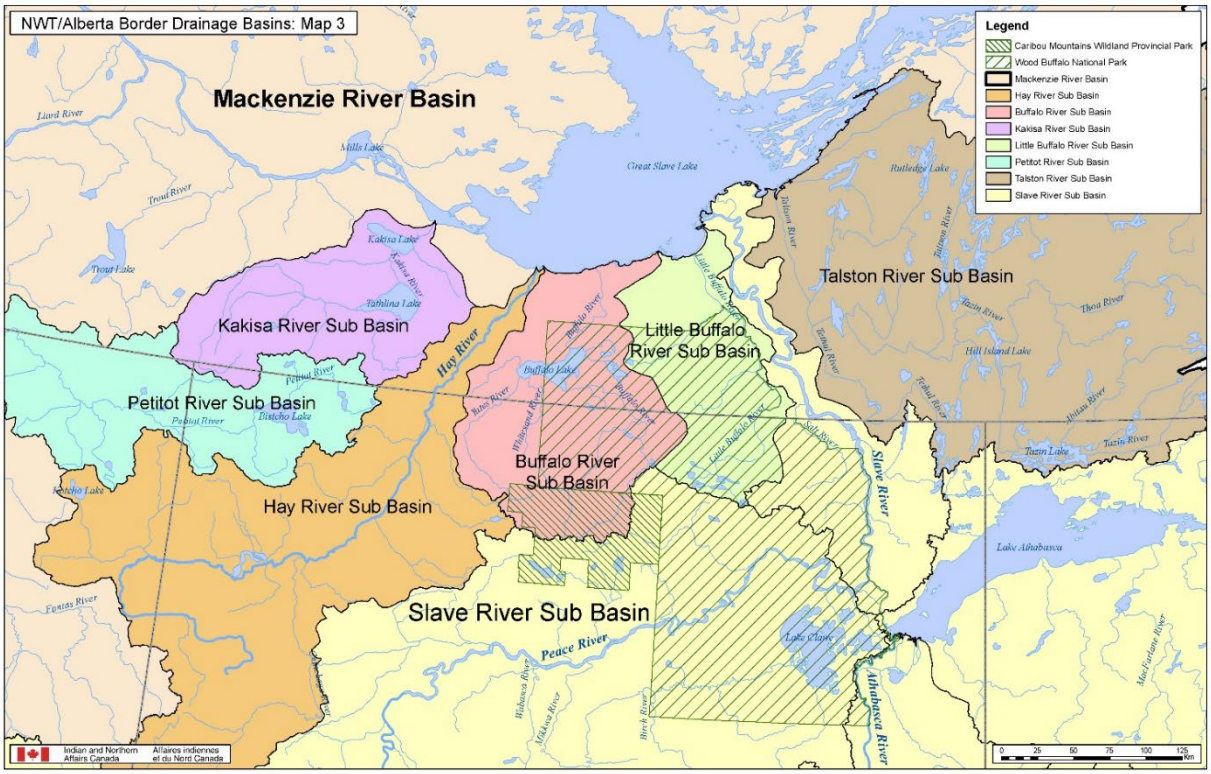

Figure 1. Drainage basins at the NWT/Alberta border, including the Hay River Sub Basin and Slave River Sub Basin. Map created by Indian and Northern Affairs Canada.

normal range and CES can be developed for low flow and high flow conditions, in order to improve accuracy of these measures of change. These additional measures will provide more tools to assess future change in these rivers and aid in the detection of any future impairment.

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 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

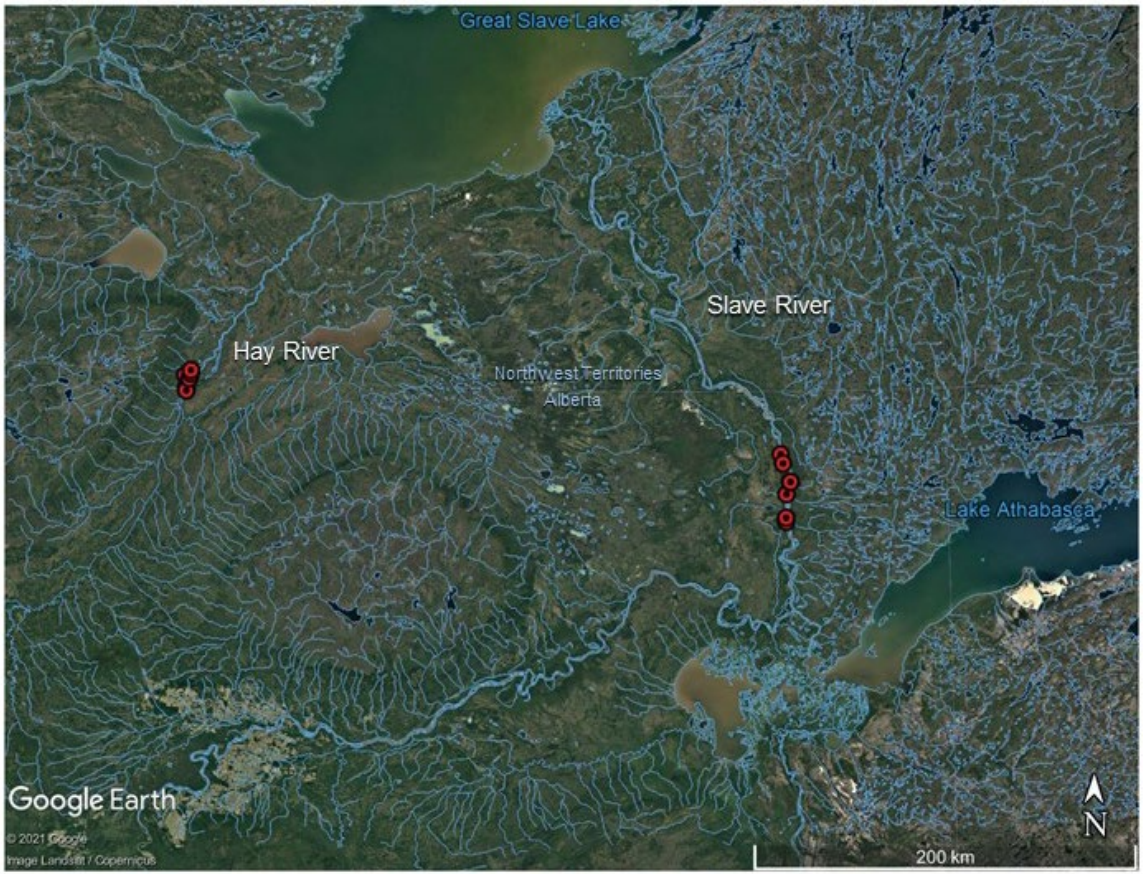

Figure 2. Map of Hay River and Slave River, showing kick-sampling reaches (red points) selected within the rivers, and an overlay of the stream network. In 2020, no sampling took place in the Hay Rive due to high water levels. Stream network layer from National Hydro Network (NHN) GeoBase Series (open.canada.ca).

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 and 2020 made sampling difficult or impossible in this river. Sampling was possible in the Slave River in all four 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 2020, it was not possible to sample the Hay River, and limited sampling took place in the Slave River from October 5-7.

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 characterize the river and achieve adequate power to detect biologically-meaningful differences among

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 2020. Only the Slave River was sampled in 2020, 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 2020	BMI sites sampled in 2020
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	1, 3, 5	1, 2, 3, 4, 5
	SR-KS2	59.4276	-111.4629	2, 3, 4	2, 3, 4
	SR-KS3	59.5350	-111.4577	1, 3, 5	1, 2, 3, 4, 5
	SR-KS4A	59.5912	-111.4195	None	None
	SR-KS4B	59.5903	-111.4225	1, 2, 3	1, 2, 3
	SR-KS6	59.6766	-111.4856	5	5
	SR-KS5	59.7182	-111.5058	5	5

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 four 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 (HR-KS4A and HR-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.

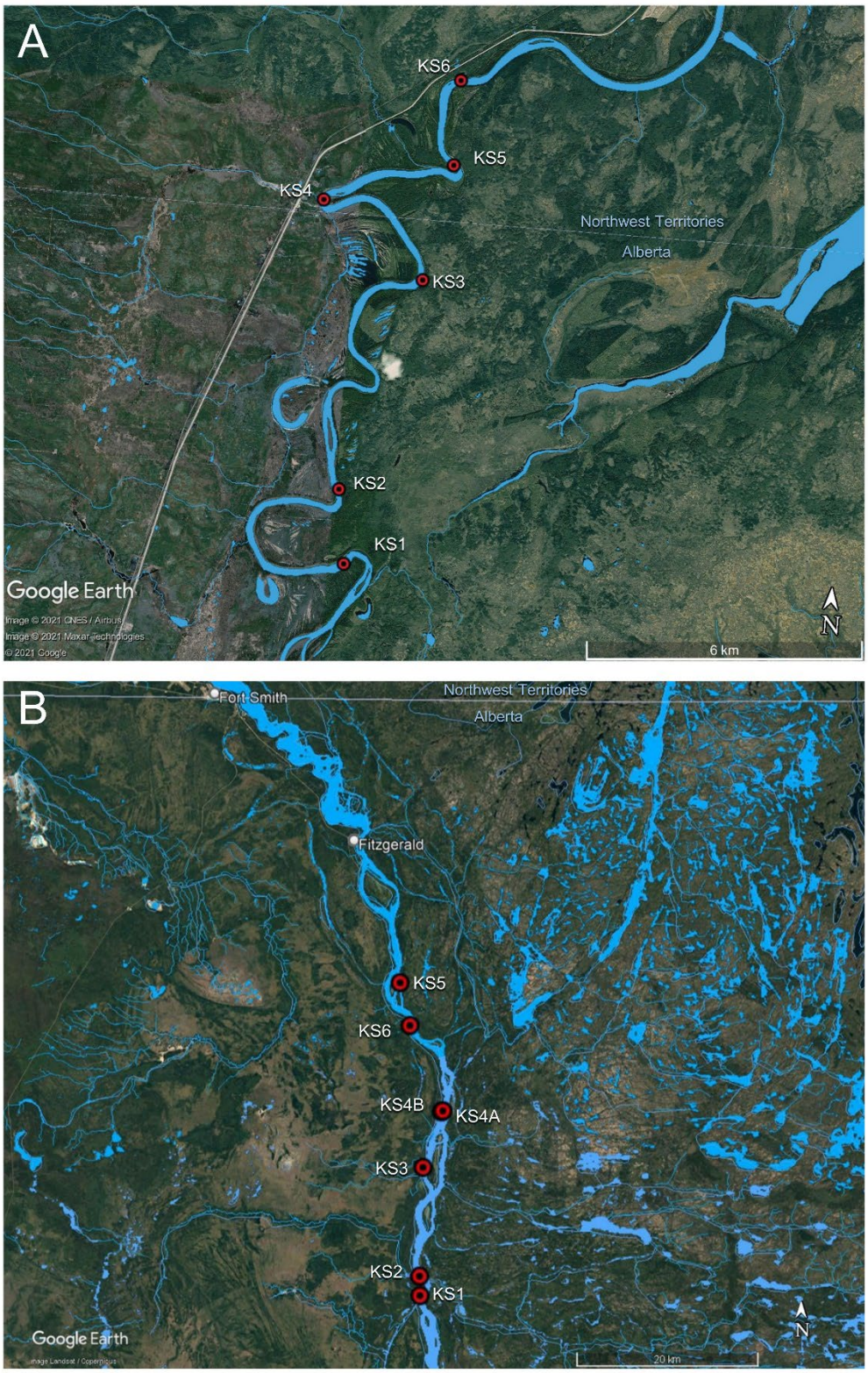

Figure 3. Kick-sample reaches (red points) in the (A) Hay River and (B) Slave River. Reaches are labeled in white text. No sampling was possible in the Hay River in 2020, and only a subset of sites within each reach of the Slave River was sampled, with the exception of Reach KS4A, which was not sampled. Water body and stream layers overlain on maps are from the National Hydro Network (NHN) GeoBase Series (open.canada.ca).

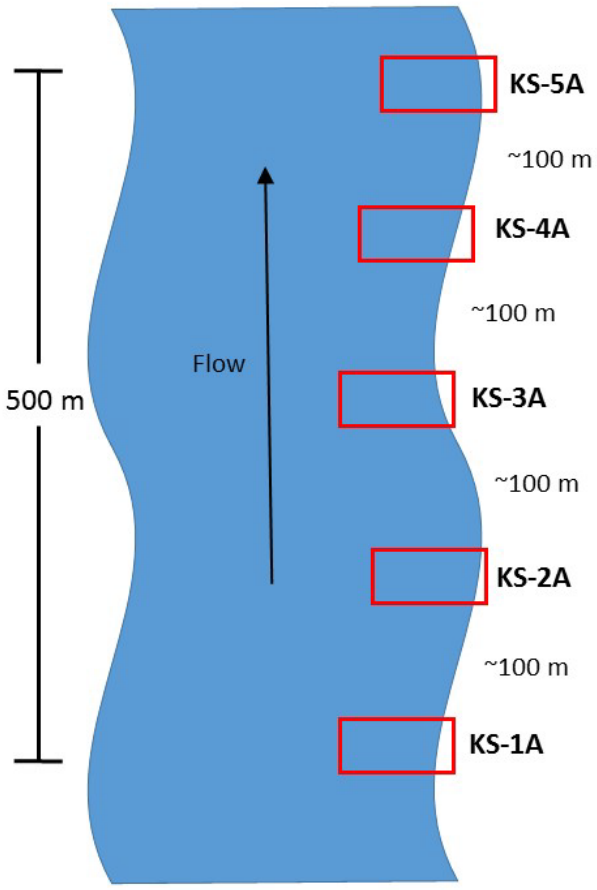

Figure 4. Example sampling design used for a single reach within the Hay River and Slave River, indicating the location of 5 sites within the 500 m reach and numbering of sites with respect to flow direction. Sampling of sites began downstream, at site KS-5A and worked upstream towards site KS-1A. Sites 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).

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).

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). Though sampling was not possible in the Hay River in 2020 and was only possible at a subset of sites in five of the Slave River reaches (Table 1), an overview of the full sampling scheme is provided here, with notes of modifications in 2020.

At three of the kick-sites in each reach (odd-numbered kick-sites; though see Table 1 for a list of adjustments to sampling 2020), water samples were collected for analysis of a standard suite of parameters, including nutrients, ions, and suspended solids. Additional water samples were collected for the analysis of metals (including mercury) at the same three kick-sites. Where water levels did not allow access to all sites in a reach, adjustments were made to the number and location of water chemistry samples as needed (Table 1). 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 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 at two sites in each reach (sites 1 and 5) and placed into jars. The number and location of sediment chemistry samples was adjusted in 2020 as needed, based on accessibility of sites. Sediment samples were kept cool and sent to ALS Labs for analysis.

BMI kick samples were collected at each kick-site (see Table 1 for details on sites sampled in 2020) 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.

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. 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

The logistical issues in 2020 that limited the amount of sampling that could take place have also had an impact on the analysis that can be completed on the data. For example, with no replication in Reach 5 or Reach 6 of the Slave River, and limited replication in two other reaches, some analyses were not possible, or would not have generated trustworthy results. Furthermore, the lack of sampling in the Hay River meant that an assessment of four years of data for that river was not possible. However, the high degree of variability that's been observed in these rivers across sampling years remains an interesting challenge, and there is an opportunity to conduct a more in-depth assessment of the data collected thus far (four years for the Slave River, and three years for the Hay River) to assess flow ecology relationships and explore new methods for characterizing assemblage variability, in addition to assessing the fourth year of data from the Slave River. The focus of this report is therefore to assess the 2020 Slave River data in relation to data from 2017-2019, but also to use data from both rivers to identify responses to flow variability and develop multivariate measures of normal range and CES. These additional analyses should inform assessments of data collected in these rivers in future years.

2.3.1. 2020 Hydrologic Conditions

High flows in both the Hay and Slave Rivers caused logistical issues for sampling in 2020, and it was not possible to sample the Hay River due to extremely high water levels. To characterize the extreme flows in 2020, 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 2020 sampling dates for the Slave River were used as a reference point for the Hay River to compare conditions between years.

2.3.2. Slave River Assessment

Data from the Slave River were analyzed in a similar manner to previous reports, including a spatial analysis of variability in the chemical/physical habitat and 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. However, as logistical constraints due to high flows led to only a portion of sites in the Slave River being sampled, there were some restrictions on the analyses that were possible, as noted below.

2.3.2.1. Spatial characterization of reaches

2.3.2.1.1. Chemical and physical habitat variation within and among reaches

Spatial variation in the chemical and physical habitat of the Slave River was assessed 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 mean ± standard deviation of chemical and physical habitat parameters for each reach. Water chemistry and sediment chemistry means were compared with CCME water and sediment quality guidelines, respectively (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, particularly given the high-flow conditions of 2020.

Box plots were used to present summaries of variation in water chemistry within and among reaches. Box plots were presented for a selection of parameters that displayed some variation among sites and where values were above detection limit for at least half of the sites (e.g., alkalinity, total suspended solids, conductivity, total phosphorus, total and dissolved nitrogen, aluminum, iron, manganese, and mercury). For creation of box plots, values below the detection limit were changed to be half the detection limit. Although variation among reaches was assessed for water chemistry in previous years using one-way analysis of variance, this analysis was not completed for 2020 data due to the lack of replication in some reaches (n = 3 for Reach 1, Reach 2, and Reach 3; n = 2 for Reach 4B; and n = 1 for Reach 5 and Reach 6). Because sediment chemistry samples were only taken at 1-2 sites per reach, only summary statistics were provided. Box plots were created using the ggplot2 package (Wickham 2016) in R Version 4.1.1 (R Development Core Team 2021).

2.3.2.1.2. Multivariate assessment of chemical/physical habitat

Multivariate analysis was used to characterize the abiotic environment of the Slave River using measured water chemistry and physical habitat parameters. Water chemistry and physical habitat parameters measured at all sites were used to assess variation and identify major gradients in the abiotic environment through Principal Component Analysis (PCA) with standardization of variable scores. The analysis included ions, nutrients, physicals, and metals (focusing on those that were above detection limit and that showed some variation among sites, i.e., were not the same across all sites), as well as water velocity and substrate composition. Only sites with water chemistry data (see Table 1) were included in the analysis. Prior to analysis, all abiotic parameters were $\log_{10}$- or logit-transformed as appropriate. Multivariate analysis was run in CANOCO (version 4.05; ter Braak and Šmilauer 2002).

2.3.2.1.3. Biotic metric variation

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 previous years, 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. Instead, the metrics abundance of *Hydra* and relative abundance of *Hydra* were added, to examine high abundances of this genus in 2020 Slave River samples. 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. Although ANOVAs were used to compare metric values among reaches in previous years, this analysis was not completed for 2020 data because of the lack of or low replication in several reaches in 2020 (Table 1).

2.3.2.1.4. Multivariate assessment of BMI assemblage 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. 2022). 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.

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. The results of this analysis for 2020 should be interpreted with caution, as the unbalanced design with unequal replication (and no replication in two reaches) can affect the results of the analysis (Anderson 2017). Variability within reaches was assessed for Reach1-Reach 4B (excluding reaches without replication) 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. Distance to centroid was plotted with a box plot to visualize within-reach variability across reaches for each river. 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). PERMANOVA and homogeneity of multivariate dispersions were run in R version 4.1.1 (R Development Core Team 2021) using the packages vegan version 2.5-7 (Oksanen et al. 2020) and pairwiseAdonis (Martinez Arbizu 2020).

2.3.2.1.5. Biotic-abiotic relationships

The relationship between the BMI assemblage data and abiotic data was tested with Redundancy Analysis (RDA), with a subset of abiotic parameters (water chemistry and physical habitat) selected for inclusion as constraining variables based on their importance in the abiotic PCA. Because there were BMI data for up to 5 sites in each reach, but water chemistry samples were not collected at all sites, this analysis used average water chemistry values for missing sites (e.g., missing data in site 2 were created by using the average of sites 1 and 3, and site 4 used the average of sites 3 and 5 for each water chemistry parameter). This approach was found to be appropriate in the analysis of data from 2018 (Lento 2020). Because there were many water chemistry and habitat variables collected at each site, correlations among all chemistry and physical habitat variables were calculated to identify those that were highly correlated ($|r| > 0.6$). When variables were found to be highly correlated, only a single variable was retained for the analysis. This approach was used in combination with the results of the water chemistry and physical habitat PCA to choose a subset of variables for the RDA. Because only 18 sites were sampled in the Slave River in 2020, the number of potential environmental variables was limited to 9 or fewer to avoid overfitting. Prior to analysis, all abiotic parameters were $\log_{10}$- or logit-transformed as appropriate, and all BMI data were $\log_{10}(x+1)$ transformed. RDA was run in CANOCO.

2.3.2.2. Temporal characterization of reaches

Analysis of temporal variation in monitoring data from 2017 to 2020 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, and 2020) for the Slave River. These plots were used for a visual assessment of major changes that occurred between sampling years. 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.

2.3.2.3. Normal range and CES

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 four years of sampling in the Slave River.

2.3.2.3.1. Within-year variability

The normal range and CES was developed using 2020 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 2020 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 2020 were also compared with CES limits calculated from the combined 2017-2020 data, to look at variation in the current year relative to all years of sampling (multi-year CES).

2.3.2.3.2. 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 (mean of means) and standard deviation, 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 (2017-2020 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, and 2020), and they were compared with the temporal CES for the reach (grand mean ± 2SD for the reach). Assessment of variability in normal range and changes across the four years of sampling was used to support conclusions and recommendations for future years of sampling.

2.3.3. Multivariate Normal Range and CES

Multivariate temporal patterns were assessed for the Hay River (2017-2019) and Slave River (2017-2020) to explore possible approaches to quantify 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 (2017-2019 for Hay River and 2017-2020 for Slave River) 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. For each river, Procrustes analysis was used to compare PCA ordinations (first two axes) for each pairwise combination of sample years.

The goal in this assessment was to use Procrustes residuals as a measure of inter-annual variability in assemblage structure at the site scale. First, site-scale variability was summarized across years by using the m_{12}^2 (sum of squared residuals) from each Procrustes analysis as a dissimilarity measure for each pairwise comparison of years, and constructing a dissimilarity matrix of those values, 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 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. Although the interpretation of change trajectories is somewhat limited with few years of data, additional sampling data in future years will improve the power of this test to identify temporal multivariate patterns. Second, Procrustes

residuals were 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.

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. 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 Hay River, Procrustes analysis was completed using only the 25 sites that were sampled in all three years (i.e., excluding Reach 6). For the Slave River, two sets of Procrustes analyses were run: (1) analysis of 2017-2019, using the 30 sites sampled in all three years (i.e., excluding Reach 6), and (2) analysis of 2017-2020, using the 17 sites sampled in all four 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).

2.3.4. Ecological Response to Flow Conditions

2.3.4.1. Characterizing flow conditions

Peters et al. (2014) describe a number of hydro-ecological variables that have the potential to influence river communities in cold regions. Relevant variables include timing, duration, and magnitude of peak flow (maximum discharge in the year) and low flow (minimum discharge in the year), the rise rate (the rate of increase of flow, typically during the spring freshet as peak flows are reached), fall rate (the rate of decrease of flow, typically a decrease from peak flow during the summer/fall), the number of reversals (changes between increasing or decreasing flow), the magnitude of median monthly flow, and the 90-day minimum and maximum flow magnitude (a measure of seasonal low and high flows) (Peters et al. 2014). These variables have the potential to affect habitat quality, availability, and connectivity, and also describe exposure to potentially stressful low-flow or high-flow conditions. Assessment of the variables for a river system can help characterize the variability in flow conditions to which organisms are exposed.

To address the issue of variability in flow conditions, long-term hydrology data for the Hay River and Slave River were assessed to evaluate flow from 2017-2020 in the context of the longer record. For the Hay River, hydrology station 07OB008 (Hay River near Alberta/NWT boundary) only had discharge records from 2017, but station 07OB001 (Hay River near Hay River), located north of the NWT border, had discharge data from 1963 to present. For the Slave River, hydrology station 07NB001 (Slave River at Ft. Fitzgerald) had sporadic discharge data from 1921 to 1960, but more consistent records from 1960 to

present. All hydrology data were obtained from Water Survey of Canada national hydrometric network database (wateroffice.ec.gc.ca). Preliminary visualization of the long-term hydrograph was completed with all data (1963-2020 for Hay River and 1960-2020 for Slave River), but subsequent analyses of hydro-ecological variables was completed with data from 1990-2020 to consider the most recent 30 years of flow data.

Visualization of the hydrograph and calculation and visualization of hydro-ecological variables was completed using Indicators of Hydrologic Alteration Version 7.1 (The Nature Conservancy 2009). Environmental flow components were plotted for the full and reduced periods of record, as well as for the period of sampling. Calculation of environmental flow components followed the default suggestions (The Nature Conservancy 2009). High and low-flow pulse thresholds were calculated as the median ± 25%. High flows were classified as flows that exceeded 75% of daily flows for the period, whereas all flows below 75% were classified as low flows. Small flood events were classified as high flows with a peak greater than the two-year return period event, whereas a large flood event was classified as high flows with a peak greater than the 10-year return period event. High flows that did not meet the criteria for small floods or large floods were classified as high flow pulses. Extreme low flows were classified as low flow below 10% of daily flows.

Other variables that were calculated and visualized included monthly median flow; 7-day, 30-day, and 90-day minimum and maximum flows; frequency, duration, and peak of small floods, large floods, and high flow pulses; and the rise and fall rate. Non-parametric calculations were chosen for regressions, recognizing the potential for non-normality in flow data. A subset of plots is presented, with the focus on those that are relevant to understanding the variability in flow conditions over the period of sampling in each river.

2.3.4.2. Flow-ecology relationships

Flow-ecology relationships were examined by testing for community change points in response to flow. The Threshold Indicator Taxa ANalysis (TITAN2) method (Baker and King 2010) quantifies the relationship between BMI abundance and environmental gradients, identifying taxa that consistently respond positively or negatively to the environmental gradient. Indicator value (IV) scores are calculated for each taxon based on the percentage of samples in which the taxon is found and the percent composition of the taxon in each sample. A subset of taxa is then identified as pure (if they show a consistent directional change in abundance along the gradient for ≥ 95% of 999 bootstrapped runs), and reliable (≥ 95% of 999 bootstrapped runs differ significantly from a random distribution), and retained for analysis of community change-points (Baker and King 2010). The IV scores of retained taxa are standardized relative to the permuted mean and SD to calculate *z*-scores, and taxa are classified as either negative responders (*z*-), which increase in abundance at low ends of the environmental gradient, and positive responders (*z*+), which increase in abundance at high ends of the environmental gradient. Sums of IV *z*-scores for negative and positive responders are then used to identify peaks along the gradient, where there is a large amount of change in community composition, and plateaus, where change along the gradient is consistent.

For both the Hay River and Slave River, TITAN2 analysis was completed by using the relative abundance of taxa from all samples (2017-2019 for the Hay River and 2017-2020 for the Slave River), retaining those samples with an on-site velocity measurement. Although discharge data from hydrologic gauges would have provided a more accurate picture of flow conditions from one year to the next, the single flow value was not sufficient to identify flow-ecology relationships. Integration of annual discharge

measurements with site-level velocity measurements was attempted by using a PCA to combine velocity and discharge into a single summary value (axis score for each sample). However, this approach was not successful, as annual discharge differences dominated the PCA and all site axis scores were the same for a particular year. Spot measurements of velocity were instead used in the TITAN2 analysis as a proxy for flow for each year, allowing for the detection of site-level differences. Analysis included taxa at the family/subfamily level (subfamily for Chironomidae, and family or higher for all other taxa), consistent with ordinations. Only taxa found in at least 5 samples were retained for analysis (the minimum required by the test). Analysis was run using the TITAN2 package (Baker et al. 2020) in R.

The results of both the characterization of flow variability and ecological thresholds were used to consider whether flow-specific normal range and CES might be possible for the Hay River and Slave River.

3. Results and Discussion

3.1. 2020 Hydrologic Conditions

3.1.1. Hay River

Water levels in the Hay River have been extremely variable across the four 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

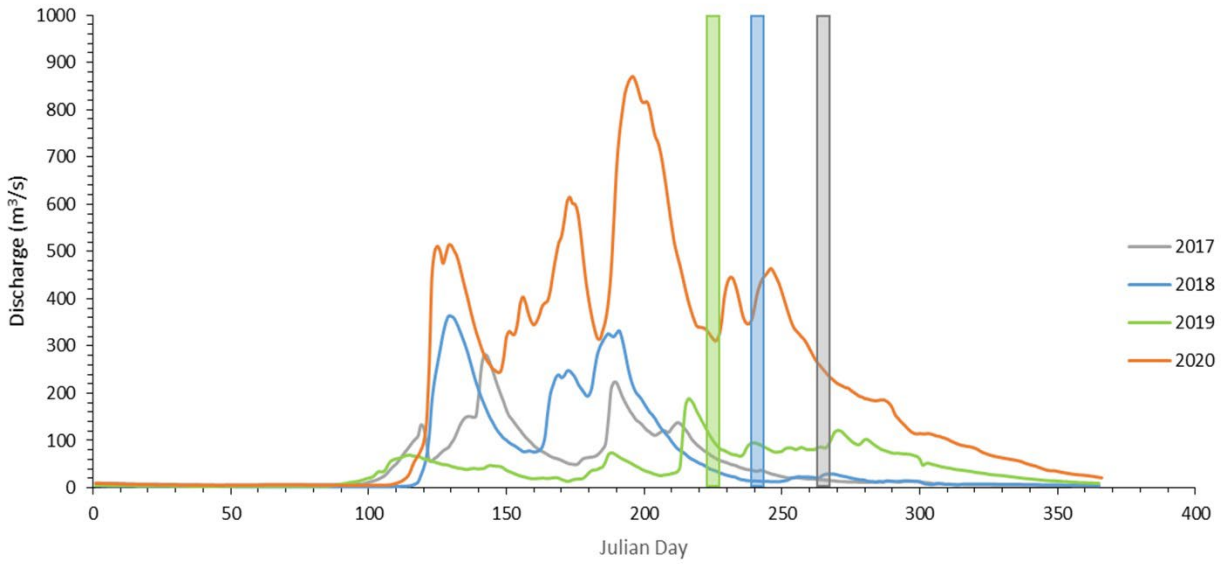

Figure 5. Hydrographs for the Hay River in 2017 (grey), 2018 (blue), 2019 (green), and 2020 (orange), with vertical shaded bars indicating the timing of sampling in each year (no vertical shaded bar for 2020 because it was not possible to sample). Data for Hay River near ALTA/NWT boundary (station 07OB008)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

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.

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 across the period of 60 days prior to sampling, median discharge in the Hay River in 2020 was 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, which reflected the consistently high water levels throughout the summer and fall of 2020. A similar pattern was observed when antecedent conditions in the 30 days prior to sampling were compared among years. Median flow in 2020 was much

Table 2. Antecedent hydrology metrics for the Hay River in 2017, 2018, 2019, and 2020, including discharge (Q (m^3/s)) at time of sampling (sample date for the Slave River used for 2020 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.

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

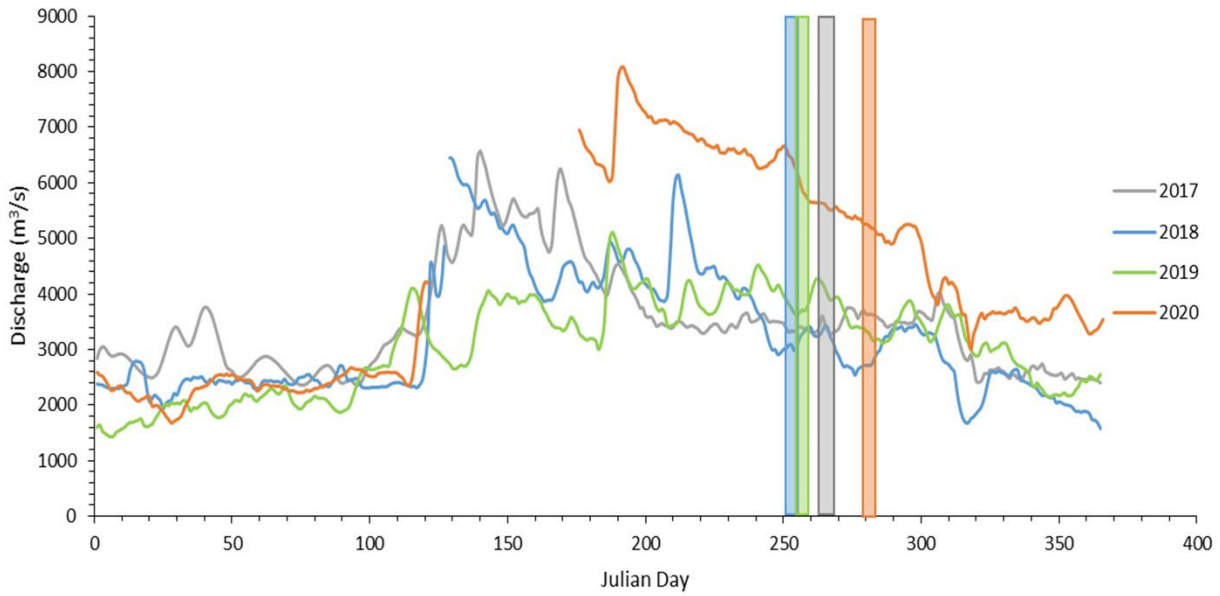

Figure 6. Hydrographs for the Slave River in 2017 (grey), 2018 (blue), 2019 (green), and 2020 (orange), with vertical shaded bars indicating the timing of sampling in each year. Data for Slave River near Fort Fitz (station 07NB001) from wateroffice.ec.gc.ca.

higher than in previous years (an order of magnitude higher than in 2017), but variability was lower (Table 2). These results highlight the dramatic change in flow conditions in the Hay River in 2020 compared to earlier years of the sampling program, and make it clear that sampling was not possible in 2020, 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 6), 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 6). 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

Table 3. Antecedent hydrology metrics for the Slave River in 2017, 2018, 2019, and 2020, 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.

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

Table 4. Summary of ion, nutrient, and physical water chemistry parameters sampled in the Slave River at six sample reaches in 2020, indicating site mean ± standard deviation for each reach where 2 or more sites were sampled. When all sites in a reach were below detection limit, the detection limit is presented. When only a subset of sites in a reach was below detection limit,

half the detection limit was used in calculations (number of sites below detection limit indicated in Parameter column). Reaches are ordered from upstream (KS1) to downstream (KS5). Three sites were sampled in Reaches 1-3, two sites in Reach 4, and one site in Reaches 5-6, with a duplicate sample collected at two sites. 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-KS4B	SR-KS6B	SR-KS5A
Alkalinity (mg/L)	77.2 ± 1.8	77.6 ± 0.6	72.7 ± 0.3	67.7 ± 0.8	70.3	66.2
Ammonia (mg/L) (3 below DL)	0.007 ± 0.001	0.006 ± 0.003	0.007 ± 0.001	0.007 ± 0.001	0.007	< 0.005
Chloride (mg/L)	12.52 ± 2.11	7.50 ± 0.00	11.23 ± 1.01	7.45 ± 0.07	7.50	6.85
Specific Conductivity (µS/cm)	234.0 ± 13.9	216.7 ± 2.1	220.3 ± 4.9	199.0 ± 0.0	202.0	198.0
Nitrate (mg/L)	0.052 ± 0.003	0.053 ± 0.006	0.050 ± 0.010	0.050 ± 0.000	0.050	0.050
Nitrite (mg/L) (14 below DL)	0.006 ± 0.001	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Dissolved N (mg/L)	0.347 ± 0.021	0.340 ± 0.010	0.323 ± 0.006	0.310 ± 0.014	0.300	0.305
Total N (mg/L)	0.418 ± 0.018	0.417 ± 0.021	0.417 ± 0.015	0.385 ± 0.007	0.380	0.390
DOC (mg/L)	9.97 ± 0.35	10.07 ± 0.23	9.60 ± 0.30	8.85 ± 0.07	9.40	8.90
TOC (mg/L)	9.53 ± 0.32	9.83 ± 0.15	9.30 ± 0.00	8.70 ± 0.14	8.90	8.60
Ortho-Phosphate (mg/L) (3 below DL)	0.0038 ± 0.0026	0.0037 ± 0.0025	0.0037 ± 0.0023	0.0030 ± 0.0000	0.0040	0.0040
pH	7.86 ± 0.01	7.88 ± 0.01	7.87 ± 0.01	7.87 ± 0.01	7.92	7.90
Dissolved P (mg/L)	0.008 ± 0.001	0.008 ± 0.000	0.008 ± 0.001	0.008 ± 0.001	0.006	0.008
Total P (mg/L)	0.065 ± 0.003	0.081 ± 0.004	0.074 ± 0.003	0.086 ± 0.008	0.076	0.084
TDS (mg/L)	153.0 ± 6.9	143.7 ± 12.9	149.7 ± 4.6	140.0 ± 11.3	148.0	138.0
TSS (mg/L)	36.0 ± 1.0	56.0 ± 4.6	55.7 ± 4.2	100.0 ± 42.4	72.0	65.0
Sulphate (mg/L)	17.7 ± 0.6	18.3 ± 0.6	17.3 ± 0.6	17.0 ± 0.0	16.0	16.0
Turbidity (NTU)	25.6 ± 0.7	34.2 ± 1.4	30.2 ± 0.1	39.6 ± 4.7	37.3	37.2

at the time of sampling in early October (Figure 6). 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.

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). 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, but relatively low variability (Table 3). Overall, flow conditions in 2020 in Slave River represented extreme high flows compared to previous years, leading to logistical complications and potential impacts on BMI assemblage structure.

3.2. Slave River Assessment

3.2.1. Spatial characterization of reaches

3.2.1.1. Chemical and physical habitat

3.2.1.1.1. 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 three or fewer sites per reach (Table 1) to account for local-scale variability in BMI assemblages in response to 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 6). In their analysis of long-term trends in water quality of the transboundary waters of the Slave River, Sanderson et al. (2012) found 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 a reflection in part 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 2020 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 analysis here is therefore primarily focused on characterizing the water chemistry conditions at the time of sampling.

The longitudinal gradient of Slave River reaches extends from Reach 1 at the south (upstream) to Reach 5 at the north (downstream; Figure 3B), with Reach 6 located upstream of Reach 5. Reach 4 had sampling sites on both banks (Reach 4A and Reach 4B), which were determined in 2017 and 2018 to have different habitat conditions (including different substrate composition). However, only Reach 4B could be sampled in 2020. Assessment of all data from the Slave River considered variation within and among reaches to account for differences due to reach location and location of sites within reaches; however, analysis was limited by the lack of replication in Reach 5 and Reach 6, as well as the reduced number of sites in Reach 2 and Reach 4B.

Water samples were collected in each river reach at one to three sites (see Table 1 for details) and analyzed for ions, nutrients, and physicals. 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 1), guidelines were available for ammonia, chloride, nitrate, nitrite, pH, TSS, and turbidity.

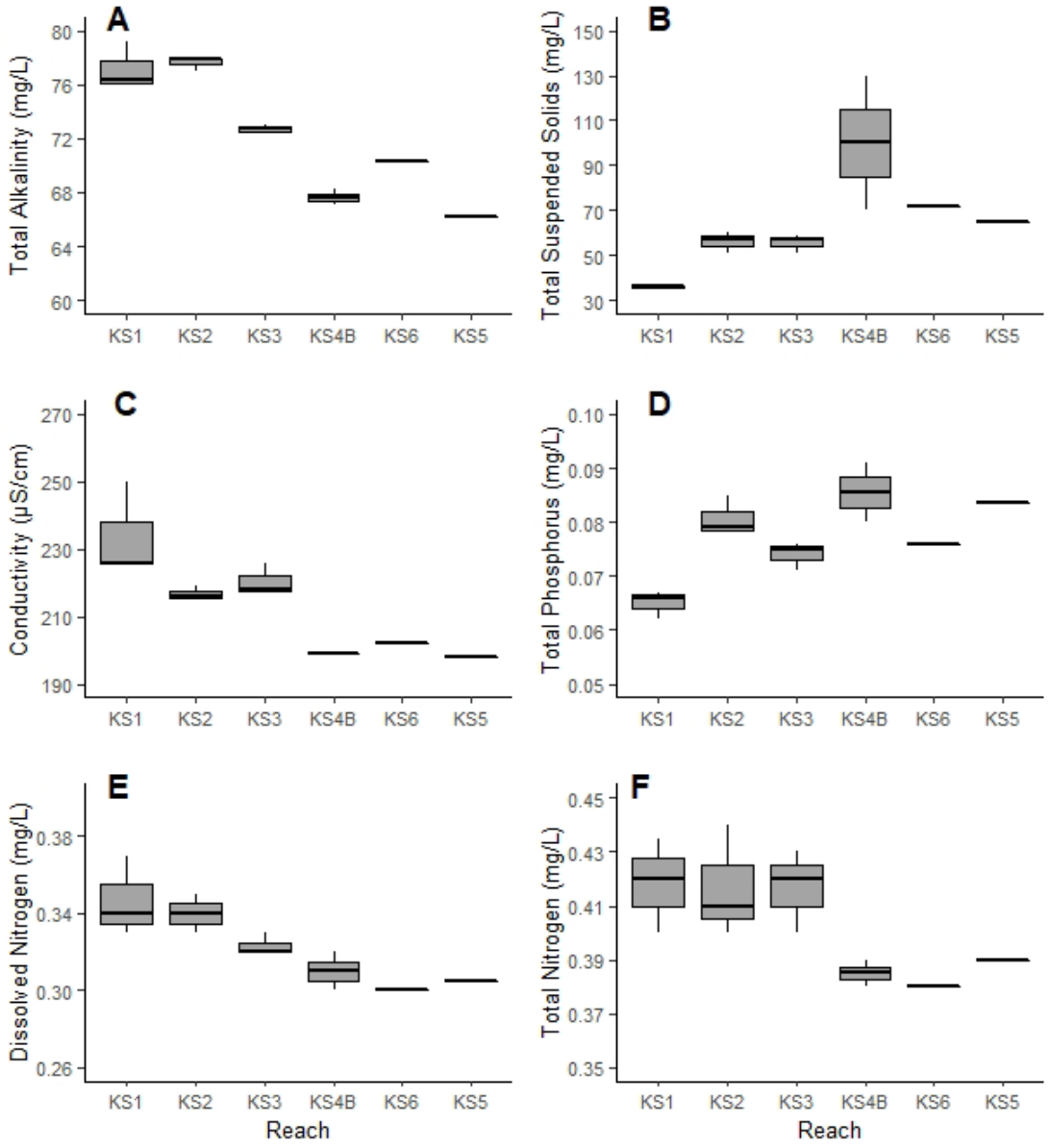

*Figure 7. Box plots of water chemistry concentrations for all reaches sampled in the Slave River in 2020, including (A) alkalinity, (B) TSS, (C) conductivity, (D) TP, (E) DN, and (F) TN. Box indicates the interquartile range, line through the box indicates the median, and whiskers indicate the range of data outside the lower and upper quartiles (1.5*interquartile range). Reaches are ordered from upstream (KS1) to downstream (KS5).*

There were no exceedances of long-term guidelines for any of the water chemistry parameters in Table 4 for which guidelines exist. Figure 7 indicated variability among reaches for several parameters, but should be interpreted with caution, as there were fewer samples collected in Reach 2 and Reach 4B, and only a single sample collected in Reach 6 and Reach 5. Most water quality parameters had similar values to those observed in 2019, including alkalinity, conductivity, and nutrients (Table 4). Some parameters like TDS and turbidity were lower in 2020, likely due to the less variable and less flashy flow conditions in 2020. TSS levels were generally similar to what was observed in 2019 and lower than long-term (1982-2010) mean of less than 100 mg/L reported for August and September at Fort Smith (Sanderson et al. 2012), though Reach 4B did have higher TSS than observed in other reaches (mean = 100 mg/L). Flow

Table 5. Summary of metal water chemistry parameters sampled in the Slave River at six sample reaches, indicating site mean ± standard deviation (for 2 or more samples) for each reach. When all sites in a reach were below detection limit, the detection

limit is indicated (parameters below DL in all reaches were omitted from the table). When only a subset of sites in a reach was below detection limit, half the detection limit was used in calculations (number of sites below detection limit indicated in Parameter column). Dissolved metal values were excluded when they exceeded total metals. 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). Three sites were sampled in Reaches 1-3, two sites in Reach 4, and one site in Reaches 5-6, with a duplicate sample collected at two sites. Reaches are ordered from upstream (KS1) to downstream (KS5).

Parameter	SR-KS1	SR-KS2	SR-KS3	SR-KS4B	SR-KS6	SR-KS5
Aluminum Diss. (µg/L)	2.95 ± 0.23	3.07 ± 0.12	2.83 ± 0.25	2.90 ± 0.28	2.80	3.00
Aluminum Total (µg/L)	**694.17 ± 44.13**	**846.33 ± 40.46**	**780.33 ± 19.55**	**967.00 ± 159.81**	**910.00**	**935.00**
Arsenic Diss. (µg/L)	0.47 ± 0.06	0.50 ± 0.00	0.43 ± 0.06	0.45 ± 0.07	0.40	0.50
Arsenic Total (µg/L)	1.08 ± 0.08	1.23 ± 0.06	1.13 ± 0.06	1.30 ± 0.14	1.20	1.30
Barium Diss. (µg/L)	40.22 ± 1.07	39.67 ± 0.85	38.40 ± 0.46	37.55 ± 0.64	38.10	38.05
Barium Total (µg/L)	60.07 ± 2.65	61.47 ± 1.80	57.57 ± 0.64	62.80 ± 7.07	57.90	61.90
Boron Diss. (µg/L)	17.82 ± 0.10	19.33 ± 0.67	18.30 ± 0.20	19.00 ± 0.14	18.80	19.10
Boron Total (µg/L)	19.55 ± 0.48	21.33 ± 0.95	19.90 ± 0.10	19.85 ± 0.92	19.60	20.55
Cesium Total (µg/L)	0.20 ± 0.00	0.20 ± 0.00	0.20 ± 0.00	0.25 ± 0.07	0.20	0.25
Chromium Diss. (µg/L) (14 below DL)	< 0.10	< 0.10	< 0.10	0.08 ± 0.04	< 0.10	< 0.10
Chromium Total (µg/L)	1.08 ± 0.08	1.33 ± 0.06	1.20 ± 0.00	1.50 ± 0.28	1.40	1.45
Cobalt Total (µg/L)	0.62 ± 0.03	0.70 ± 0.00	0.63 ± 0.06	0.90 ± 0.14	0.70	0.80
Copper Diss. (µg/L)	0.90 ± 0.00	1.00 ± 0.00	1.05 ± 0.07	1.00 ± 0.00	1.10	1.00
Copper Total (µg/L)	2.37 ± 0.38	2.47 ± 0.06	2.27 ± 0.06	**2.85 ± 0.35**	2.50	2.65
Iron Diss. (ug/L)	74.83 ± 0.29	75.67 ± 5.51	63.00 ± 0.00	54.50 ± 0.71	51.00	52.50
Iron Total (µg/L)	**1570.0 ± 72.1**	**1873.3 ± 66.6**	**1656.7 ± 28.9**	**2150.0 ± 311.1**	**1880.0**	**2035.0**
Lead Diss. (µg/L) (14 below DL)	< 0.10	< 0.10	0.23 ± 0.32	< 0.10	< 0.10	< 0.10
Lead Total (µg/L)	0.77 ± 0.06	0.90 ± 0.00	0.80 ± 0.00	1.10 ± 0.14	0.90	1.05
Lithium Diss. (µg/L)	5.38 ± 0.24	5.57 ± 0.12	5.27 ± 0.15	5.35 ± 0.07	5.30	5.35
Lithium Total (µg/L)	6.28 ± 0.25	6.70 ± 0.30	6.20 ± 0.00	6.30 ± 0.14	6.20	6.45
Manganese Diss. (µg/L)	4.68 ± 1.10	3.13 ± 0.23	2.37 ± 0.12	2.50 ± 0.57	1.70	2.20
Manganese Total (µg/L)	42.70 ± 2.57	49.67 ± 1.50	43.60 ± 1.06	55.05 ± 6.15	48.00	51.95
Mercury Diss. (UL) (ng/L)	0.55 ± 0.05	0.57 ± 0.15	0.53 ± 0.06	0.55 ± 0.07	0.50	0.50
Mercury Total (UL) (ng/L)	4.18 ± 0.97	4.10 ± 0.17	3.83 ± 0.31	6.55 ± 1.06	5.30	5.60
Mercury Total (µg/L) (7 below DL)	0.019 ± 0.001	0.008 ± 0.003	0.008 ± 0.003	0.008 ± 0.004	< 0.005	< 0.005
Molybdenum Diss. (µg/L)	0.60 ± 0.00	0.60 ± 0.00	0.60 ± 0.00	NA	0.60	0.55
Molybdenum Total (µg/L)	0.60 ± 0.00	0.60 ± 0.00	0.60 ± 0.00	0.50 ± 0.00	0.60	0.55
Nickel Diss. (µg/L)	1.10 ± 0.00	1.17 ± 0.06	1.10 ± 0.00	1.10 ± 0.00	1.10	1.10
Nickel Total (µg/L)	2.63 ± 0.12	2.93 ± 0.12	2.73 ± 0.06	3.30 ± 0.42	2.90	3.15
Rubidium Diss. (µg/L)	0.80 ± 0.00	0.77 ± 0.06	0.77 ± 0.06	0.70 ± 0.00	0.70	0.75
Rubidium Total (µg/L)	2.52 ± 0.10	2.90 ± 0.17	2.70 ± 0.00	3.15 ± 0.35	3.00	3.10
Strontium Diss. (µg/L)	135.67 ± 5.69	131.33 ± 2.52	128.33 ± 3.06	123.50 ± 0.71	NA	124.50

Parameter	SR-KS1	SR-KS2	SR-KS3	SR-KS4B	SR-KS6	SR-KS5
Strontium Total (µg/L)	140.33 ± 4.04	135.67 ± 3.51	131.00 ± 1.00	124.00 ± 2.83	121.00	126.00
Titanium Diss. (µg/L)	0.18 ± 0.03	0.23 ± 0.06	0.20 ± 0.00	0.20 ± 0.00	0.20	0.20
Titanium Total (µg/L)	13.63 ± 1.10	16.63 ± 1.17	15.57 ± 0.06	19.65 ± 1.34	20.60	17.55
Uranium Diss. (µg/L)	0.30 ± 0.00	0.30 ± 0.00	0.30 ± 0.00	0.30 ± 0.00	0.30	0.30
Uranium Total (µg/L)	0.37 ± 0.06	0.37 ± 0.06	0.30 ± 0.00	0.35 ± 0.07	0.30	0.35
Vanadium Diss. (µg/L)	0.20 ± 0.00	0.20 ± 0.00	0.20 ± 0.00	0.20 ± 0.00	0.20	0.20
Vanadium Total (µg/L)	2.45 ± 0.13	2.97 ± 0.21	2.73 ± 0.06	3.35 ± 0.49	3.20	3.25
Zinc Total (µg/L)	4.73 ± 1.93	6.43 ± 0.68	5.63 ± 0.15	7.70 ± 1.13	6.30	7.00

remained high throughout the spring and summer of 2020, and sediment transport was likely steady throughout this period.

Estimates of mean TP in the Slave River were all lower than 0.100 mg/L, and reaches were classified as eutrophic based on the Canadian Guidance Framework (Canadian Council of Ministers of the Environment 2001b). In their analysis of long-term trends in phosphorus (total and dissolved) in the Slave River, Glozier et al. (2009) found elevated levels in the Slave River relative to the Athabasca and Peace Rivers that flow into the Slave. The spot measurements of TP collected in the Slave River during the 2020 sampling event were similar to the median of 0.078 mg/L from long-term routine monitoring data (Glozier et al. 2009). There was somewhat higher TP and lower DN in downstream reaches relative to upstream reaches (Figure 7), consistent with patterns observed in 2019.

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 (Table 5). 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). Four dissolved metal concentrations were omitted because they were higher than values for the total metal (dissolved copper in Reach 5 duplicate sample, dissolved molybdenum in Reach 4B (both sites), and dissolved strontium in Reach 6 (Table 5)).

Total metal concentrations were generally low for most metals, and there were few exceedances of long-term water quality guidelines (Table 5). Total aluminum concentrations exceeded long-term

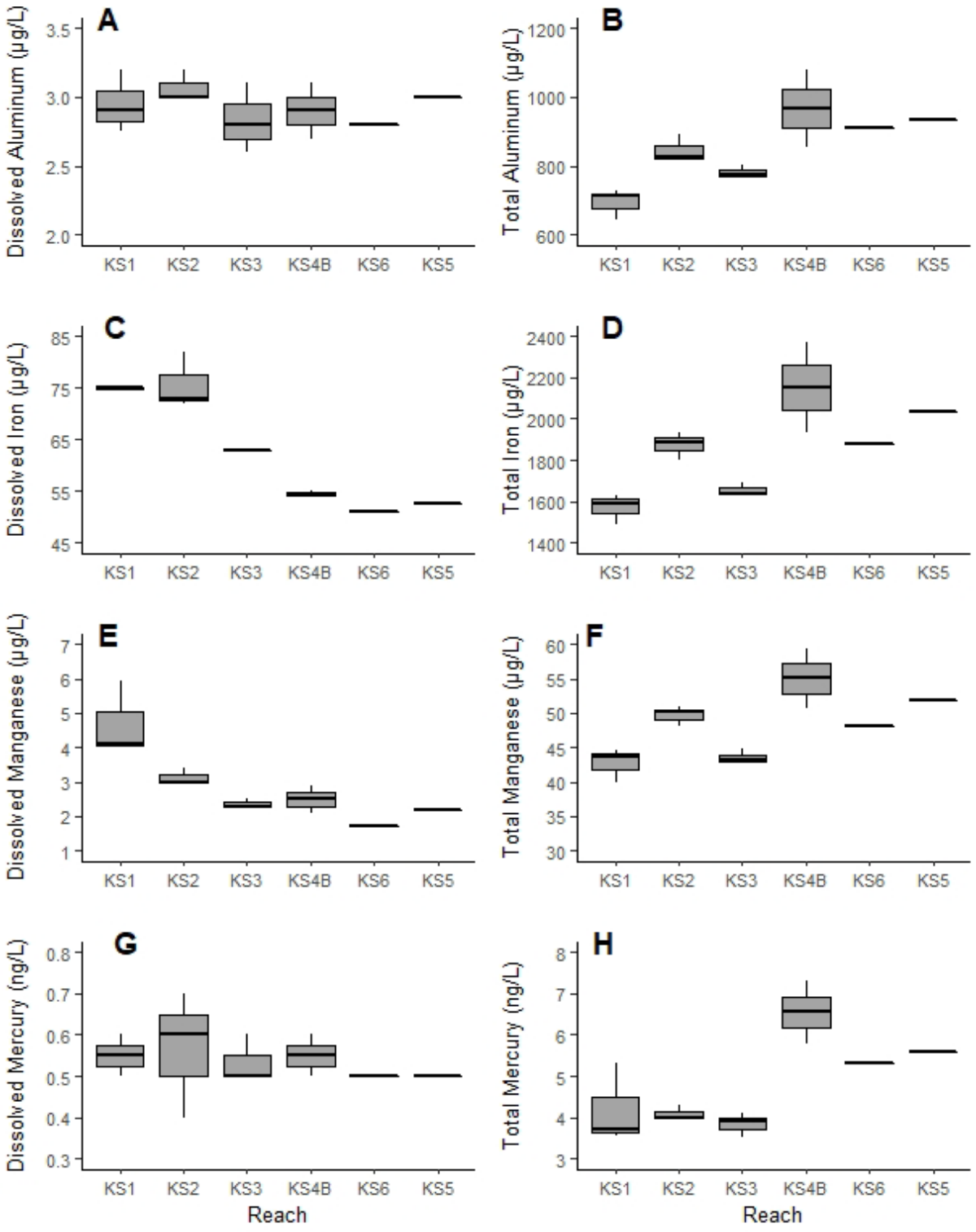

*Figure 8. Box plots of dissolved and total metal concentrations for all reaches sampled in the Slave River in 2020, including (A) dissolved aluminum, (B) total aluminum, (C) dissolved iron, (D) total iron, (E) dissolved manganese, and (F) total manganese, all measured in µg/L, and (G) dissolved mercury and (H) total mercury, both measured in ng/L. 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). Reaches are ordered from upstream (KS1) to downstream (KS5).*

Table 6. Physical habitat variables measured in the Slave River in 2020, summarized by reach. Velocity (spot measurement) is presented as mean ± standard deviation for reaches where > 1 site was sampled; 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-KS4B	SR-KS6	SR-KS5
Velocity (m/s)	0.40 ± 0.20	0.28 ± 0.03	0.25 ± 0.20	0.33 ± 0.13	0.34	0.61
Dominant streamside vegetation	deciduous trees	shrubs	shrubs	deciduous trees	NA	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
Substrate - sand (%)	0	0	0	0	0	0
Substrate - gravel (%)	2	2	0	0	0	5
Substrate - pebble (%)	58	20	49	19	20	70
Substrate - cobble (%)	40	50	50	70	80	15
Substrate - boulder (%)	0	3	0	8	0	10
Substrate - bedrock (%)	0	25	1	3	0	0

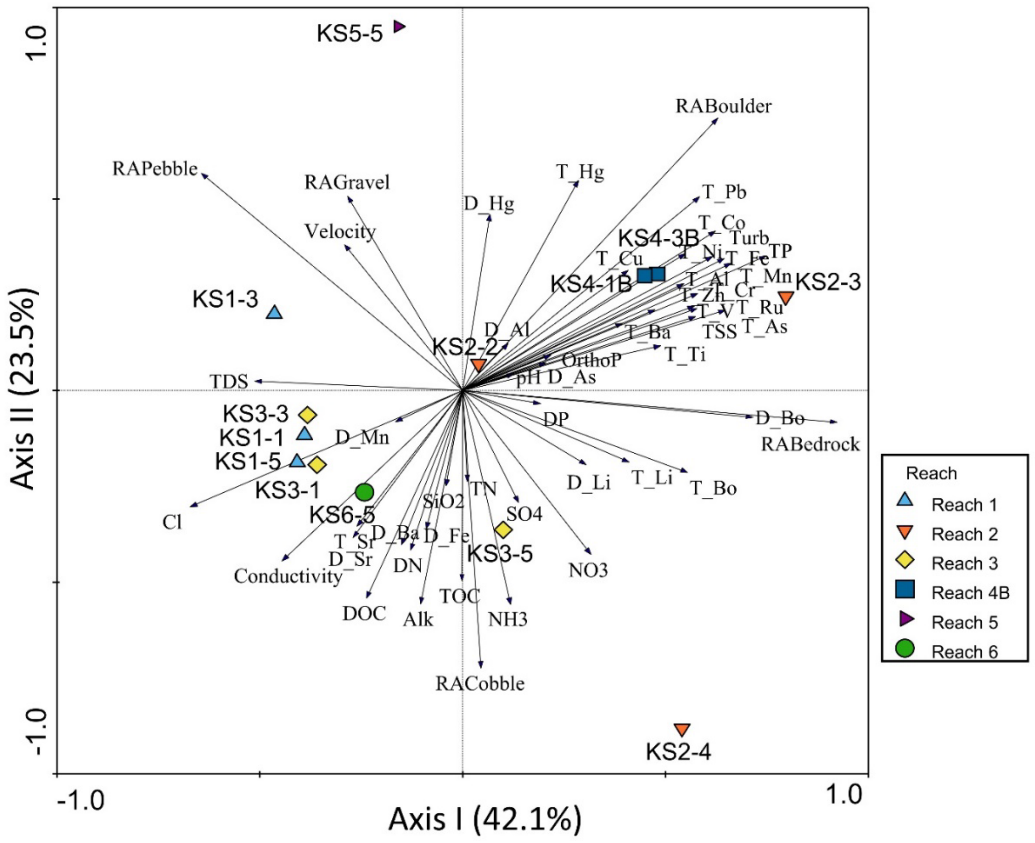

Figure 9. PCA ordination of water chemistry and habitat variables at Slave River kick-sites in 2020, with sites colour-coded based on reach number. Arrows point in the direction of increasing values of parameters, and correlations of sites with parameters are indicated by the location of kick-site points in proximity to arrows. Kick-sites on the same end of a gradient are positively correlated, whereas samples on opposite ends of gradients are negatively correlated. Kick-sites and parameters at right angles through the origin are uncorrelated. Kick-site points located near the origin have similar correlations with all measured parameters. "D-" in front of metals indicates dissolved form, and "T-" indicates total metals.

exposure water quality guidelines (100 µg/L), but reach averages, which ranged from 694 to 967 µg/L (Table 5), were much lower than the long-term median value of 4360 µg/L reported for the Slave River

at Fort Smith (Sanderson et al. 2012), and were lower than observed in 2019. Concentrations of total aluminum were generally similar among all reaches (Figure 8). The average concentration of total copper was slightly higher than the long-term guideline level of 3.66 μg/L, but this guideline level was calculated based on the hardness values measured in 2019, as hardness was not included in the suite of parameters measured by the lab in 2020. Moreover, total copper concentrations were lower in all reaches than was observed in 2019. The remaining exceedances of long-term water quality guidelines for were for total iron, which was above the CCME long-term exposure guideline of 300 μg/L in all reaches (Table 5). Total iron, which ranged from 1570 to 2150 μg/L on average across reaches, was 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 4653 μg/L based on a DOC concentration of 9.5 mg/L and a pH of 7.9 (as observed in the Slave River samples in 2020). Following the federal guideline from ECCC, all concentrations of total iron were well within acceptable limits.

3.2.1.2. Physical Habitat

Measurements were taken at each site to characterize they physical habitat in BMI sampling locations, including variables such as velocity, streamside vegetation, in-stream periphyton cover, and substrate composition (Table 6). Because water levels were high in 2020, bankfull width and wetted width were too large to be measured, and these values were omitted from the summary table. Velocity ranged from 0.25 to 0.61 m/s on average across reaches (Table 6), and was similar to averages observed in 2019 (range: 0.21 to 0.54 m/s on average). Substrate composition in reaches was predominantly a combination of pebble and cobble size classes (Table 6). 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 2019.

3.2.1.2.1. Characterizing the chemical and physical habitat

Multivariate analysis of abiotic data was used to characterize patterns in water chemistry and physical habitat within and among reaches in the Slave River, and to identify parameters that might be used for analysis of biotic-abiotic relationships. Combining water chemistry and physical habitat data in the ordination allowed for comparison of their relative importance in characterizing gradients in the abiotic habitat.

The PCA ordination of water chemistry and physical habitat variables indicated a similar contribution of both types of variables to the environmental gradients along which reaches varied (Figure 9). Whereas low variability in water chemistry in 2019 resulted in stronger loadings of physical habitat variables on the first and second axes, axis loadings were more similar for chemical and physical habitat variables in 2020. The majority of sites were spread along the first axis gradient, which explained 42.1% of the variance among sites. On the positive end of the first axis, sites in Reach 4B and site KS2-3 were positively correlated with metal concentrations (dissolved and total), and with the relative abundance of boulder and bedrock (Figure 9). On the negative end of the gradient, sites in Reach 1, Reach 3, and Reach 6 were positively correlated with conductivity, chloride, DOC, alkalinity, and TDS, and with the exception of Reach 6, they were also positively correlated with velocity and smaller substrate size (pebble, gravel; Figure 9).

The second axis gradient, which explained an additional 23.5% of variation among sites, represented a strong separation of site KS5-5 from sites KS3-5 and KS2-4 (Figure 9). Reach 5 was positively correlated with velocity (it had the highest velocity among all reaches), and with the relative abundance of gravel,

pebble, and boulder (it had the highest relative abundance of each of these substrate sizes, and the lowest relative abundance of cobble; Table 6). On the other end of the second axis gradient, sites KS3-5 and KS2-4 were positively correlated with nutrients, alkalinity, and the relative abundance of cobble.

3.2.1.2.2. Sediment chemistry

Sediment chemistry samples were collected from one to two sites in each Slave River reach 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

Table 7. Summary of sediment chemistry parameters sampled in the Slave River in 2020, indicating site mean ± standard deviation for each reach (2 samples collected in Reaches 1-3, 1 in remaining reaches, and 3 duplicates). When all samples in a reach were below detection limit, the detection limit is indicated. When only one sample in a reach was below detection limit, half the detection limit was used in calculations (number of samples below DL indicated in Parameter column). 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-KS4B	SR-KS6	SR-KS5
Physicals						
Moisture %	27.8 ± 2.3	31.9 ± 0.9	22.2 ± 2.1	27.4	39.8	24.1
Metals (mg/kg)						
Aluminum (Al)	5680.0 ± 664.7	6250.0 ± 70.7	7435.0 ± 2835.5	5470.0	10100.0	9640.0
Antimony (Sb)	0.318 ± 0.025	0.378 ± 0.032	0.270 ± 0.042	0.350	0.450	0.435
Arsenic (As)	5.28 ± 0.56	**6.04 ± 0.40**	5.04 ± 0.13	5.50	**8.00**	**7.01**
Barium (Ba)	244.5 ± 34.6	293.3 ± 25.1	203.5 ± 135.1	280.0	308.0	261.5
Beryllium (Be)	0.315 ± 0.021	0.358 ± 0.004	0.430 ± 0.198	0.300	0.530	0.565
Bismuth (Bi)	< 0.2	< 0.2	< 0.2	< 0.2	< 0.2	< 0.2
Boron (B) (5 below DL)	5.4 ± 2.3	3.2 ± 0.9	6.2 ± 5.2	< 5	6.9	7.4
Cadmium (Cd)	0.303 ± 0.045	0.415 ± 0.036	0.240 ± 0.083	0.329	0.582	0.436
Calcium (Ca)	50150.0 ± 43204.2	19175.0 ± 954.6	67300.0 ± 67458.0	18800.0	17200.0	12300.0
Chromium (Cr)	12.85 ± 1.20	13.98 ± 0.04	20.35 ± 9.55	14.00	21.50	18.85
Cobalt (Co)	5.80 ± 0.69	6.49 ± 0.08	5.93 ± 0.47	6.38	9.12	7.89
Copper (Cu)	9.30 ± 0.08	11.50 ± 1.70	7.08 ± 1.24	8.51	19.20	17.40
Iron (Fe)	14475.0 ± 1520.3	16075.0 ± 318.2	15850.0 ± 1909.2	15100.0	22200.0	19550.0
Lead (Pb)	5.42 ± 0.30	6.46 ± 0.51	5.60 ± 0.57	5.80	8.64	8.63
Lithium (Li)	7.23 ± 0.32	8.00 ± 0.28	10.05 ± 4.45	6.80	12.70	12.90
Magnesium (Mg)	7902.5 ± 1255.1	6512.5 ± 173.2	37300.0 ± 42992.1	6150.0	7330.0	6125.0
Manganese (Mn)	219.3 ± 0.4	262.8 ± 37.1	229.0 ± 15.6	214.0	307.0	263.0
Molybdenum (Mo)	0.840 ± 0.170	0.768 ± 0.018	1.180 ± 0.806	0.640	1.060	0.950
Nickel (Ni)	17.65 ± 1.06	19.00 ± 0.28	21.95 ± 8.13	17.70	26.80	23.20
Phosphorus (P)	667.3 ± 90.9	731.3 ± 30.1	538.0 ± 250.3	732.0	760.0	658.5
Potassium (K)	862.5 ± 31.8	862.5 ± 17.7	1600.0 ± 1187.9	740.0	1300.0	1485.0
Selenium (Se) (1 below DL)	0.258 ± 0.025	0.388 ± 0.074	0.180 ± 0.113	0.240	0.560	0.465
Silver (Ag) (7 below DL)	< 0.1	0.100 ± 0.028	< 0.1	< 0.1	0.160	0.145
Sodium (Na)	92.5 ± 10.6	83.0 ± 0.0	103.5 ± 34.6	74.0	95.0	91.5
Strontium (Sr)	70.25 ± 27.37	51.90 ± 0.28	51.90	51.60	56.80	46.70
Sulfur (S) (11 below DL)	< 1000	< 1000	3600.0 ± 4384.1	< 1000	< 1000	< 1000
Thallium (Tl)	0.104 ± 0.008	0.109 ± 0.008	0.076 ± 0.017	0.085	0.145	0.165
Tin (Sn)	< 2	< 2	< 2	< 2	< 2	< 2
Titanium (Ti)	77.33 ± 24.36	63.45 ± 18.46	58.85 ± 65.27	135.00	73.60	90.00

Parameter	SR-KS1	SR-KS2	SR-KS3	SR-KS4B	SR-KS6	SR-KS5
Tungsten (W)	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5	< 0.5
Uranium (U)	0.848 ± 0.129	0.811 ± 0.007	0.959 ± 0.284	0.806	0.972	0.982
Vanadium (V)	22.45 ± 2.62	23.88 ± 0.39	22.30 ± 1.41	25.10	35.10	32.80
Zinc (Zn)	52.28 ± 6.33	62.18 ± 2.44	44.60 ± 8.49	52.10	78.90	71.15
Zirconium (Zr)	3.33 ± 0.18	3.73 ± 0.25	3.25 ± 0.07	3.60	3.70	4.75
Polycyclic Aromatic Hydrocarbons (PAHs) (mg/kg)						
1-Methylnaphthalene (1 below DL)	0.017 ± 0.002	0.025 ± 0.007	0.011 ± 0.008	0.018	0.040	0.026
2-Methylnaphthalene (1 below DL)	0.020 ± 0.004	**0.029 ± 0.007**	0.013 ± 0.011	**0.022**	**0.044**	**0.030**
Acenaphthene	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005
Acenaphthylene	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005	< 0.005
Anthracene	< 0.004	< 0.004	< 0.004	< 0.004	< 0.004	< 0.004
BaP Total Potency Equivalent	< 0.01	< 0.01	< 0.01	< 0.01	0.016	< 0.01
Benz[a]anthracene	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Benzo[a]pyrene	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Benzo[b&j]fluoranthene (1 below DL)	0.017 ± 0.002	0.020 ± 0.006	0.009 ± 0.006	0.014	0.029	0.022
Benzo[b+j+k]fluoranthene (4 below DL)	0.013 ± 0.008	0.020 ± 0.006	< 0.015	< 0.015	0.029	0.022
Benzo[g,h,i]perylene (1 below DL)	0.025 ± 0.010	0.024 ± 0.006	0.010 ± 0.006	0.014	0.032	0.026
Benzo[k]fluoranthene	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Chrysene (variable DL)	< 0.01	< 0.013	< 0.01	< 0.01	< 0.017	< 0.014
Dibenz[a,h]anthracene (variable DL)	< 0.0055	< 0.005	< 0.005	< 0.005	< 0.006	< 0.005
Fluoranthene (10 below DL)	< 0.01	0.008 ± 0.004	< 0.01	< 0.01	0.012	< 0.01
Fluorene	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
IACR (CCME) (6 below DL)	0.093 ± 0.053	0.108 ± 0.074	< 0.11	< 0.11	0.220	0.140
Indeno[1,2,3-c,d]pyrene (10 below DL)	0.009 ± 0.006	< 0.01	< 0.01	< 0.01	0.011	< 0.01
Naphthalene (1 below DL)	0.012 ± 0.001	0.018 ± 0.004	0.009 ± 0.005	0.015	0.027	0.017
Phenanthrene (1 below DL)	0.022 ± 0.002	0.033 ± 0.008	0.013 ± 0.011	0.023	**0.048**	0.033
Pyrene (1 below DL)	0.011 ± 0.002	0.017 ± 0.004	0.009 ± 0.005	0.012	0.022	0.016
Quinoline (1 below DL)	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Total PAHs BC Sched 3.4 (1 below DL)	0.063 ± 0.005	0.102 ± 0.031	0.042 ± 0.031	0.072	0.095	0.164
Total PAHs (EPA 16) (1 below DL)	0.093 ± 0.019	0.118 ± 0.035	0.046 ± 0.037	0.078	0.113	0.192

incorporated into sediments in the benthic habitat due to sorption to particulate matter and subsequent settling in the sediment (Canadian Council of Ministers of the Environment 2001a, 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 2001a). 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).

To determine whether levels of metals or PAHs were elevated beyond recommended levels in Slave River samples, mean values for each site 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, and zinc, and the PAHs 2-methylnaphthalene, acenaphthene, acenaphthylene, anthracene, benz[a]anthracene, benzo[a]pyrene, chrysene, dibenz[a,h]anthracene, fluoranthene, fluorine, 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 contaminants.

Concentrations of most metals in sediments were below the guidelines for the protection of aquatic life (Table 7). Arsenic was the only metal to exceed the ISQG (three reach averages above the ISQG of 5.9 mg/kg; Table 7), 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 2019, indicating that this did not reflect an increase from the previous year. Other metals 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 7). However, concentrations of 2-methylnaphthalene were elevated above the ISQG in four reaches, including the three downstream reaches. 2-methylnaphthalene is an LMW-PAH, thus representing an acutely toxic species for benthic organisms (Canadian Council of Ministers of the Environment 1999). In Reach 6, phenanthrene (another LMW-PAH with acute toxicity) was also found to exceed the ISQG. 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, some HMW-PAHs (carcinogenic compounds) that were found to exceed ISQGs in 2019 (e.g., benzo[b&j] fluoranthene and chrysene) were below guidelines in 2020. There is not a CCME guideline for total PAHs; however, McGrath et al. (2019) presented guideline levels from a number of different sources, and the total PAHs measured in the Slave River samples were below the recommended guidelines for each source. The BaP Total Potency Equivalent, which is a measure of cancer risk to humans, and the IACR, which measures threats to drinking water, were below guideline levels in all reaches (Canadian Council of Ministers of the Environment 2010).

3.2.1.3. Benthic macroinvertebrates

3.2.1.3.1. Biotic metric variation

Biotic metrics were used to compare abundance, relative abundance, and taxonomic richness of key organism groups among sites and reaches in the Slave River. Total abundance was high in Slave River samples in 2020, particularly compared to the low abundances observed in 2019. Total abundance ranged from 2295 to 8466 individuals on average per reach, though there was moderate to high variability within reaches (Table 8). EPT abundance was fairly high across reaches, and Chironomidae abundance was nearly an order of magnitude higher than observed in 2019. The highest total abundances were observed in Reach 2 and 3, and this was driven by extremely high abundance of the genus *Hydra* (Table 8, Figure 10), a freshwater cnidarian that is related to sea anemone, jellyfish, and corals. *Hydra* were present at unusually high abundances across most reaches, and the genus accounted

for 51.2 to 91.5% of individuals in samples on average in Reaches 2, 3, 4B, 6, and 5 (it only made up 19.3% of the sample on average in Reach 1; Table 8). Thus, although abundances of EPT and Chironomidae were high, they made up a very small proportion of the samples in these reaches in 2020 (Table 8, Figure 11). Despite the greater numbers of Chironomidae than in the last year of sampling, they accounted for only 2.1 to 7.1% of individuals collected, on average.

The genus *Hydra* is included as part of CABIN sample enumeration, but is not often the focus of research on BMI assemblages, because it is rarely the dominant taxon in kick samples. In international sampling and sorting protocols for BMI, *Hydra* are often excluded, as their numbers are considered to be

Table 8. Summary of biotic metrics for kick-site reaches sampled in the Slave River in 2020, including the mean ± standard deviation (when > 1 site was sampled in a reach) 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-KS4B	SR-KS6	SR-KS5
Total Abundance	3443 ± 1819	8466 ± 3476	8007 ± 2676	3882 ± 1785	3733	2295
EPT abundance	2221 ± 804	491 ± 97	1181 ± 565	927 ± 1045	433	926
Chironomidae abundance	207 ± 220	121 ± 73	130 ± 55	130 ± 124	100	63
Diptera + Oligochaeta abundance	265 ± 249	167 ± 76	169 ± 63	183 ± 189	144	163
Hydra abundance	847 ± 1232	7784 ± 3320	6591 ± 2741	2765 ± 1024	3144	1175
Percent EPT	70.7 ± 21.0	6.1 ± 1.3	16.1 ± 10.0	20.9 ± 15.6	11.6	40.4
Percent Chironomidae	4.9 ± 3.5	1.5 ± 1.0	1.8 ± 1.1	2.9 ± 1.7	2.7	2.7
Percent Diptera + Oligochaeta	6.5 ± 3.6	2.1 ± 1.2	2.3 ± 1.1	4.1 ± 2.7	3.9	7.1
Percent *Hydra*	19.3 ± 17.8	91.5 ± 2.3	80.8 ± 11.4	74.9 ± 18.4	84.2	51.2
Taxonomic Richness	20.4 ± 4.1	12.3 ± 2.9	13.8 ± 2.6	15.0 ± 6.1	15.0	21.0
Richness of EPT	6.8 ± 1.1	4.7 ± 0.6	6.0 ± 1.0	5.3 ± 1.5	6.0	7.0
Richness of Chironomidae	6.0 ± 2.3	3.3 ± 2.1	3.4 ± 1.1	6.0 ± 3.6	4.0	7.0
Richness of Diptera + Oligochaeta	9.2 ± 3.1	5.3 ± 2.1	5.0 ± 1.9	8.0 ± 4.4	7.0	11.0

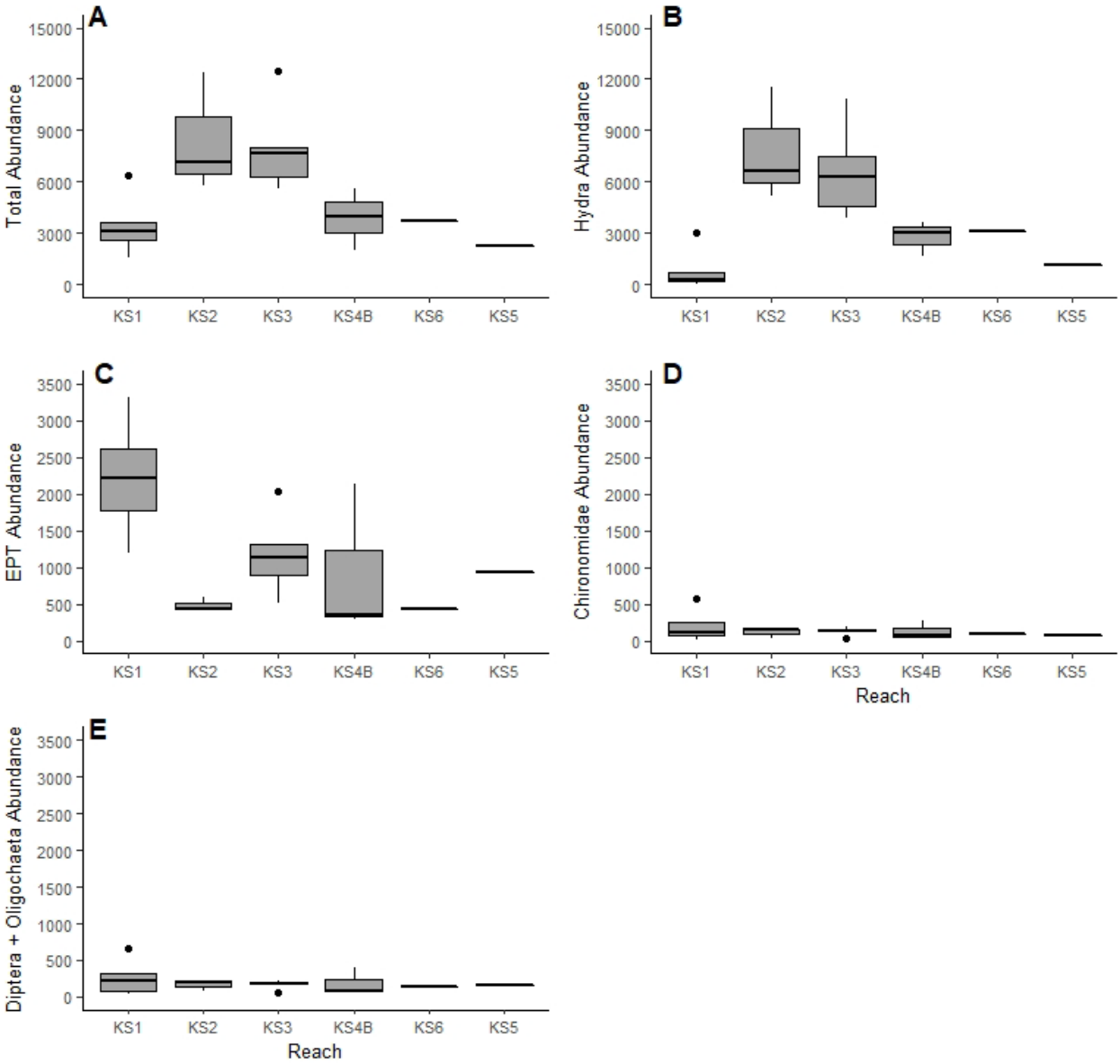

Figure 10. Box plots of abundance BMI metrics for the Slave River reaches sampled in 2020, including (A) total abundance, (B) abundance of Hydra*, (C) abundance of Ephemeroptera, Plecoptera, and Trichoptera (EPT), (D) abundance of Chironomidae (midges), (E) abundance of Diptera (true flies) + Oligochaeta (segmented worms). Panels A and B have the same scale on the y-axis, whereas panels C, D, and E have a different scale for ease of interpretation. 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).*

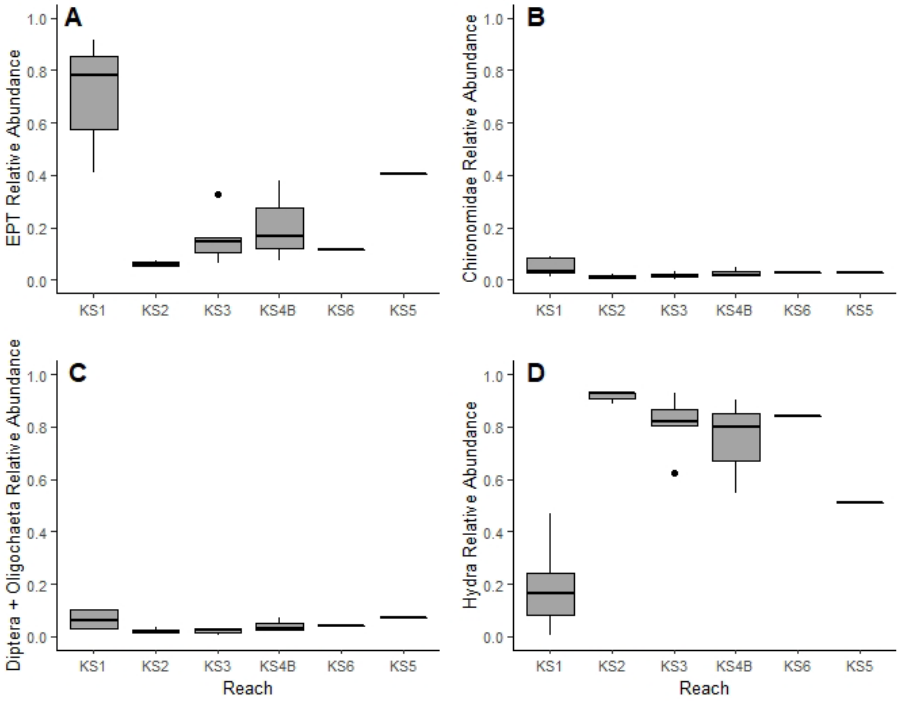

Figure 11. Box plots of relative abundance metrics for the Slave River reaches sampled in 2020, including (A) EPT, (B) Chironomidae (midges), (C) Diptera (true flies) + Oligochaeta (segmented worms), and (D) Hydra*. Box indicates the interquartile range, line through the box indicates the median, and 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).*

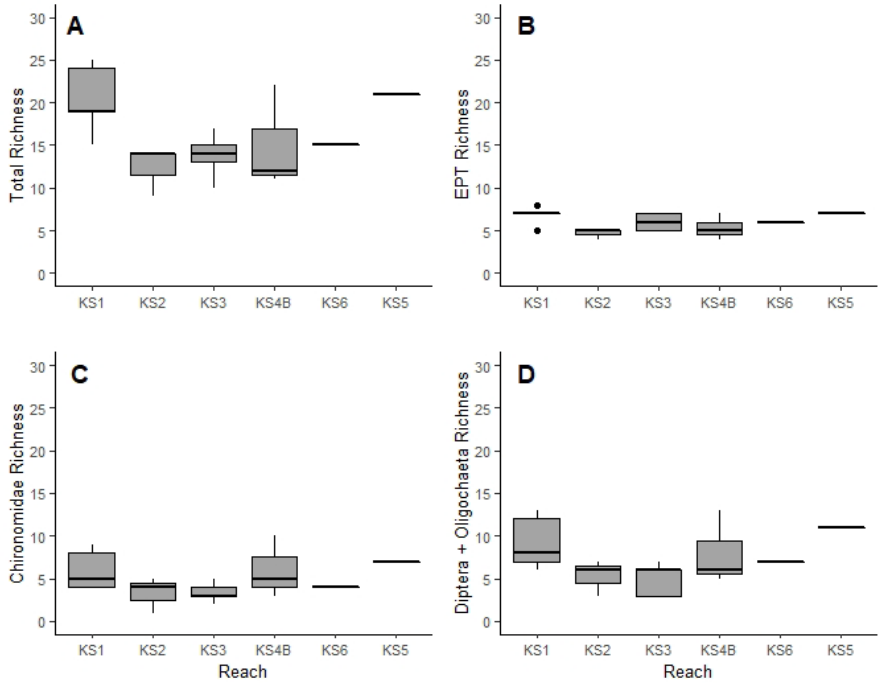

*Figure 12. Box plots of richness BMI metrics for the Slave River reaches sampled in 2020, including (A) total richness, (B), richness of EPT, (C) richness of Chironomidae (midges), and (D) richness of Diptera (true flies) + Oligochaeta (segmented worms). Box indicates interquartile range, line through 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).*

underestimated when sampled with a kick net. As a result, little research exists on freshwater *Hydra* in the context of BMI community structure, and there is much about the importance of this group to benthic ecology that is unknown (Di Camillo et al. 2017). Di Camillo et al. (2017) note that hydroids may become quite dominant in benthic habitats, and their abundance and importance to habitat structure is often underestimated. Hydroids are clonal organisms that can form multiserial colonies that spread laterally and form benthic "animal forests", becoming part of the benthic habitat and adding complexity to that habitat, altering flows and light penetration across the habitat, and providing food and shelter to other benthic organisms (Di Camillo et al. 2017). They are classified as predators, feeding on zooplankton in the water column. They can easily adapt to different environmental conditions, and respond to a variety of stressors through adaptation of their growth, reproduction, and behaviour (see review in Di Camillo et al. 2017 and references therein). In their analysis of BMI data from a large river,

Angradi et al. (2006) found that *Hydra* was an indicator taxon for distinguishing between shoreline samples (where they were abundant) and channel samples (where they were less abundant or absent). Another study on the effects of water withdrawals in a regulated river found that greater export of plankton downriver of the dam with water release favoured non-insect taxa, including planktivores, and identified *Hydra* as one of the common taxa found at downriver sites (Murphy et al. 2021). *Hydra* was one of three taxon groups that contributed to the greatest distinction between upriver and downriver sites in that study, and Murphy et al. (2021) indicated that all three taxon groups were small, multivoltine sessile predators with development that was not seasonally-timed.

The extremely high abundance *Hydra* in Slave River samples in 2020 may have reflected their ability to adapt to the changing hydrologic conditions and increased flows in both 2019 and 2020. Moreover, a review of the data collected in the Slave River in previous years indicates that *Hydra* have consistently been common in the benthic samples, and that their abundance has changed with variation in flows. *Hydra* accounted for 13-16% of individuals in the Slave River on average across all samples in 2017 and 2019, and increased to 30% on average in 2018, when antecedent flow conditions indicated elevated discharge in the river prior to sampling. Average relative abundance of *Hydra* was strongly positively correlated with antecedent discharge in the four years of sampling, with a correlation of 0.95 (though it should be noted that this correlation is likely inflated due to only four years of data). Given the changes in abundance of *Hydra* over time and its apparent adaptability, this genus may represent a strong indicator of changes to habitat conditions in the river, with high abundances signalling high flows or altered or adverse conditions.

Taxonomic richness ranged from 12 to 21 taxa on average per reach (Table 8), with higher richness in Reach 1 and Reach 5, both of which had lower abundance of *Hydra* than other reaches (Figure 12). Reach 1 and Reach 5 had higher taxonomic richness of Diptera + Oligochaeta than other reaches. In contrast, richness of EPT and Chironomidae were similar across all reaches on average.

3.2.1.3.2. 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.

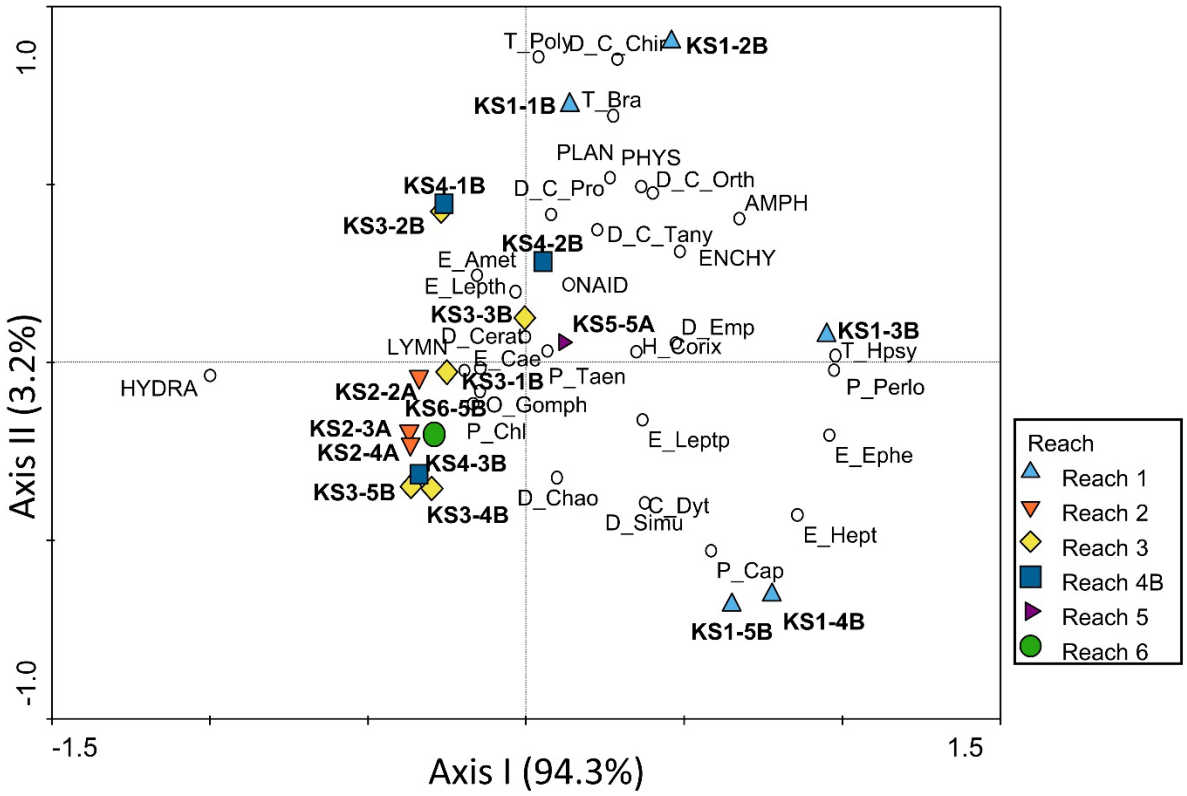

Figure 13. PCA ordination of BMI from kick samples in the Slave River in 2020, with sample points coloured by reach and both sample and taxa points labelled. Kick-sites in close proximity have similar assemblages, whereas samples on opposite ends of gradients have differences in their assemblages. Samples 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.

The PCA identified an extremely strong gradient along the first axis that explained 94.3% of the variability in assemblage structure among sites. The first axis gradient separated sites in Reach 1 from all other sites due to a positive correlation of all other reaches with *Hydra* (Figure 13). Given the low-moderate abundance of *Hydra* in Reach 1 and the high abundance of this taxon across the other reaches, it is not surprising that *Hydra* would play an important role in distinguishing among sites, though the % variance explained by the first axis was still much higher than is generally observed in such assessments. Along the second PCA axis, which explained an additional 3.2% of the variation among sites, there was separation among the sites in Reach 1 that was primarily due to differences in the composition of EPT taxa (Figure 13). Sites in Reach 1 were associated with taxa that have adaptations for fast flows, including the caddisfly Hydropsychidae, the mayflies Ephemerellidae and Heptageniidae, and the stonefly Perlodidae. Because most other sites were dominated by *Hydra*, they were largely clustered together near the origin, indicating little variability in assemblage structure.

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 ($F = 7.27$, $p = 0.004$). Pairwise PERMANOVA was used to identify which reaches had statistically significant differences in assemblages

Table 9. 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, but no p-values were significant at the FDR-corrected α.

	KS1	KS2	KS3	KS4B	KS5	KS6
KS1						
KS2	0.018					
KS3	0.008	0.123				
KS4B	0.041	0.2	0.702			
KS5	0.5	0.25	0.1667	0.25		
KS6	0.1667	0.25	0.6667	0.75	1	

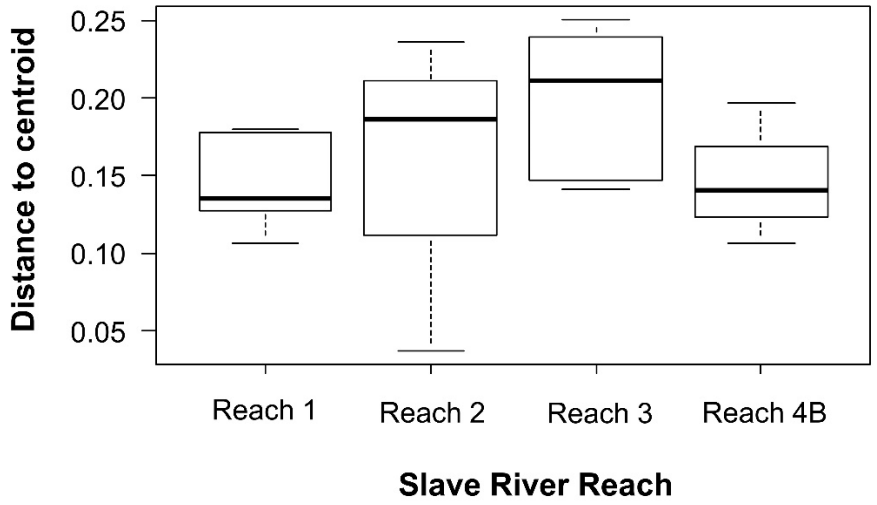

Figure 14. Results of homogeneity of multivariate dispersions analysis of Slave River BMI assemblages for reaches in which more than one site was sampled in 2020, showing the median distance to centroid for each reach (black bar), 25th and 75th percentiles (lower and upper bounds of box, respectively), 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.

(significant at an FDR-corrected α-level, based on the rank of each *p*-value). Although *p* was below 0.05 (a typical α-level) for comparisons of Reach 1 with Reach 2, Reach 3, and Reach 4B, no comparisons were significant at the FDR-corrected α-level (Table 9). The results may have been affected in part by unequal replication among reaches, and should be interpreted with caution. Reach 1 clearly did differ from other reaches with respect to its relatively low abundances of *Hydra* and its higher abundances and richness of other taxa.

Within-reach variability in assemblage composition was similar for each reach of the Slave River that had more than one site sampled (homogeneity of multivariate dispersions $F = 0.0.859$, $p = 0.47$). The median distance to centroid was similarly low for all reaches, though it was lowest in Reach 1 and Reach 4B, indicating greater similarity among sites (Figure 14). Reach 2 showed high variability in the distance to centroid, whereas Reach 3 had the highest median distance to centroid of all the reaches. Reach 2 and Reach 3 included the sites with the highest abundance of *Hydra*, but not all sites in these reaches had

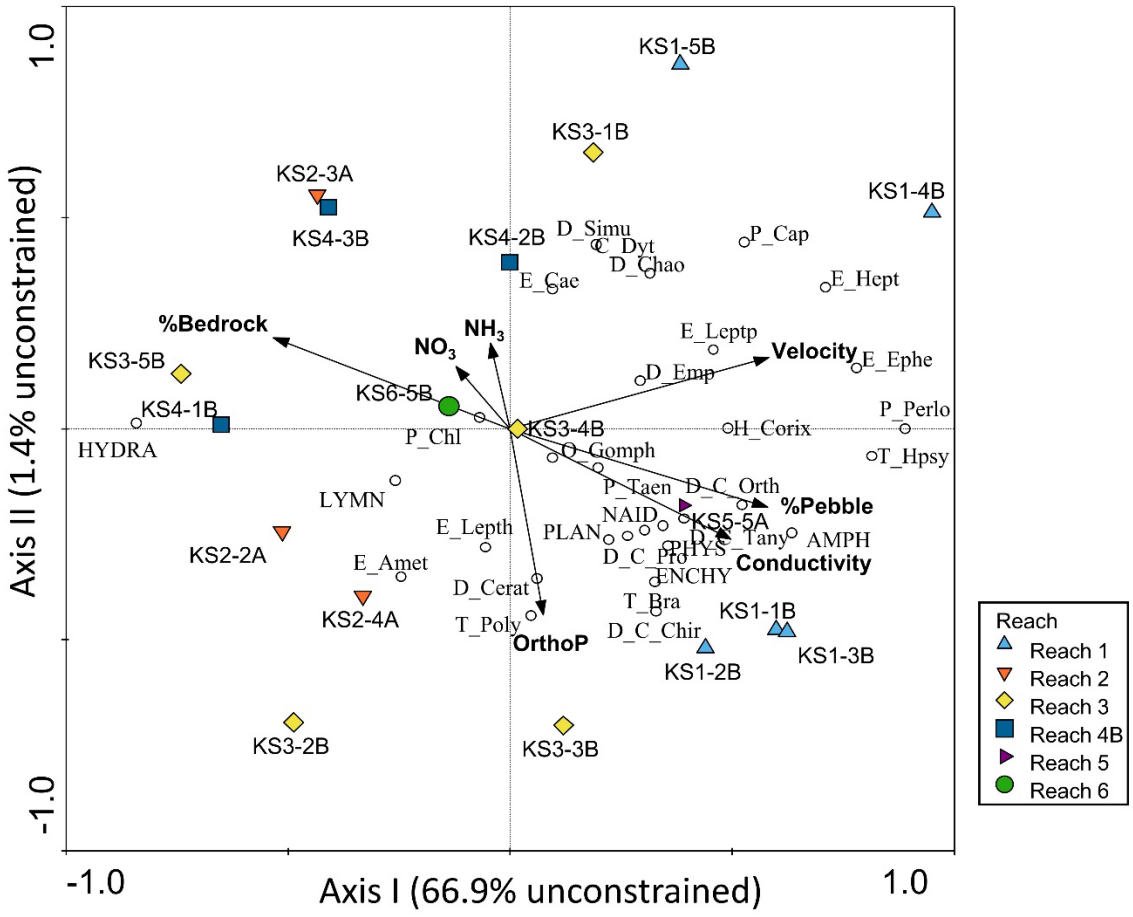

Figure 15. RDA ordination of 2020 Slave River BMI data constrained by physical and chemical habitat data at each site, with sites coloured by reach. Even-numbered sites use average values of chemical parameters from neighbouring odd-numbered sites (with the exception of Reach 2, where even-numbered sites were sampled). Kick-sites at the same end of a gradient are positively correlated with similar taxa and chemical/physical variables, whereas samples on opposite ends of gradients are negatively correlated. Samples at right angles through the origin are uncorrelated. Vectors indicate direction of change of physical and chemical parameters, and sites are ranked along these vectors based on the strength of their correlation with each parameter.

such a strong dominance of the taxon, contributing to greater variability in composition among sites (and thus a higher distance to centroid).

3.2.1.3.3. Biotic-abiotic relationships

Relationships between BMI assemblages and abiotic parameters were tested with Redundancy Analysis (RDA) to explore potential drivers of assemblage structure in the Slave River. RDA is a multivariate approach that uses environmental variables to constrain the spatial arrangement of sites based on BMI relative abundance. The test acts as a multiple regression on assemblage data, and assesses the amount of variation in the unconstrained ordination (the PCA of BMI samples) that is explained by relating the data to chosen environmental variables, identifying major abiotic gradients in the data. Prior to analysis, correlations between environmental parameters were examined in combination with the abiotic PCAs to pick out important drivers of differences among sites that were uncorrelated with each other (low correlations between environmental parameters were chosen to avoid multicollinearity). This also

worked to reduce the number of environmental parameters in the analysis and avoid over-fitting the data. As suggested by the abiotic PCA (Figure 9), many of the water chemistry parameters were highly correlated. In particular, concentrations of metals were generally all correlated, and they were also correlated with alkalinity and conductivity. Rather than choosing one particular metal to include in the analysis, conductivity was chosen to reflect ionic composition in the water as well as the many metals that were correlated with it. The final RDA for water chemistry and physical habitat variables included velocity, % pebble, % bedrock, conductivity, nitrate (NO_3), ammonia (NH_3), and orthophosphate (OrthoP). Other ions, nutrients, physical measures, and total and dissolved metals were highly correlated with the chosen variables. Thus, any patterns described for these parameters also apply to the correlated parameters.

The first axis and all axes of the RDA of BMI relative abundance and chemical and physical parameters were statistically significant (Monte Carlo permutation test: first axis $F = 20.21$, $p = 0.028$; all axes $F = 3.16$, $p = 0.030$), and the first three axes explained 99.7% of the constrained variation among Slave River samples (68.7% of the unconstrained variation). Similar to the PCA, the constrained ordination primarily separated sites in Reach 1 from most sites in other reaches along the first axis (Figure 15), which explained 66.9% of unconstrained variation in assemblage structure (97.1% of constrained variation). Sites in Reach 1 were positively correlated with velocity, conductivity (and thus metals), and the % pebble. The positive correlation with velocity was logical given that the taxa that were most strongly associated with this reach are generally found in faster flows (Monk et al. 2008). Even with higher discharge and water levels in 2020, velocity differences between reaches appeared to have an impact on assemblage structure. The sites that were positively correlated with *Hydra* on the opposite end of the first axis gradient were negatively associated with velocity and conductivity, but also positively associated with bedrock (Figure 15). Along the second axis of the RDA, which only explained 1.4% of the unconstrained variance in assemblage structure, sites appeared to vary along a nutrient gradient.

The statistical significance of chemical and physical habitat variables in the RDA was tested using Monte Carlo permutational tests. The variables with the strongest effect on the fit of the model were conductivity ($F = 6.88$, $p = 0.016$) and velocity ($F = 4.75$, $p = 0.038$). No other variables were significant, though the % bedrock and % pebble had the next highest conditional effects in the model.

3.2.2. Temporal characterization of BMI assemblages

3.2.2.1. Benthic macroinvertebrate composition

Compositional changes from 2017 to 2020 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. Previously, this assessment focused on insect groups and major non-insect groups excluding *Hydra*. However, given the dominance of this taxon in 2020, it was important to assess changes relative to *Hydra* across the four years of sampling. The relative abundance of *Hydra* was highest in 2018, when it accounted for more than one-quarter of the assemblage on average, and in 2020, when it accounted for more than half of the assemblage on average (Figure 16). The relative abundance of *Hydra* was similar and much lower in both 2017 and 2019, when antecedent median discharge was lower. Also apparent was the decline in relative abundance of Diptera (true flies) from 2017 to 2018, and the concurrent increase in relative abundance of Trichoptera (caddisflies; Figure 16). Increases in relative abundance of other mobile taxa such as Ephemeroptera (mayflies), Plecoptera (stoneflies), and Hemiptera (true bugs)

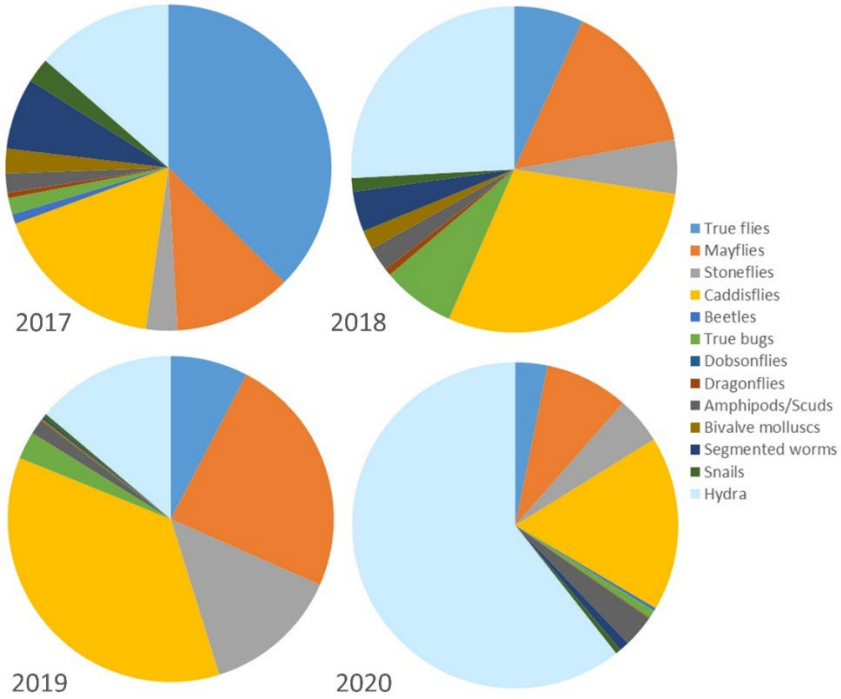

Figure 16. Average relative abundance of major taxonomic groups in Slave River kick samples collected in 2017, 2018, 2019, and 2020. Taxa are grouped as true flies (Diptera), mayflies (Ephemeroptera), stoneflies (Plecoptera), caddisflies (Trichoptera), beetles (Coleoptera), true bugs (Hemiptera), dobsonflies (Megaloptera), dragonflies (Odonata), amphipods/scuds (Amphipoda), bivalve molluscs (Bivalvia), segmented worms (Oligochaeta), snails (Gastropoda), and Hydra.

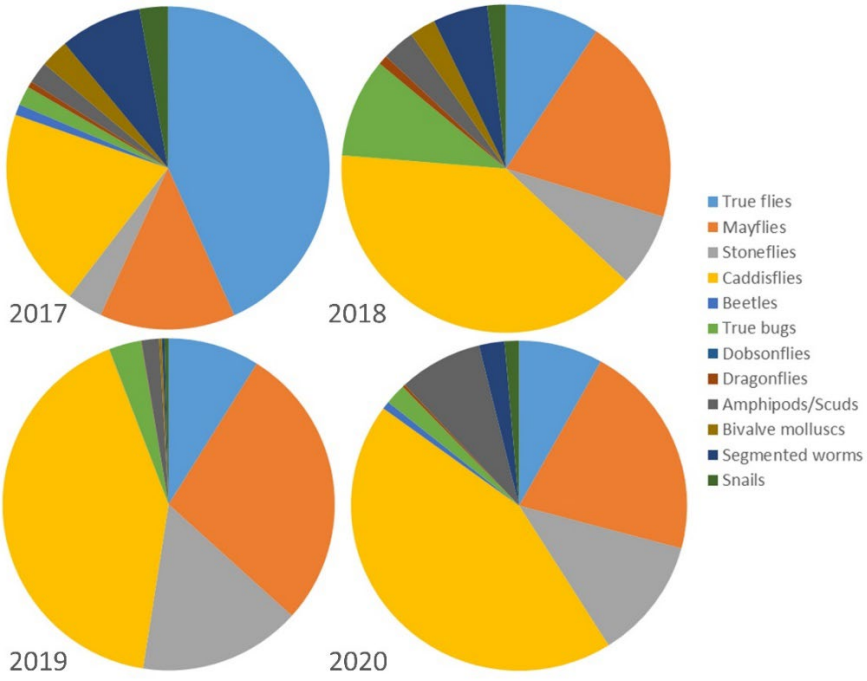

Figure 17. Average relative abundance of major taxonomic groups in Slave River kick samples collected in 2017, 2018, 2019, and 2020, calculated after the exclusion of Hydra. Taxa are grouped as true flies (Diptera), mayflies (Ephemeroptera), stoneflies (Plecoptera), caddisflies (Trichoptera), beetles (Coleoptera), true bugs (Hemiptera), dobsonflies (Megaloptera), dragonflies (Odonata), amphipods (Amphipoda), bivalve molluscs (Bivalvia), segmented worms (Oligochaeta), and snails (Gastropoda).

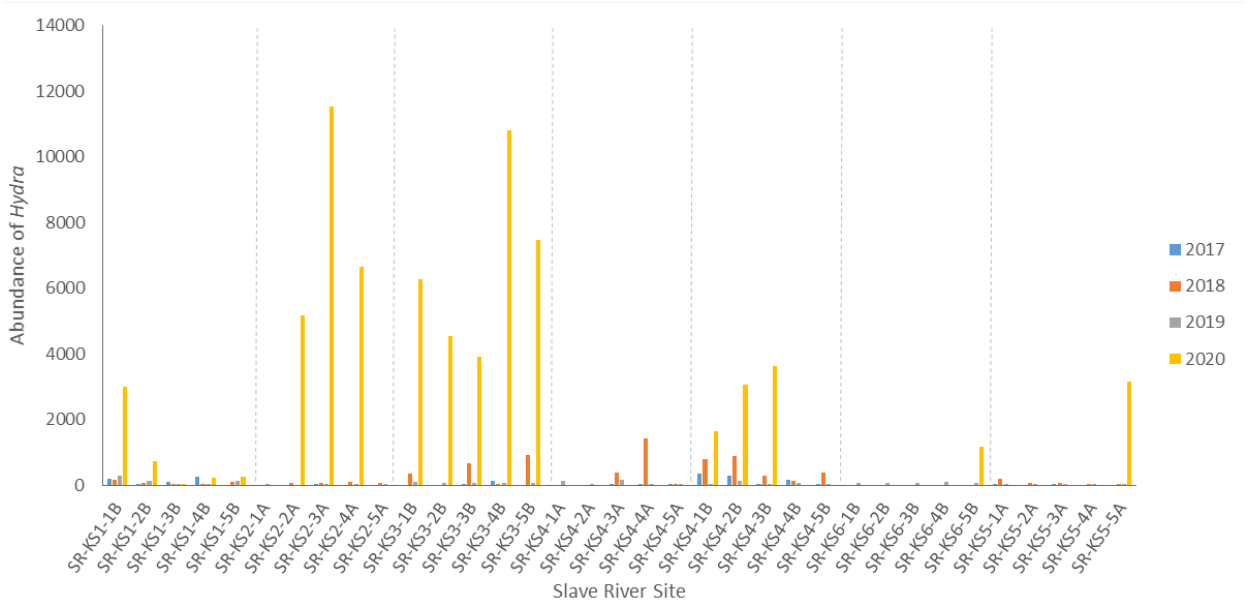

Figure 18. Abundance of Hydra *in Slave River kick sites in 2017 (blue bars), 2018 (orange bars), 2019 (grey bars), and 2020 (yellow bars). Light grey vertical dashed lines denote reaches.*

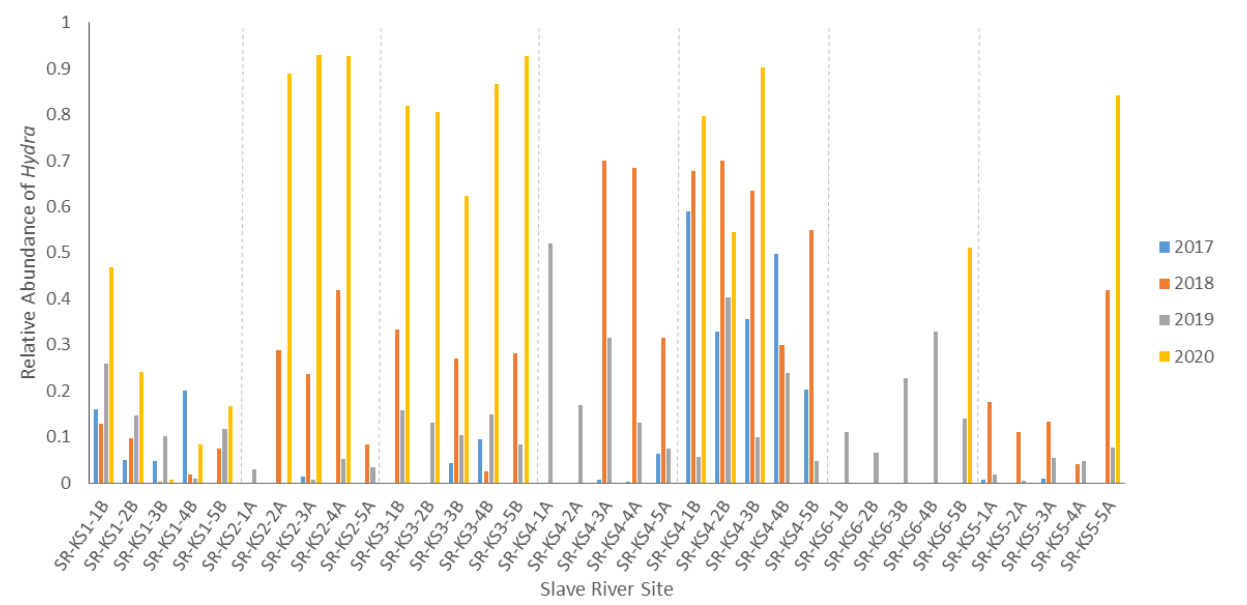

Figure 19. Relative abundance of Hydra *in Slave River kick sites in 2017 (blue bars), 2018 (orange bars), 2019 (grey bars), and 2020 (yellow bars). Light grey vertical dashed lines denote reaches.*

were also evident in 2018. Average relative abundance of several taxonomic groups were similar in 2019 to what was found in 2018, particularly for true flies. However, there was an increase in relative abundance of EPT groups (mayflies, stoneflies, and caddisflies), and a decrease in true bugs and non-insects (including molluscs, snails, and segmented worms) relative to 2018 (Figure 16). With the dramatic increase in abundance of *Hydra* in 2020, relative abundances of all other taxonomic groups declined.

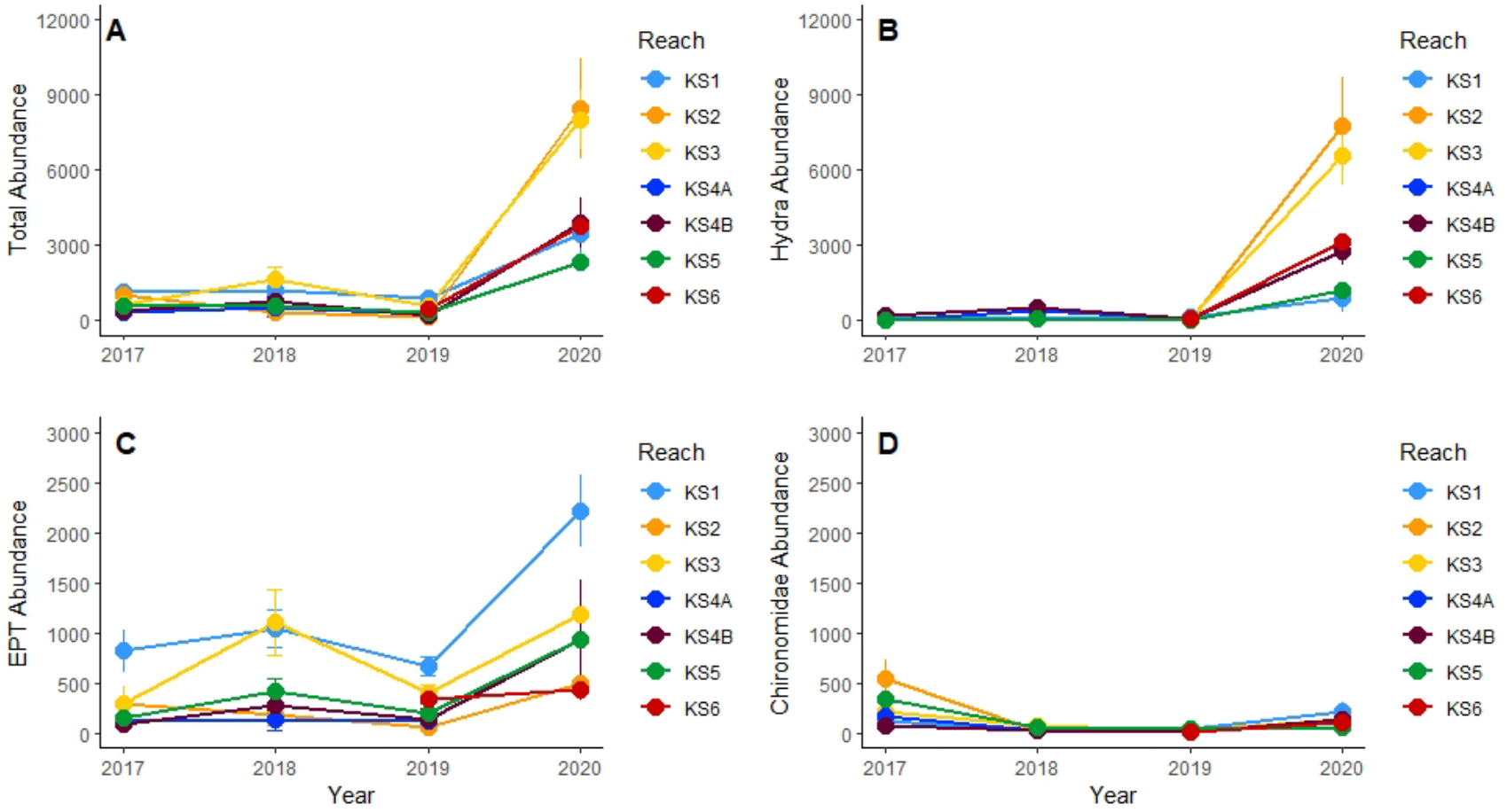

Figure 20. Line plots of changes over time in (A) total abundance, and abundance of (B) Hydra*, (C) Ephemeroptera, Plecoptera, Trichoptera (EPT), and (D) Chironomidae (midges) in the Slave River, showing the mean ± SE for each reach in 2017, 2018, 2019, and 2020.*

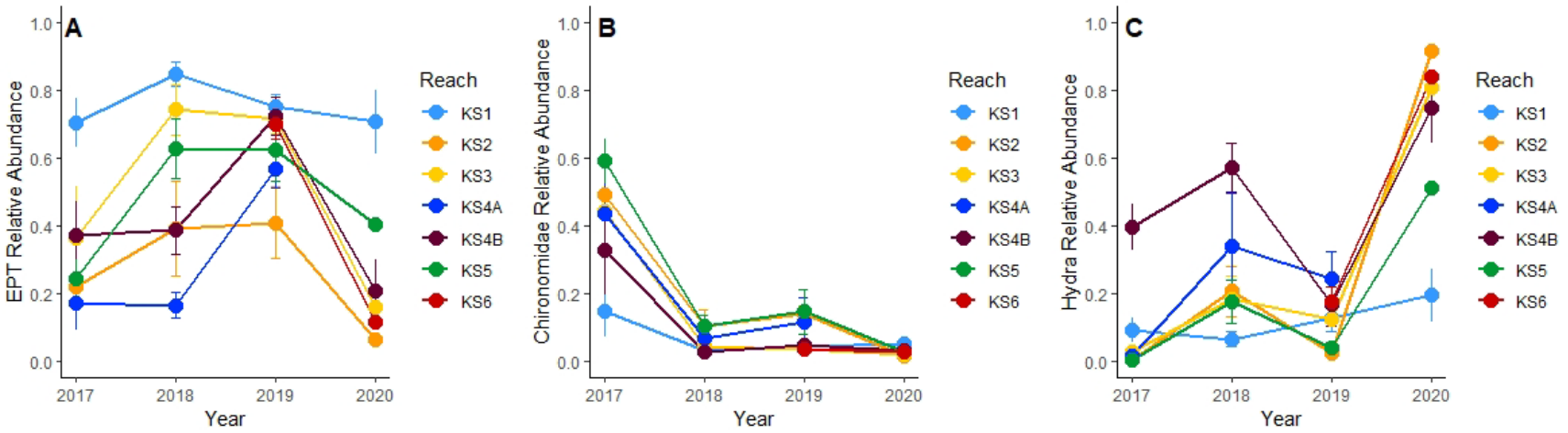

Figure 21. Line plots of changes over time in relative abundance of (A) Ephemeroptera, Plecoptera, and Trichoptera (EPT), (B) Chironomidae (midges), and (C) Hydra *in the Slave River, showing the mean ± SE for each reach in 2017, 2018, 2019, and 2020.*

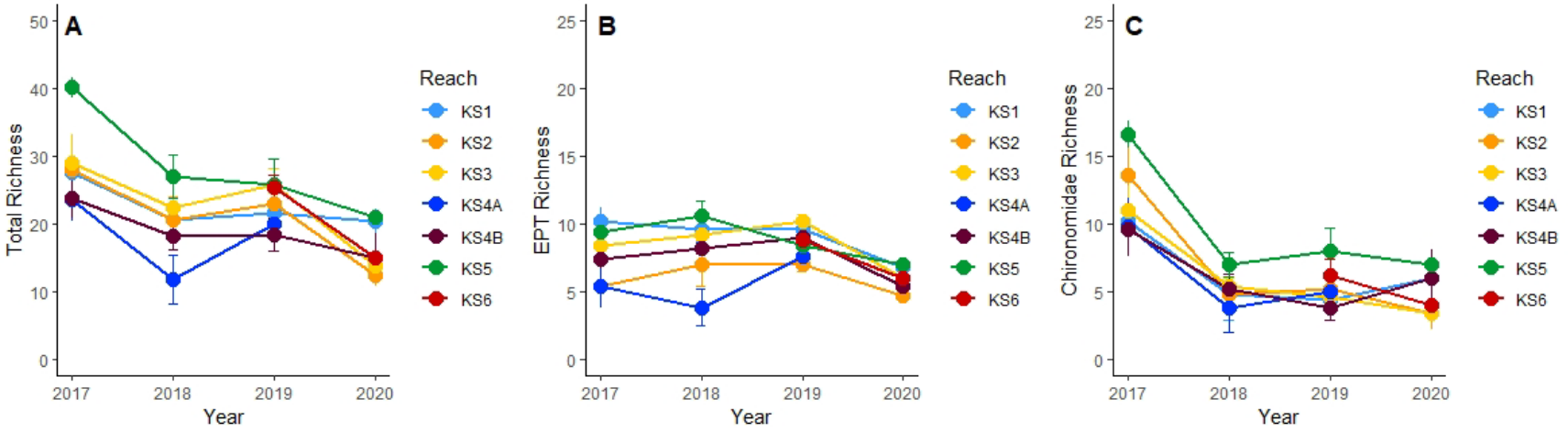

Figure 22. Line plots of changes over time in (A) total taxonomic richness and richness of (B) Ephemeroptera, Plecoptera, and Trichoptera (EPT), and (C) Chironomidae (midges), in the Slave River, showing the mean ± SE for each reach in 2017, 2018, 2019, and 2020.

Relative abundance of major taxonomic groups was also assessed in the absence of *Hydra* to better visualize relative shifts over time (Figure 17). In the absence of *Hydra*, Diptera (true fly) remained at a similar abundance relative to other major non-*Hydra* taxa from 2018 to 2019, as did Trichoptera (caddisflies). Ephemeroptera (mayflies) were more variable among years, with the highest abundance relative to other groups in 2019. Amphipods, which are strong swimmers and typical of deep, slow-flowing water, increased in abundance relative to other groups in 2020, which may have reflected the deep water conditions in that year. However, Figure 17 must be interpreted with the understanding that patterns for 2020 in particular represent a small portion of the assemblage once *Hydra* is considered.

Changes in the dominance of *Hydra* over time were further assessed through plots of the abundance (Figure 18) and relative abundance (Figure 19) of this genus in each site across the four sampling years. The abundance of this genus was sufficiently high in 2020 that it did not appear to be comparable with previous years (Figure 18). When expressed as relative abundance, it was clear that *Hydra* increased the most in 2020 relative to previous years in Reach 2 and Reach 3, as well as in the single sites of Reach 6 and Reach 5 that were sampled (Figure 19). Both Reach 2 and Reach 3 displayed low relative abundance of *Hydra* in 2017 and 2019, increased relative abundance in 2018, and a very large increase in 2020 that represented at least a doubling of 2018 levels. In contrast, relative abundance of *Hydra* in Reach 4B appeared to be high in all years of sampling, which suggests that this reach contains ideal habitat for *Hydra*.

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 20). Mean abundance appeared to increase in all reaches in 2020 relative to earlier years, and 2017-2019 appeared to be relatively invariable compared to the change in 2020 (Figure 20A). The increased abundance was clearly driven by the increase in abundance of *Hydra*, as indicated in Figure 20B. Although at a much smaller scale, the abundance of EPT also increased in most reaches in 2020 relative to earlier years (Figure 20C), while the abundance of Chironomidae showed a decline from 2017 to 2018 and a slight increase in 2020 (Figure 20D). Although Lento (2021) described these metrics as being variable over time when only data from 2017-2019 were considered, this variability appeared to be minor relative to the changes that took place in 2020.

Relative abundance metrics showed more variability over time than abundance metrics. The relative abundance of EPT was particularly variable, with little consistency among reaches with the exception of a sharp decline in 2020 in most reaches (except Reach 1, which maintained a similar relative abundance across all years, and Reach 5 which had a smaller decline in relative abundance of EPT than other reaches; Figure 21A). In contrast, most reaches showed the same pattern over time for the relative abundance of Chironomidae, with a sharp decrease in 2018 and similar values from 2018-2020 (Figure 21B). The patterns in relative abundance of *Hydra* over time were also somewhat consistent among sites, with an increase in most sites in 2018, a decline in 2019, and a sharp increase in most sites in 2020 (Figure 21C).

Richness metrics showed similar trends over time in most reaches. Total richness declined in 2018 and again in 2020, coinciding with the loss of Chironomidae in 2018 and the increased abundance of *Hydra* in 2020 (Figure 22A). Similarly, Chironomidae richness declined in 2018, but remained fairly similar from 2018-2020 (Figure 22C). The richness of EPT remained similar from 2017-2019 in most reaches, but

declined in all reaches in 2020 (Figure 22B). Together, these patterns point to a loss of diversity in 2020 when water levels were high, and *Hydra* dominated most reaches.

3.2.3. Normal range and CES

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.

Initial normal range estimates were developed by Lento (2021) using data from 2017 to 2019, and this report expands those estimates to include data from 2020. 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 separate the data to develop flow-based normal range criteria. This idea is considered in Section 3.4, as flow-ecology relationships are explored. However, in this section, normal range and CES estimates are evaluated with all four years of data included, to assess which metrics were most affected by variability in 2020.

Previously, 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 consideration of additional metrics. Both abundance of *Hydra* and relative abundance of *Hydra* were added to the analysis to consider variability over time. 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 was expected to cover a large range for most metrics.

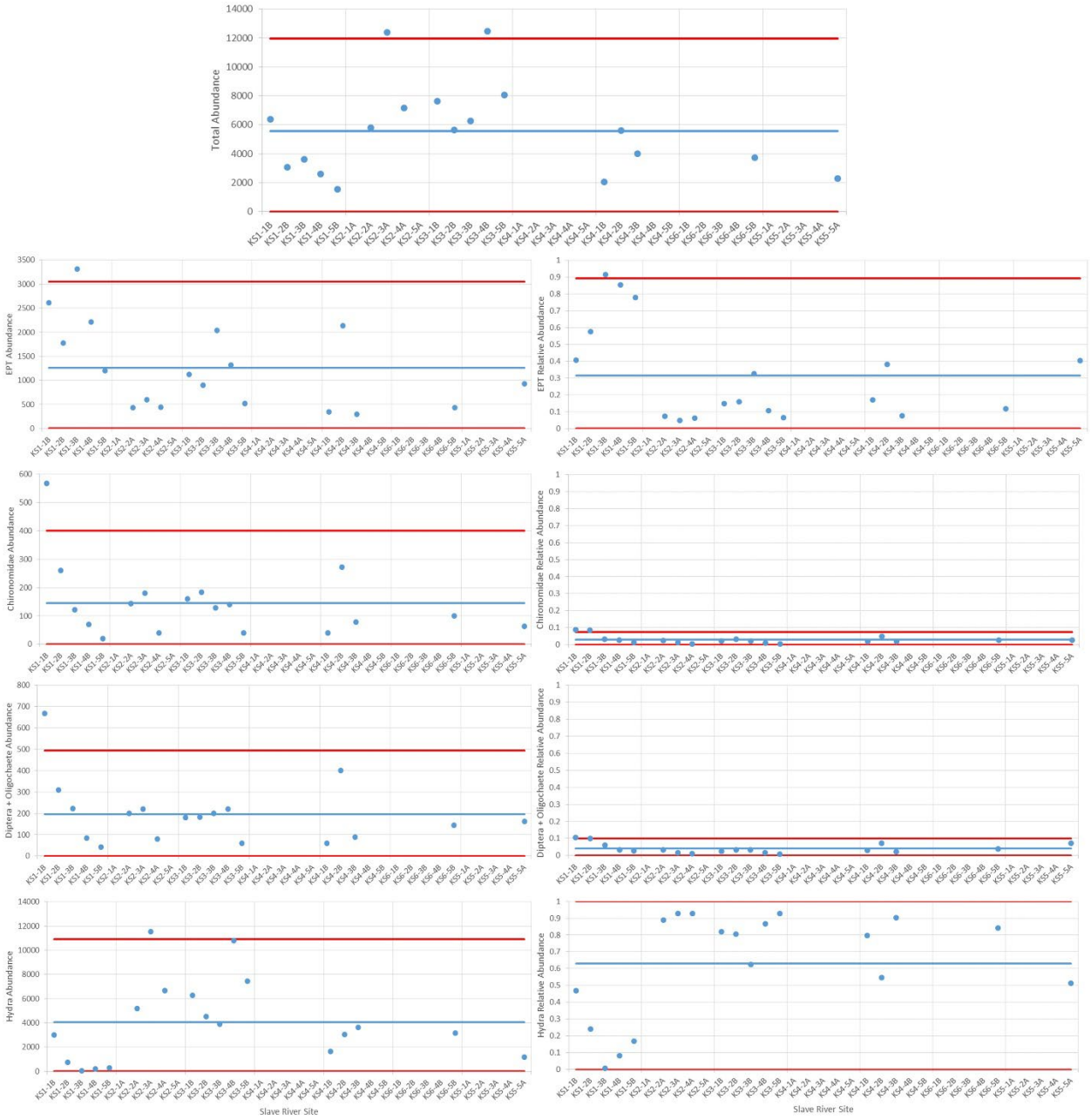

Figure 23. Abundance-based biotic metrics plotted for each site in the Slave River with the 2020 mean (blue line) and the upper and lower single-year critical effect size (CES; red lines), calculated as mean ± 2SD (calculated based on 2020 data). Each point represents a kick-site, moving from reach 1, site 1 (far left) to reach 5, site 5 (far right) on each plot, with gaps for those sites not sampled in 2020. 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.*

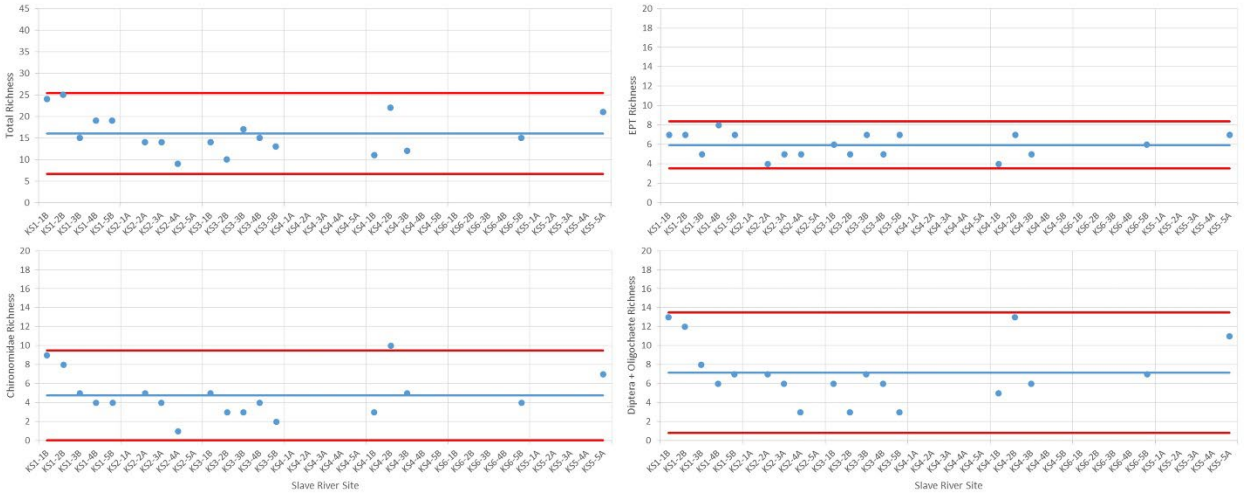

Figure 24. Richness-based biotic metrics plotted for each site in the Slave River with the 2020 mean (blue line) and the upper and lower single-year critical effect size (CES; red lines), calculated as mean ± 2SD (calculated based on 2020 data). Each point represents a kick-site, moving from reach 1, site 1 (far left) to reach 5, site 5 (far right) on each plot, with gaps for those sites not sampled in 2020. Metrics include (left column) total richness, EPT richness, (right column) Chironomidae richness, and Diptera + Oligochaeta richness.

3.2.3.1. 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 2020, and the multi-year CES, calculated based on the mean and standard deviation of combined 2017-2020 data. Single-year and multi-year CES boundaries were compared with site data from 2020 to evaluate how much the samples collected in 2020 varied amongst each other and relative to all previous years.

Samples collected in the Slave River in 2020 generally fell within the CES boundaries developed using only 2020 data, though the boundaries were also large for some metrics due to variability among reaches (Figure 23, Figure 24). For example, total abundance was much higher in Reach 2 and Reach 3 than in the other reaches, and the CES therefore ranged from 0 to 12000 (Figure 23). The CESs for other abundance metrics were similarly large, but there were still exceedances of upper boundaries for EPT abundance and Chironomidae abundance in Reach 1, and exceedances of the upper boundary for *Hydra* abundance in Reach 2. Relative abundance metrics also had a wide calculated normal range for *Hydra* and EPT, with both covering nearly the full range of possible values. In both cases, this was due to the high relative abundance of EPT/low relative abundance of *Hydra* at Reach 1 and the high relative abundance of *Hydra*/low relative abundance of EPT at the remaining reaches, which resulted in a wide spread of possible values across all reaches (Figure 23). In contrast, the normal range was narrow for the relative abundance of both Chironomidae and Diptera+Oligochaeta, both of which were low across all reaches. However, it should be noted that due to low abundances of both metrics, only an increase (above the upper CES) would be detectable, as the lower boundary of the normal range was at zero. Normal range boundaries were also more narrow for richness metrics, as variability in each of these metrics was much lower across all reaches (Figure 24). In particular, the normal range for EPT richness

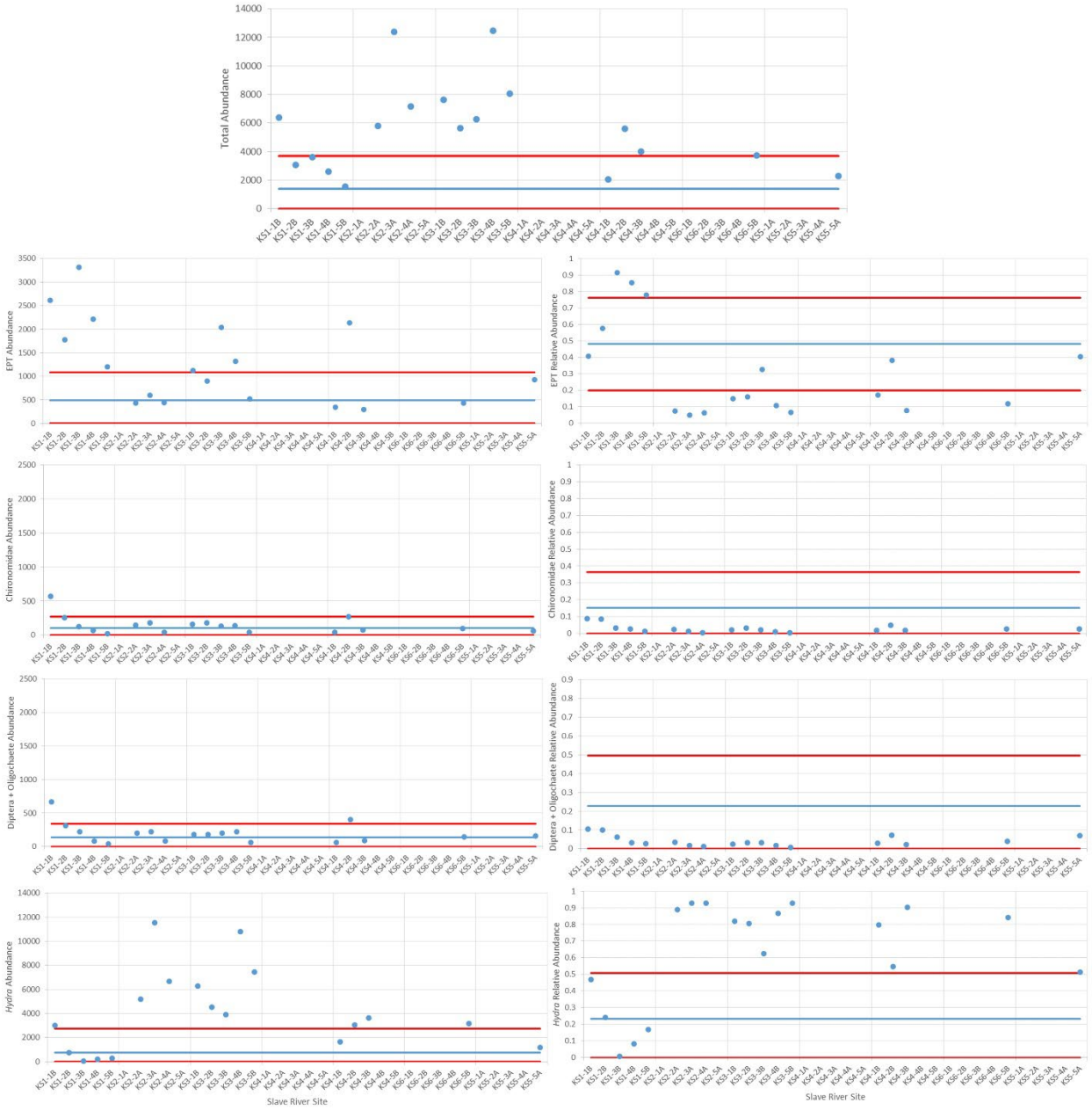

Figure 25. Abundance-based biotic metrics plotted for each site in the Slave River with 2020 data (blue points), a multi-year mean (mean of 2017-2020 data; blue line) and the upper and lower multi-year critical effect size (CES; red lines), calculated as mean ± 2SD (calculated based on 2017, 2018, 2019, and 2020 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.*

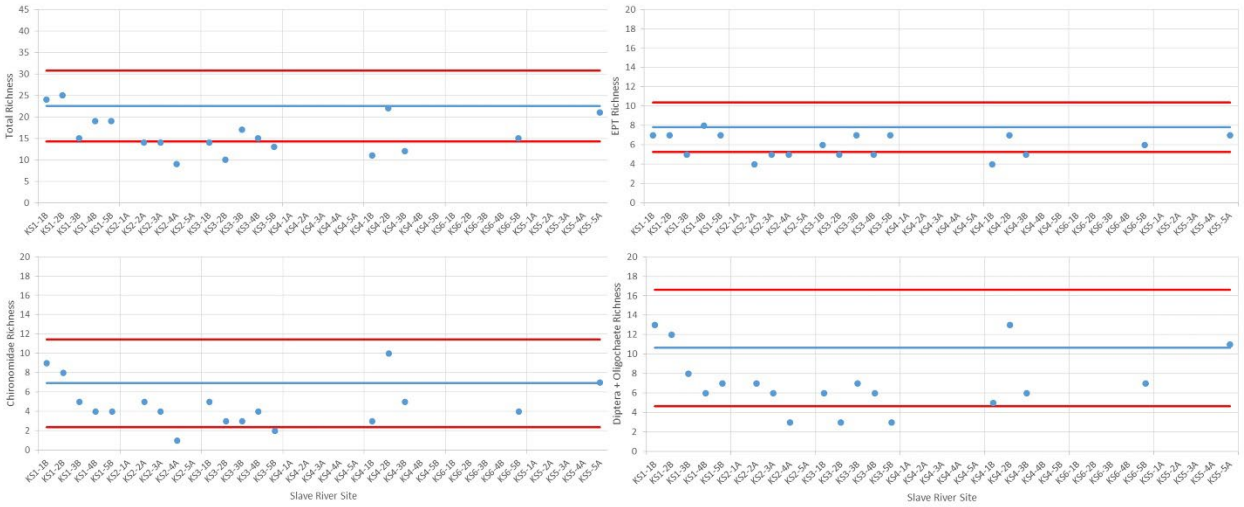

Figure 26. Richness-based biotic metrics plotted for each site in the Slave River with 2020 data (blue points), a multi-year mean (mean of 2017-2020 data; blue line) and the upper and lower multi-year critical effect size (CES; red lines), calculated as mean ± 2SD (calculated based on 2017, 2018, 2019, and 2020 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 low, varying between approximately 4 and 8 taxa, allowing for detection of values above or below the normal range.

Comparison of 2020 data with the multi-year CES developed based on data from 2017-2020 provided a better picture of how variable samples collected in 2020 were relative to previous years. There were a large number of deviations from the normal range when 2020 data were compared with the multi-year CES. For example, all samples from Reach 2 and Reach 3, as well as two of the samples from Reach 4B exceeded the upper CES for total abundance (Figure 25). EPT abundance and *Hydra* abundance also exceeded the upper CES for a number of sites, with EPT abundance being high in Reach 1 and *Hydra* abundance being high in Reach 2, Reach 3, and Reach 4B. In contrast, both the abundance of Chironomidae and abundance of Diptera+Oligochaeta generally fell within the narrow bounds of the multi-year CES. This comparison of 2020 data with multi-year CES for Chironomidae and Diptera+Oligochaeta abundance metrics suggests there is possible utility in the use of these metrics for quantifying normal range in the river, despite the large differences in abundance that were observed from 2017-2018. Abundances of both groups have been relatively stable since then.

Relative abundance metrics most clearly showed the effect of the high abundance of *Hydra* in 2020, with large deviations from normal range evident for the relative abundance of EPT (most sites below the lower CES) and *Hydra* (most sites above the upper CES), and the relative abundance of Chironomidae and Diptera+Oligochaeta near the lower boundary of the multi-year normal range (Figure 25). The relative utility of relative abundance and absolute abundance, the latter of which was formerly considered to be the more variable, differed greatly in 2020 when there was such a strong change in composition and abundance.

Taxonomic richness was noted to be low in several reaches in earlier analyses, and these patterns were confirmed when 2020 samples were compared with multi-year CES. Richness values for 2020 fell below

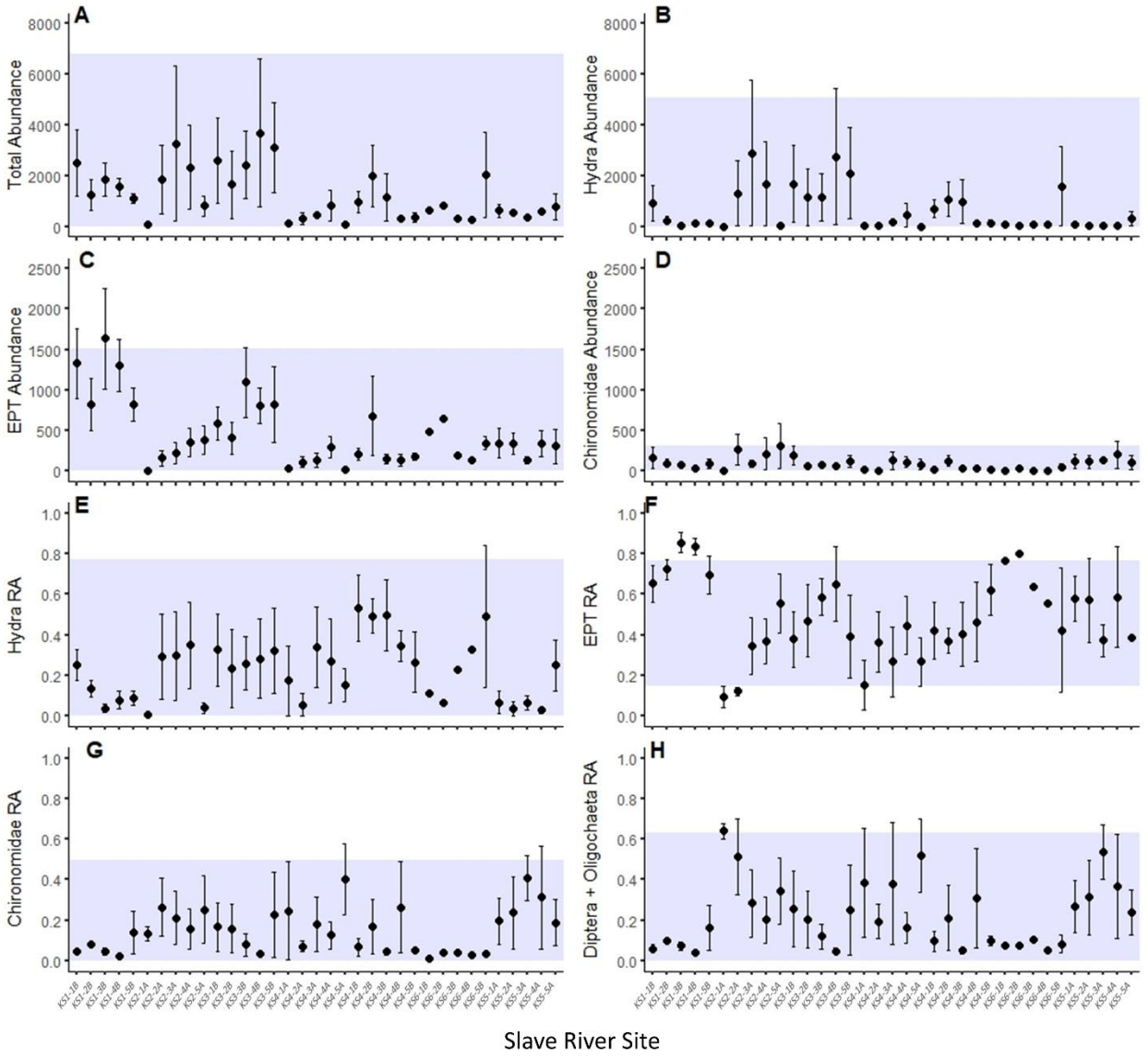

Figure 27. Site-scale temporal variability in abundance-based biotic metrics in the Slave River, showing mean ± SE for 2017-2020 for each site, with the grand mean (mean of annual means for the river) ± 2SD (normal range for the river) indicated by the shaded area, including (A) total abundance, (B) 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).*

the lower CES for all four richness metrics in several sites in Reach 2 and 3 in particular (Figure 26), possibly reflecting the influence of the high abundance of *Hydra* in these reaches.

3.2.3.2. Among-year variability

3.2.3.2.1. Site-scale variability

Temporal variation at the site scale was assessed by comparing the 2017-2020 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, and 2020 means of all sites) ± 2SD. The analysis visualizes temporal variability within sites

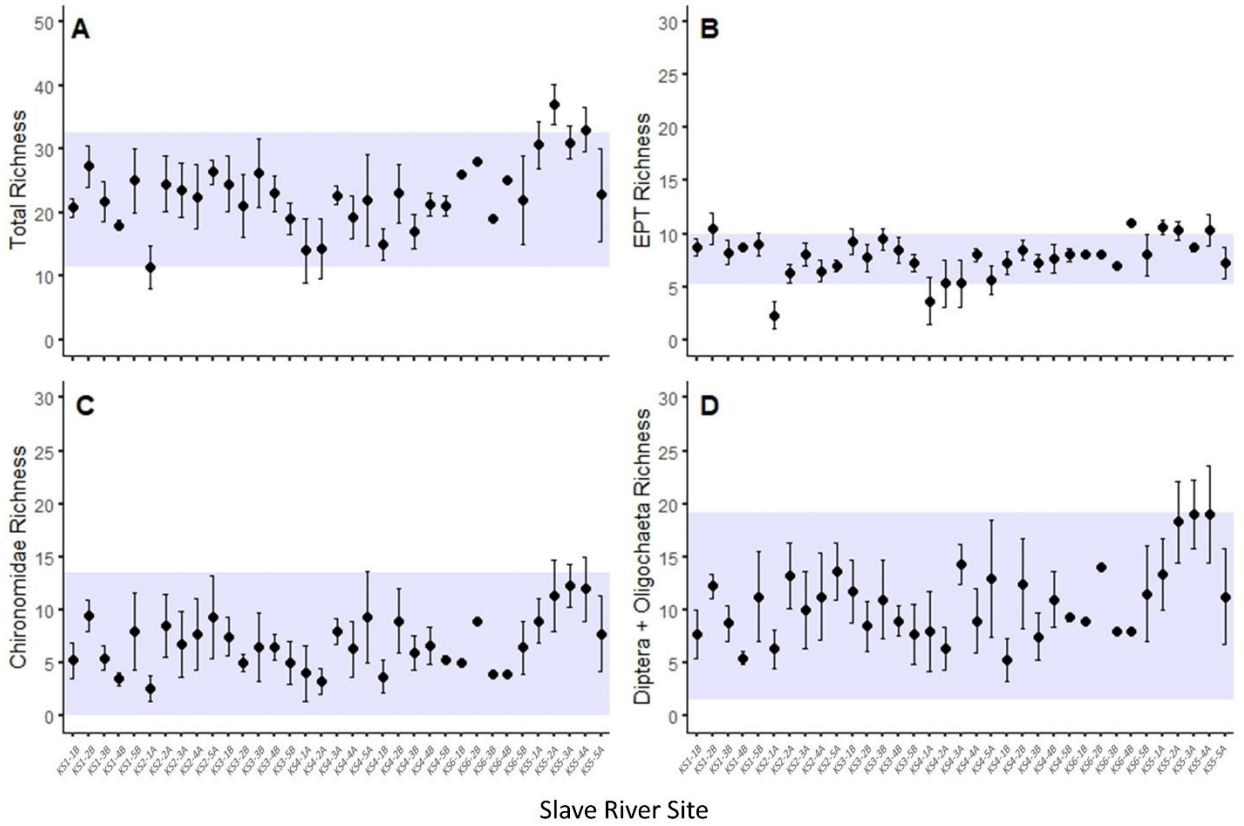

Figure 28. Site-scale temporal variability in richness-based biotic metrics in the Slave River, showing mean ± SE for 2017-2020 for each site, with the grand mean (mean of annual means for the river) ± 2SD (normal range for the river) indicated by the shaded area, including (A) total richness, (B) EPT richness, (C) Chironomidae richness, and (D) Diptera + Oligochaeta richness. Sites are ordered from upstream (KS1-1B, left) to downstream (KS5-5A, right).

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 system. This analysis was completed for abundance metrics, relative abundance metrics, and richness metrics.

Annual means across all sites in the Slave River varied widely between 2017 and 2020 for most metrics, contributing to wide preliminary normal ranges for the river. For example, the preliminary CES boundaries for total abundance ranged from 0 to nearly 7000 due to high abundances in Reach 2 and 3 in 2020 (Figure 27A). The preliminary CES boundaries for *Hydra* abundance were similarly wide (Figure 27B). Boundaries for abundance of EPT were wide due mostly to high abundance in Reach 1, as most other reaches were less variable for this metric (Figure 27C). Only the boundaries for Chironomidae abundance were narrow, suggesting some possible utility to this metric (Figure 27D). Relative abundance metrics were much more variable, both in terms of the width of the estimated normal range and in terms of site means ± SE, which often covered nearly the full range of possible values (Figure 27E-H). Additional years of sample data would help to improve estimates of normal range for these metrics, though the division of years into sub-analyses based on flow or other conditions should be considered to support the development of more narrow normal range criteria. The sharp contrasts in composition between years with differing flow regimes suggests that such an approach would improve the usefulness of normal range estimates by making them more narrow.

The preliminary estimate of the normal range of variability in site-scale analyses was more narrow for richness metrics, but did suffer from wide inter-annual variability in site means across the entire river. The exception was EPT richness, which had an extremely narrow preliminary normal range based on the grand mean (Figure 28), but several sites fell outside the CES boundaries because the normal range was so narrow.

Of the metrics tested for the Slave River, Chironomidae abundance and EPT richness appeared to have the greatest initial potential for developing monitoring and management triggers. Chironomidae abundance would allow for detection of increases in the metric that are outside of the normal range (the lower CES boundary was at zero, and therefore would not allow for detection of a decrease in this metric). In contrast, 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).

3.2.3.2.2. 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, and 2020) ± 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 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 were expected to result in wider normal range boundaries than were observed previously for these reaches.

At the reach scale, there was a great deal of variability in terms of the width of the estimated normal range for each abundance-based metric (Figure 29), reflecting the large changes in abundance in 2020 for some reaches. In particular, the normal range for total abundance was very wide for Reach 2 and Reach 3, which reflected the sharp increase in abundance of *Hydra* in 2020. Most other reaches had much smaller normal ranges for this metric. Similarly, the normal range for EPT abundance was very wide for Reach 1, which reflected the higher abundance of EPT taxa in 2020 (Figure 29). In contrast, all other reaches had very narrow boundaries for EPT abundance, which suggested that this metric might work well to detect temporal changes at the reach scale. The preliminary normal range for Chironomidae abundance was narrow across most reaches as well (with the exception of Reach 2 and Reach 5, both of which had much higher abundance of this group in 2017), and could be considered as a useful metric as more data are collected and the normal range is refined. 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 30). Interestingly, normal range boundaries were more narrow for relative abundance metrics 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, and data would have to be split to

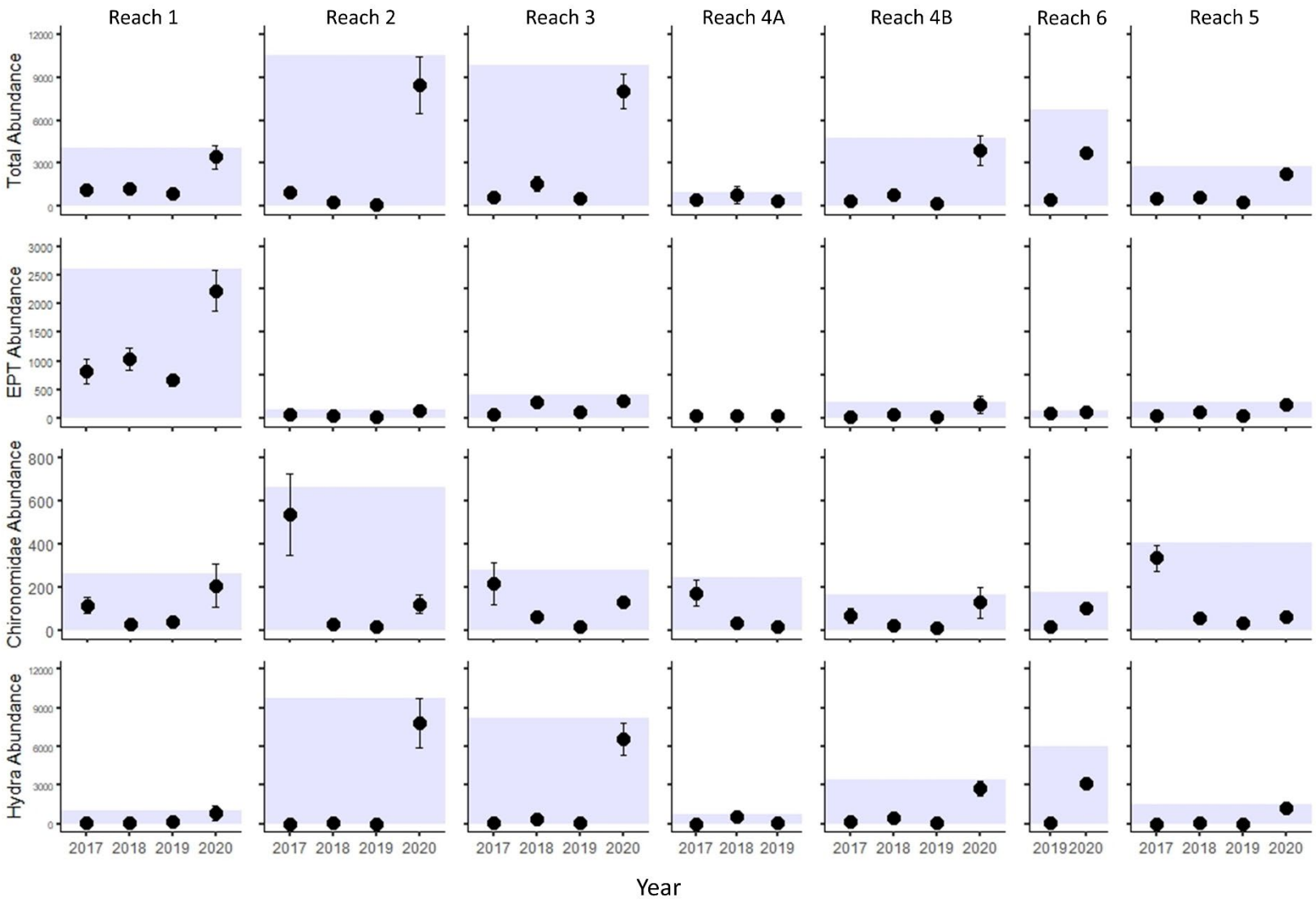

Figure 29. Reach-scale temporal assessment of normal range and critical effect size for abundance-based metrics in the Slave River, including total abundance and abundance of EPT, Chironomidae, and Hydra*. Points represent the mean ± SE across all sites in a reach, plotted for each year (2017, 2018, 2019, and 2020), 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.*

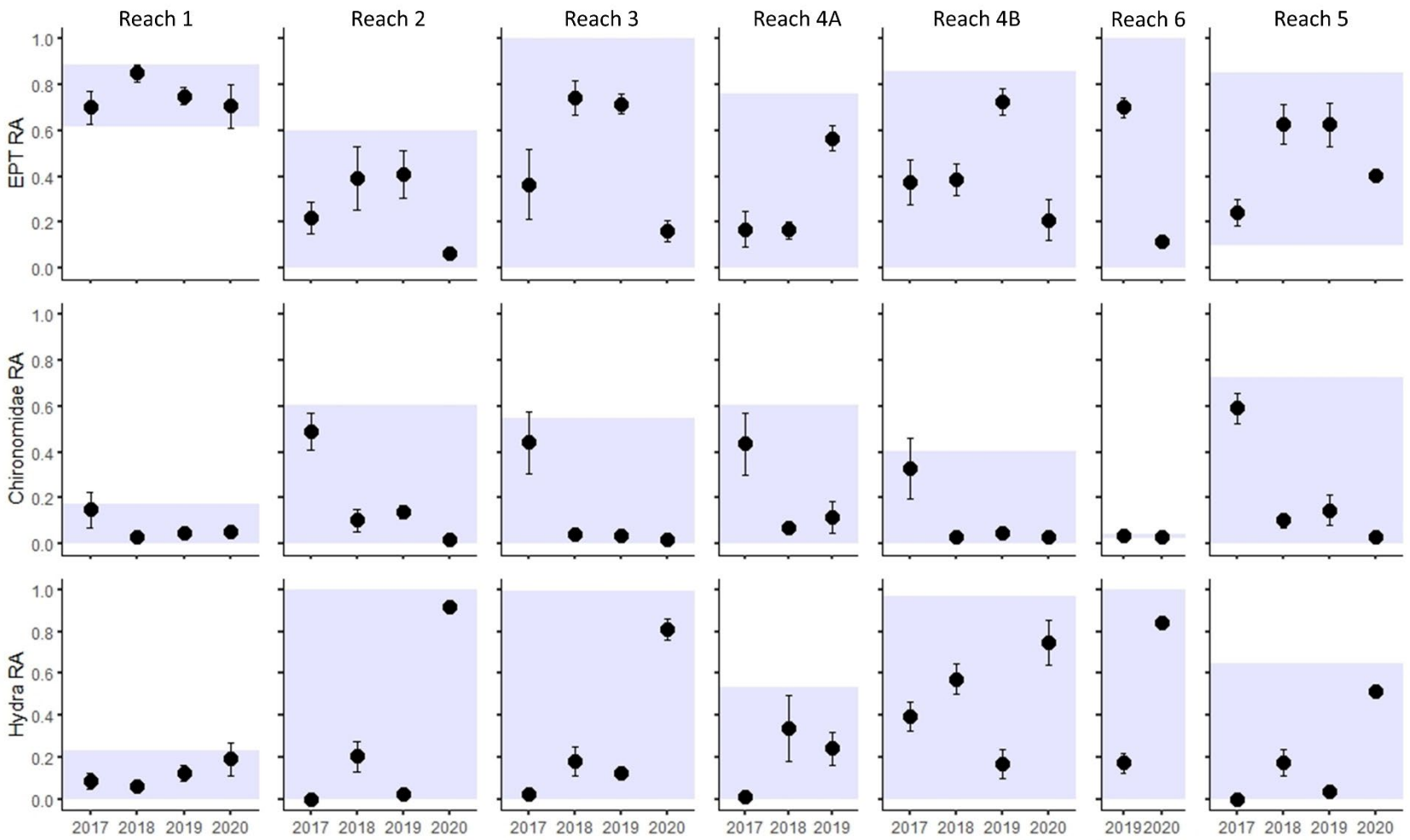

Figure 30. Reach-scale temporal assessment of normal range and critical effect size for relative abundance-based metrics in the Slave River, including relative abundance EPT, Chironomidae, and Hydra. *Points represent the mean ± SE across all sites in a reach, plotted for each year (2017, 2018, 2019, and 2020), 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.*

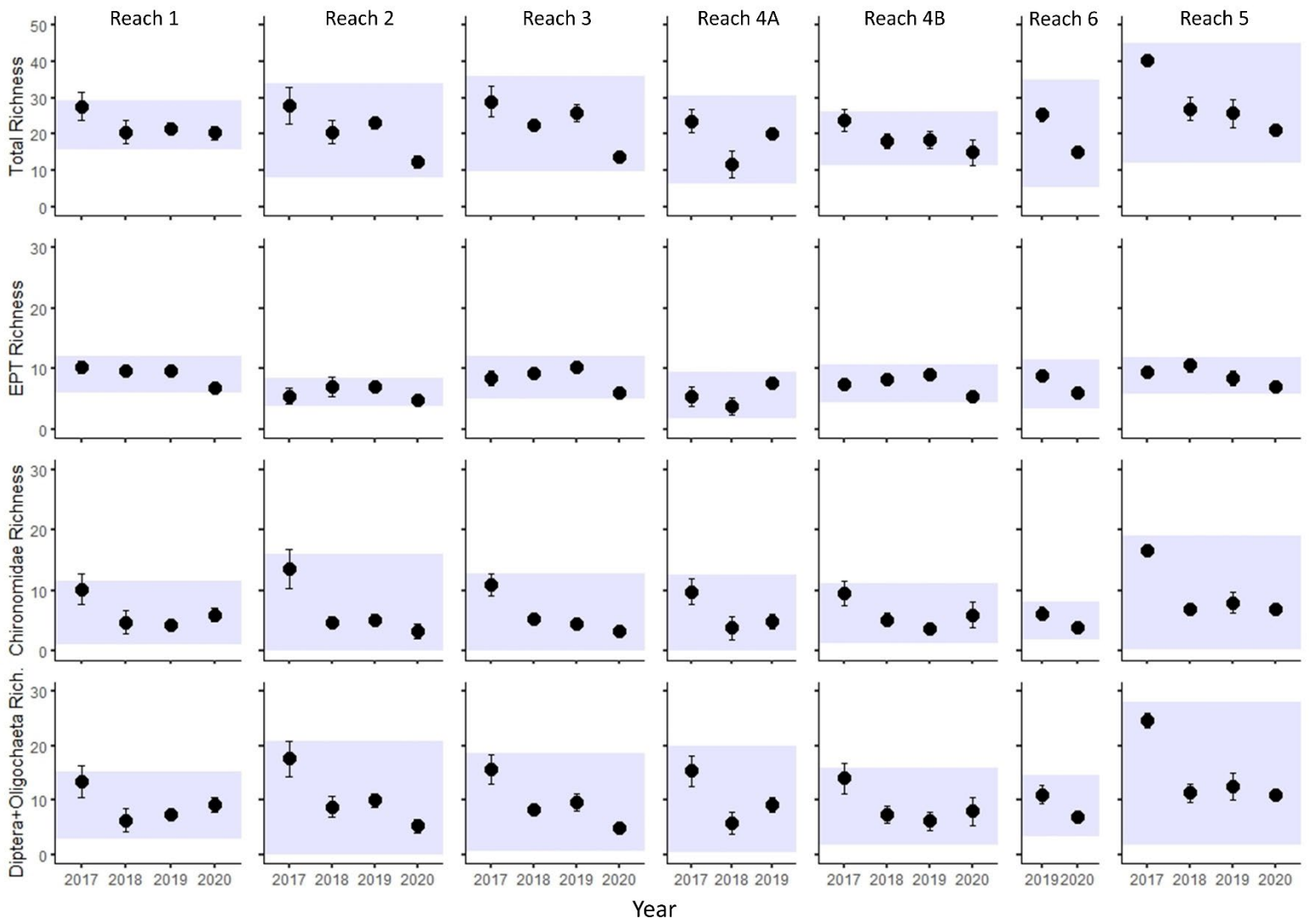

Figure 31. Reach-scale temporal assessment of normal range and critical effect size for richness-based metrics in the Slave River, including total richness, EPT richness, Chironomidae richness, and Diptera + Oligochaeta richness. Points represent the mean ± SE across all sites in a reach, plotted for each year (2017, 2018, 2019, and 2020), 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.

represent different flow conditions or several more years of data would need to be collected to refine and improve their diagnostic power.

Preliminary estimates of the normal range of variability were somewhat more narrow for richness-based metrics, 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 31). Total richness and Chironomidae richness also had fairly low variability in several reaches (with the exception of Reach 5; Figure 31). Much of the variability observed in richness metrics was due to higher richness in most reaches in 2017, and additional years of data will likely help to refine these preliminary normal range estimates.

3.3. Multivariate Normal Range and CES

3.3.1. Hay River

The PCA of Hay River samples from 2017-2019 indicated a strong overlap of samples from 2017 and 2018, and a greater spread and separation of samples from 2019 (Figure 32). The 95% normal probability ellipse for 2017 was the smallest, which indicated greater within-year similarity among samples. The ellipse for 2017 was also completely contained within the ellipse for 2018, which indicated that the composition of BMI in kick samples was extremely similar between those two years. In contrast, the normal probability ellipse for 2019 only partially overlapped with the other two years, and many samples fell well outside the ellipses for 2017 and 2018 (Figure 32). This pattern suggested that the BMI composition of many samples differed significantly in 2019 compared to normal range for the two earlier years of sampling. 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 a stronger association of some taxa with samples collected in 2019 relative to the two earlier years of sampling, and indicates a degree of temporal variation in BMI composition in the Hay River.

At the site level, Procrustes analysis of Hay River data indicated a similar degree of dissimilarity in composition among years, with the sum of squared residuals (m_{12}^2) ranging from 0.27 to 0.31 across all pairwise comparisons. The m_{12}^2 was low for all pairwise comparisons of years, indicating a low degree of change in the position of sites in multivariate space across years. These results indicate that none of the sampling years showed a particularly strong change in individual sites relative to other sites. The comparison of ordinations between years using Procrustes analysis indicates whether the placement of sites changes relative to one another, and would result in high residuals for any site that 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 indicated 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.

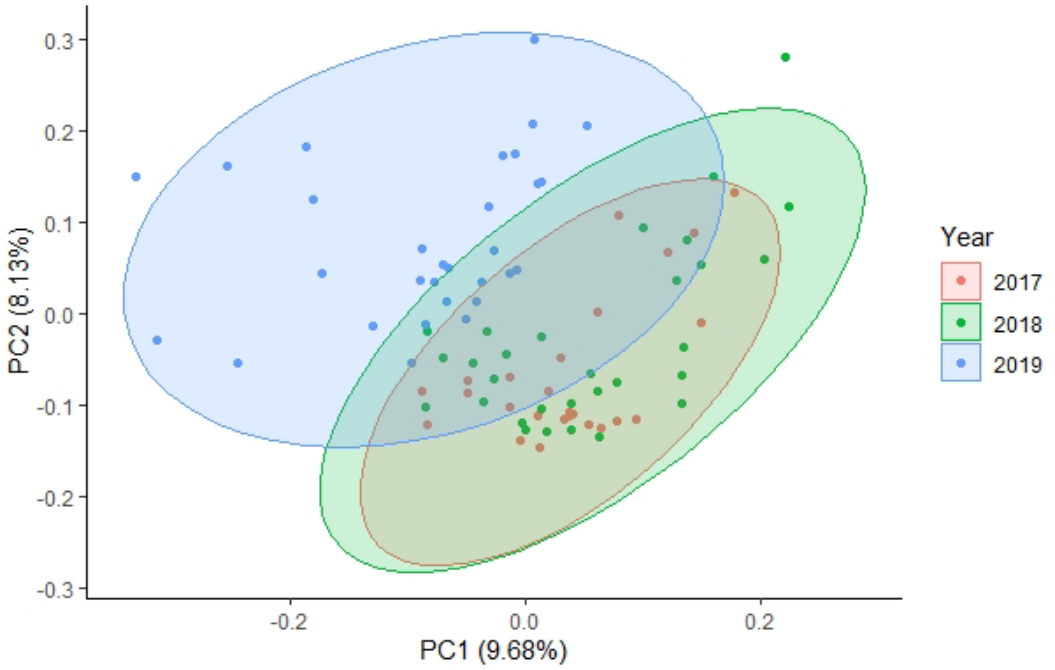

Figure 32 PCA ordination of Hay River samples from 2017 (red), 2018 (green), and 2019 (blue), with 95% normal probability ellipses for each year. Normal probability ellipses indicate the area of multivariate space in which there's a 95% probability of a sample falling if it is representative of the population of samples collected in that year. Overlap of ellipses indicates similar composition in two years of sampling.

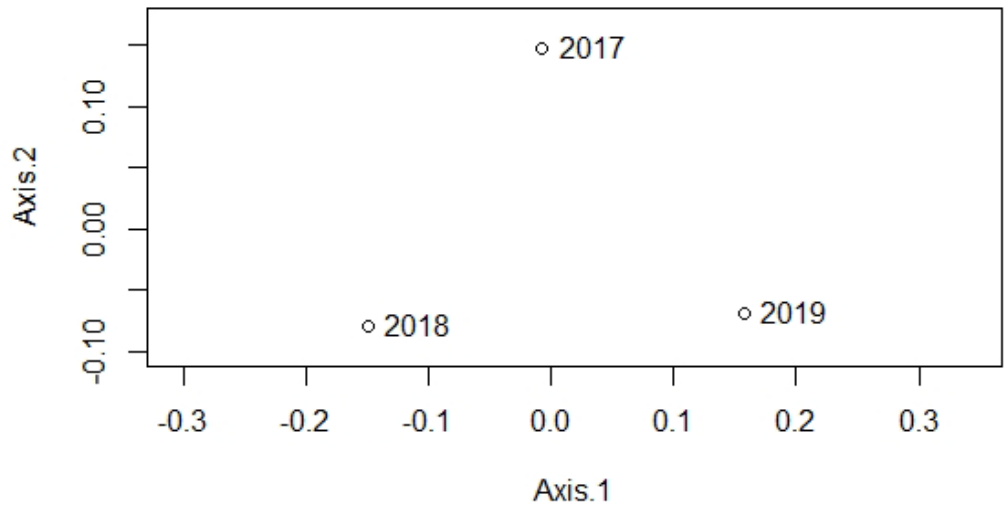

Figure 33 PCoA ordination of a dissimilarity matrix of m_{12}^2 values from pairwise comparisons of years in Procrustes analysis of Hay River samples. Distance between years on the ordination biplot is representative of dissimilarity of samples between years.

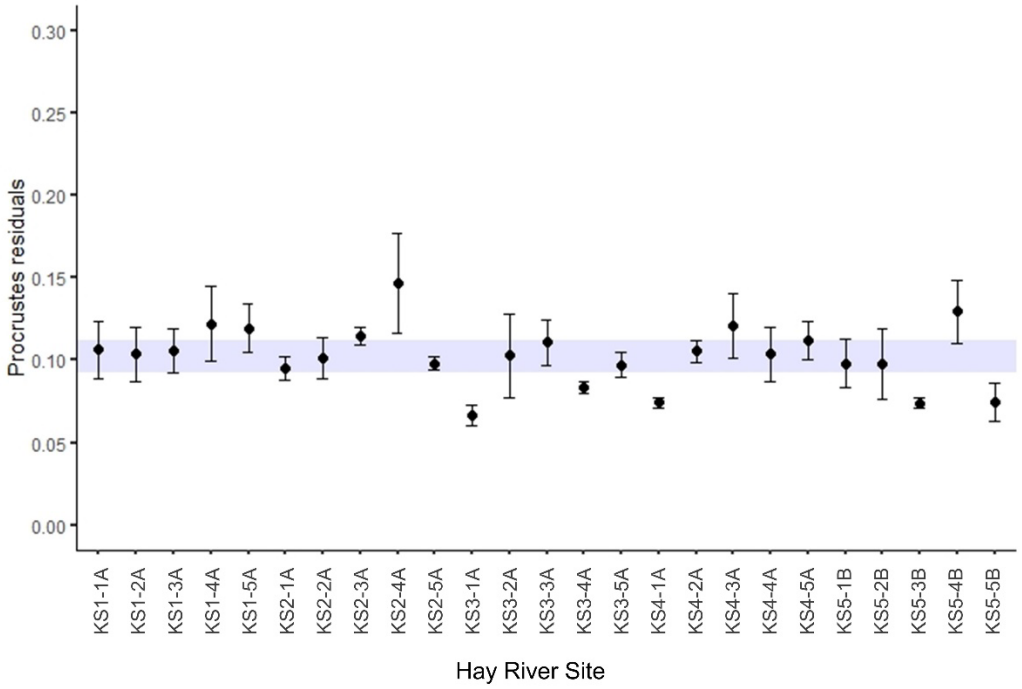

Figure 34 Multivariate normal range and CES boundaries for the Hay River, with the mean ± SE Procrustes residual (2017-2019) plotted for each site and the grand mean (mean of means for each pairwise year comparison) ± 2SD Procrustes residual indicated by the grey shaded area. Only sites sampled in all three years are included.

The generally low but similar m_{12}^2 values among years resulted in a PCoA plot that indicated a similar degree of dissimilarity among all three sampling years (Figure 33). Each point in the PCoA represented one year of sampling, with distance between years in multivariate space reflecting dissimilarity among years, as determined by the matrix of m_{12}^2 values. No sampling year plotted out close to any other sampling year; however, the first axis gradient separated 2018 from 2019, which indicated a greater degree of dissimilarity between these two years. Three years of data is a very small amount for this analysis approach, and with additional years of sampling data, a change trajectory may become more apparent in the plot.

An initial estimate of a multivariate normal range and CES boundaries was created based on the grand mean Procrustes residual (mean of mean residuals for each pairwise year comparison) ± 2 SD, and mean ± SE Procrustes residuals were plotted for each site to identify site-scale temporal variability relative to the normal range (Figure 34). Because the degree of dissimilarity in pairwise comparisons was similar across years, the normal range was narrow, and the mean residual for 11 of the 25 sites fell outside of CES boundaries (above or below CES; Figure 34). Of these sites, 6 had a mean residual above the upper CES, indicating greater temporal change in BMI composition than expected based on the normal range. This included two sites each in Reach 1 and Reach 2, and one site each in Reach 4 and Reach 5 (Figure 34). In contrast, two sites each in Reach 3 and Reach 5 and one site in Reach 4 had low mean residuals

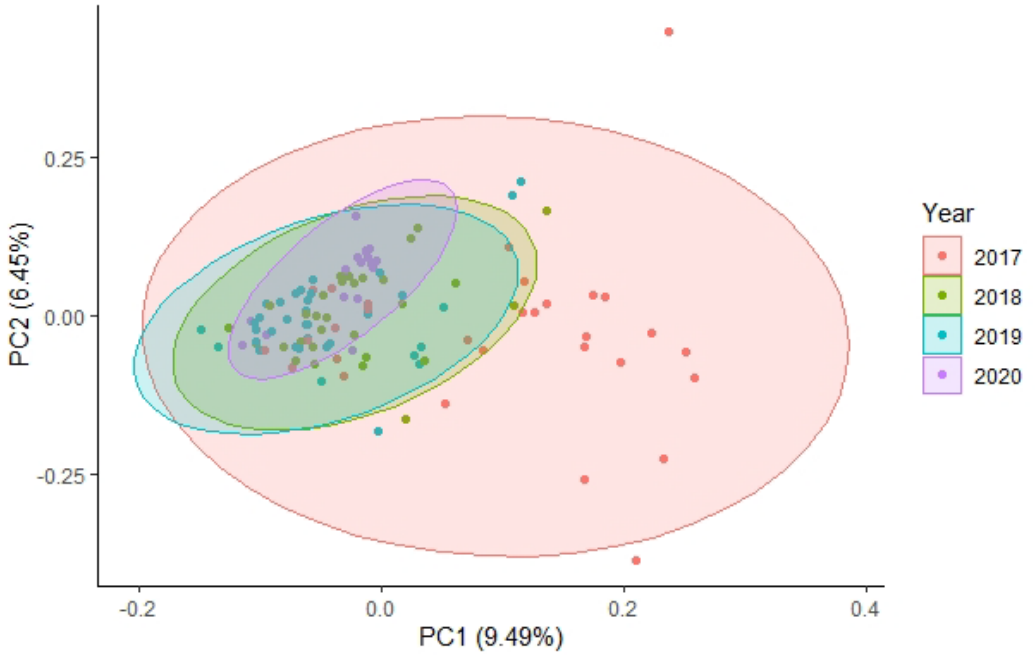

Figure 35 PCA ordination of Slave River samples from 2017 (red), 2018 (green), 2019 (blue), and 2020 (purple) with 95% normal probability ellipses for each year. Normal probability ellipses indicate the area of multivariate space in which there's a 95% probability of a sample falling if it is representative of the population of samples collected in that year. Overlap of ellipses indicates similar composition in two years of sampling.

that fell below the lower CES, indicating less temporal change in composition than expected based on the normal range (Figure 34). Though some sites were more variable than others (larger SE), most sites in Hay River had low mean Procrustes residuals that were within or below the normal range of variability, indicating generally limited site-scale temporal multivariate change. With additional years of data, it will be possible to continue to assess the degree of temporal stability in composition of sites relative to one another, and potentially identify any sites that change composition substantially relative to previous years.

3.3.2. Slave River

Multivariate analysis of temporal variability in Slave River samples yielded stronger patterns that reflected the large compositional shifts that were described by temporal analysis of metrics. The PCA of all sampling years (2017-2020) highlighted the differences between 2017 samples, which included abundant and diverse Chironomidae assemblages, and samples from all other years (Figure 35). 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 2018-2020 overlapped, indicating strong similarity in composition among years. The ellipse for 2020 was the smallest, which reflected both the smaller number of samples (18, compared to 30-35 in other years) and the strong similarity that was observed among sites that were dominated by *Hydra*. Normal probability ellipses for 2018 and 2019 were extremely similar in size and position, indicating a strong similarity in BMI composition among years (Figure 35). The ellipses for 2018-2020 were all contained within the normal probability ellipse for 2017, which was at least twice as large as the other ellipses. This pattern suggests partial similarity of 2018-2020 samples with a subset of samples from 2017, which may

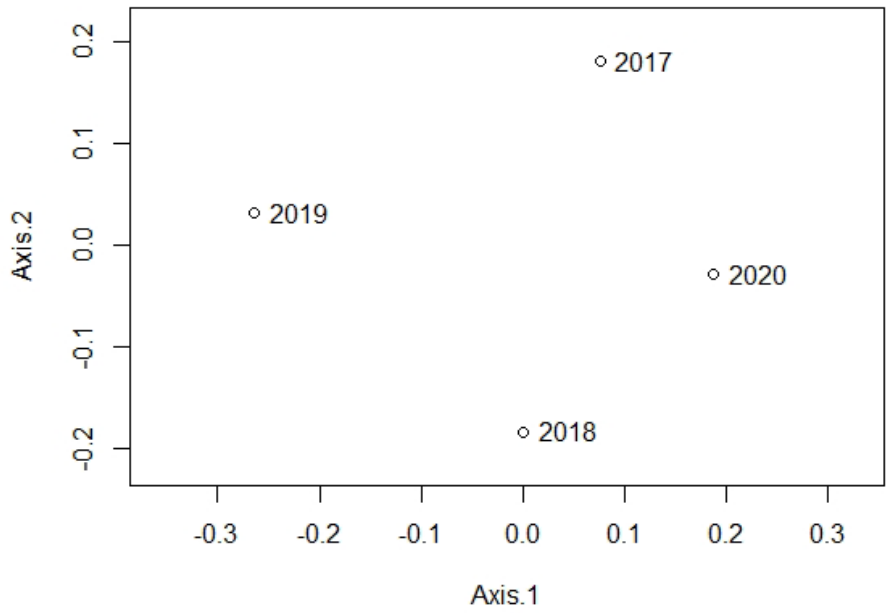

Figure 36 PCoA ordination of a dissimilarity matrix of m_{12}^2 values from pairwise comparisons of years in Procrustes analysis of Slave River samples. Distance between years on the ordination biplot is representative of dissimilarity of samples between years.

be due to associations with taxa that remained abundant in 2018 and later. The varying size and position of normal probability ellipses in the PCA reflected the loss of diversity in 2018 and later years compared to 2017, and indicated low levels of multivariate temporal variability from 2018-2020.

Though overall composition did not vary greatly from 2018-2020, Procrustes analysis indicated site-scale changes among years. The sum of squared Procrustes residuals (m_{12}^2) ranged from 0.29 to 0.46 across all pairwise comparisons, with the two most similar years being 2017 and 2020, and the two most dissimilar years being 2019 and 2020. Values of m_{12}^2 were more variable and covered a wider range of values than was observed for the Hay River, which indicated greater temporal variability in the position of sites in multivariate space. Such variability would occur if BMI composition did not change in the same way in all sites from one year to the next, or if there were temporal changes in composition in some sites but not in others.

The results of the Procrustes analysis indicated that the greatest shift in the relative position of sites in multivariate space occurred between 2019 and 2020, and that the relative position of sites in multivariate space in 2020 was more similar to what was observed in 2017. This pattern was reflected in the PCoA, which showed a large change in position along the second axis gradient from 2017 to 2018, followed by shift along the first axis in 2019, and a larger shift in 2020 that brought the trajectory closer to 2017 than in previous years (Figure 36). When the pattern is examined in the context of the known changes to composition among years, it may reflect the following change trajectory: (1) sites in 2017 were equally dominated by both EPT taxa and Chironomidae; (2) in 2018, Chironomidae abundance declined significantly and composition varied relative to 2017; (3) in 2019, there were shifts in relative abundance of some taxonomic groups, but Chironomidae remained at low abundance relative to other taxa; and (4) in 2020, the dominance of *Hydra* across most sites meant that relative abundances of EPT taxa and Chironomidae were once again similar, and sites reflected 2017 patterns. Although BMI

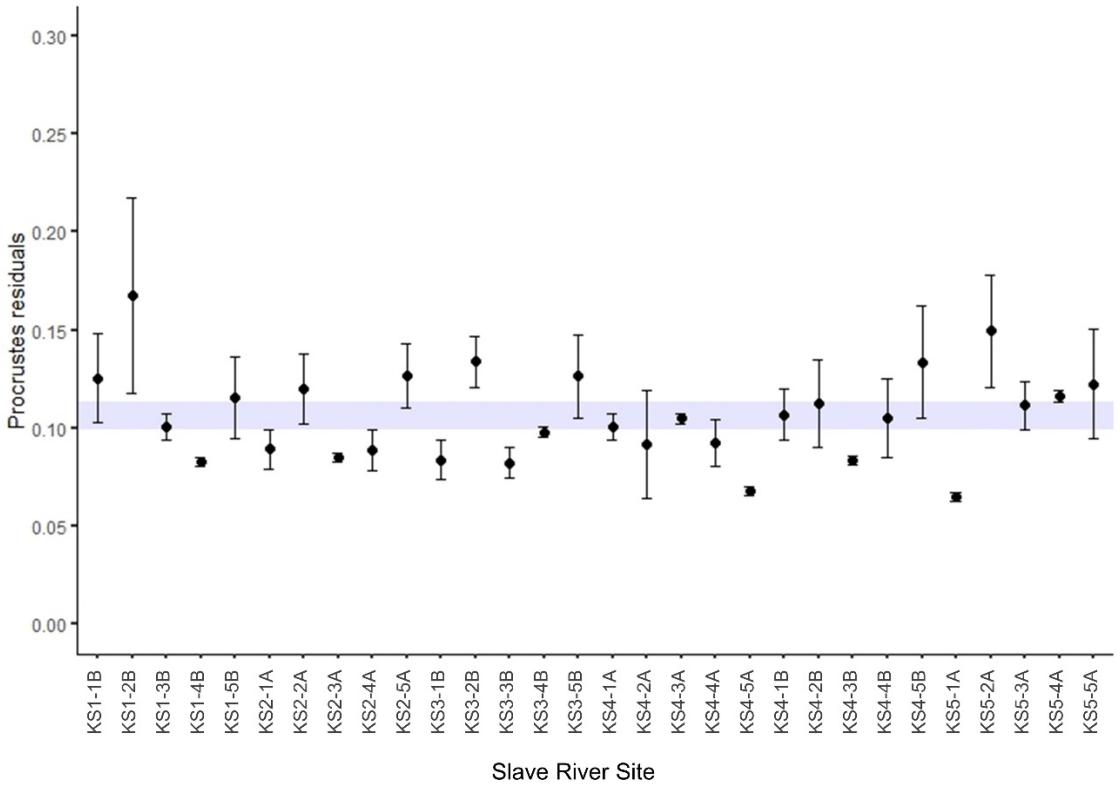

Figure 37 Multivariate normal range and CES boundaries for the Slave River, with the mean ± SE Procrustes residual (2017-2019) plotted for each site and the grand mean (mean of means for each pairwise year comparison) ± 2SD Procrustes residual indicated by the grey shaded area. Only sites sampled in all three years are included.

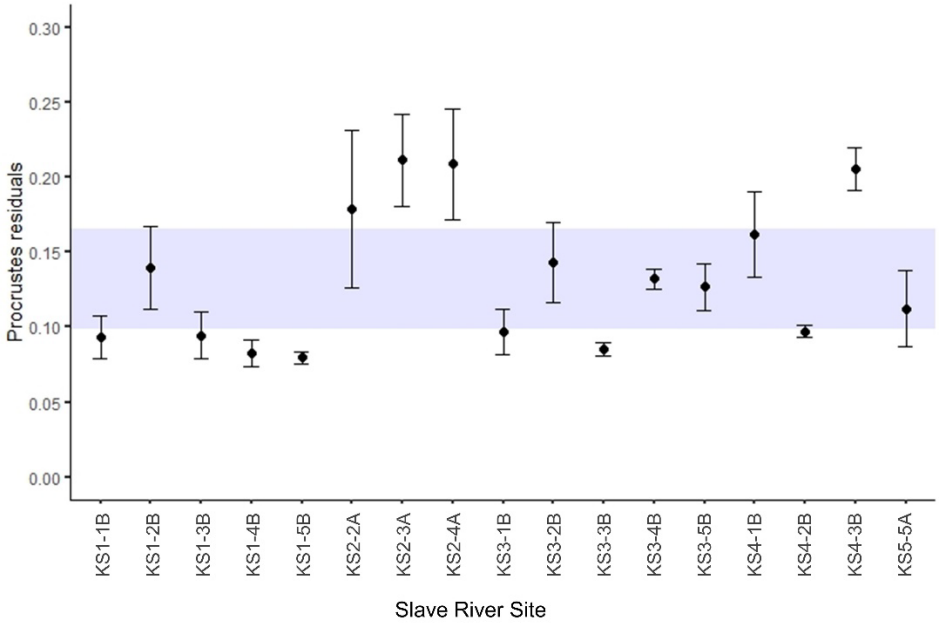

Figure 38 Multivariate normal range and CES boundaries for the Slave River, with the mean ± SE Procrustes residual (2017-2020) plotted for each site and the grand mean (mean of means for each pairwise year comparison) ± 2SD Procrustes residual indicated by the grey shaded area. Only sites sampled in all four years are included.

composition differed between 2017 and 2020, relative differences in composition among sites were more similar between these two years.

An initial estimate of a multivariate normal range and CES boundaries was created for the Slave River based on data from the period 2017-2019 (30 sites) and 2017-2020 (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 2017-2019, when more sites were sampled, the normal range was narrow and similar to that observed for the Hay River (Figure 37). However, in contrast to the Hay River, the mean residuals for few sites fell within the normal range, and error bars were wide for most sites. The means for 7 of 30 sites were within the CES boundaries, whereas the means for 11 sites were above the upper CES (indicating greater change across years than expected based on the normal range) and 12 sites were below the lower CES (indicating less change across years than expected; Figure 37). The high variability at the site scale is not surprising, given the strong compositional changes that were observed across years. The most variable sample was KS1-2B, and Reach 4A and Reach 4B were generally the least variable, with nearly all site means within normal range or below the lower CES (Figure 37).

Assessment of initial normal range for 2017-2020 was limited to only 17 sites that were sampled across all four years. The normal range for this subset of sites over all four years of sampling was much wider (Figure 38), which indicated greater temporal variability in the spatial arrangement of sites in multivariate space in the reduced ordinations. Sites in Reach 2 had much higher mean residuals and were more variable when 2020 was included, whereas sites in Reach 1 had slightly lower residuals. This may have reflected the relative stability of BMI assemblages in Reach 1 in 2020 (where *Hydra* did not dominate) and the large shift in composition in Reach 2 (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. Other sites sampled in 2020 were within the normal range or below the lower CES, with the exception of site KS4-3B, which had a high mean residual value that was above the upper CES (Figure 38). 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, when nearly twice as many sites were included in the ordinations. However, if the full set of sites in the Slave River can be sampled in future years, it will be possible to revisit the analyses and better refine estimates of the normal range of Procrustes residuals.

3.4. Ecological Response to Flow Conditions

3.4.1. Hay River

3.4.1.1. Characterizing flow conditions

The long-term hydrograph for the Hay River showed a history of periodic small floods and large floods, with the periods of highest flow (large floods, exceeding the peak flows of 2020) occurring prior to 1990 (Figure 39A). In the context of these earlier extreme high flows (discharge greater than 1000 m^3/s), the discharge in 2020 was categorized as a small flood, similar to other high-flow events in the last 30 years. However, the high flow pulses in between flood events in the first half of the record notably peaked at

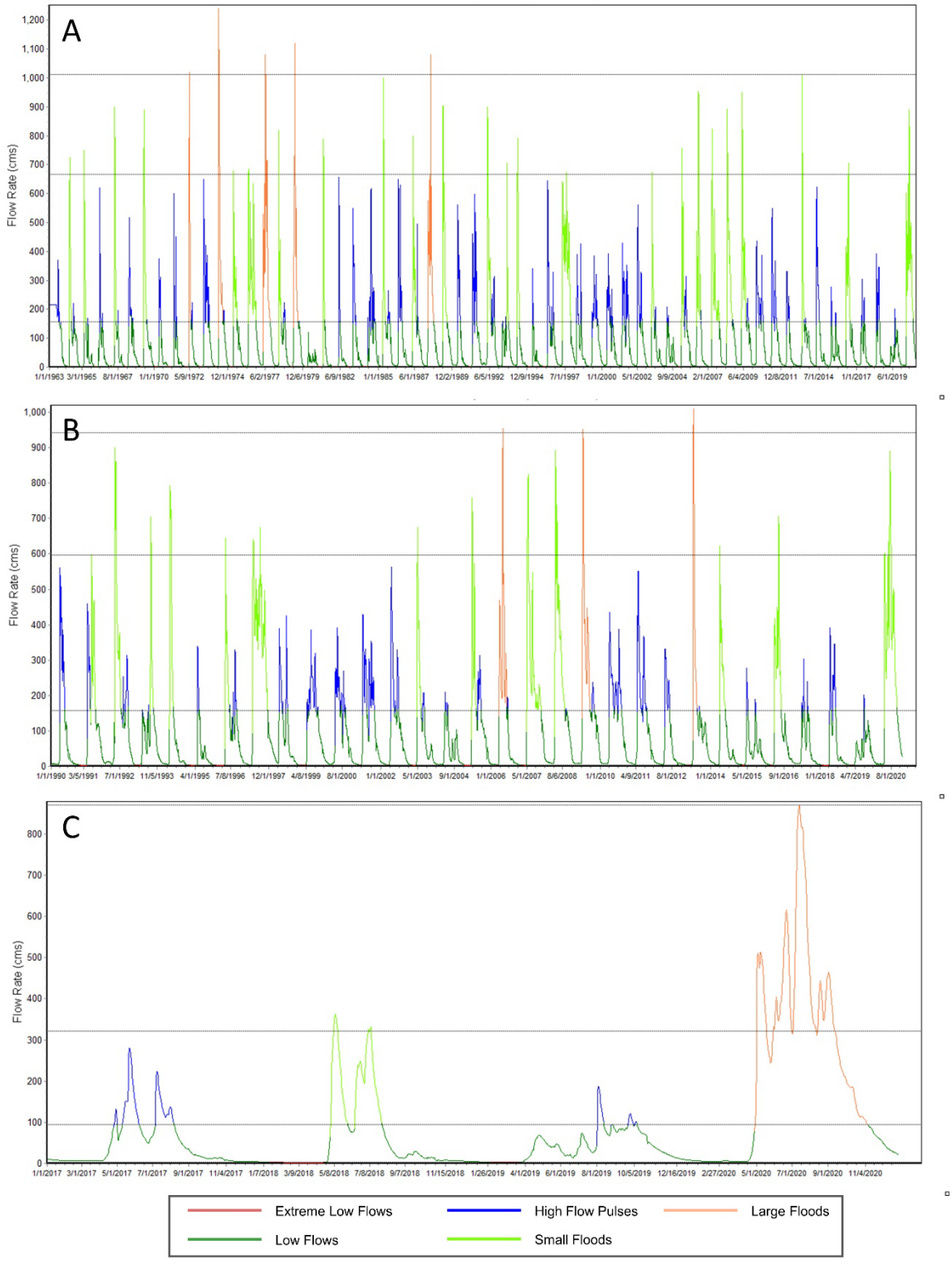

Figure 39 Environmental flow components of the Hay River hydrograph, calculated and plotted over three time periods: (A) 1963-2020, (B) 1990-2020, and (C) 2017-2020. Colours indicate periods of extreme low flows, low flows, high flow pulses, small floods, and large floods (see methods for definitions). Data in (A) and (B) from station 07OB001 (Hay River near Hay River) and data in (C) from station 07OB008 (Hay River near Alta/NWT boundary); all data from wateroffice.ec.gc.ca.

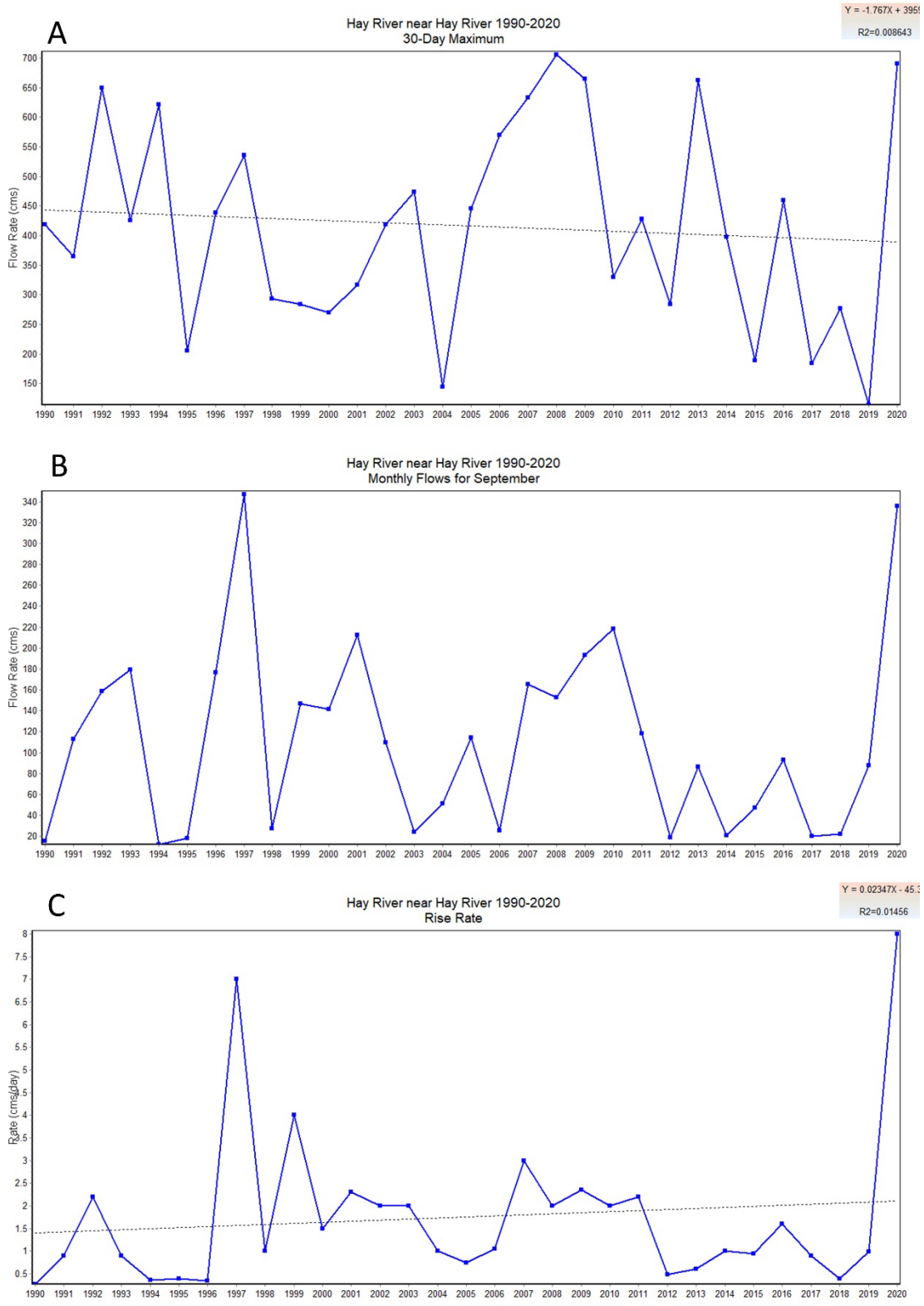

Figure 40 Hydro-ecological variables plotted for the Hay River for the period 1990-2020, including (A) 30-day maximum flows, (B) median monthly flows for September, and (C) rise rate. Dashed line is a non-parametric trend line fit to the data. Data for station 07OB001 (Hay River near Hay River) from wateroffice.ec.gc.ca.

higher maximum discharge than was observed in 2017-2019. The small flood event in 2020 represented a greater shift in flow conditions because peak flows were relatively low in the years prior. This pattern was also evident when the hydrograph from 1990-2020 was examined (Figure 39B). The change in peak flows from 2019 to 2020 was greater than that observed in most previous years, with the exception of the large flood in 2013. Furthermore, peak flows in 2017 and 2018 were relatively low compared to previous years (Figure 39B) and the spring freshets in 2018 and 2019 were preceded by extreme low flows (Figure 39B, C). Therefore, although the high flows in 2020 may have reached peak levels similar to or lower than previous years, this period of high flow was preceded by years where flows reached extreme lows and peaked at a lower discharge level.

The 30-day maximum for the period 1990-2020 (maximum flow from 30-day moving averages taken for every possible period of the year), indicated low maximum flow levels in 2019 and a sharp increase in 2020, representing the largest difference in 30-day maximum flows over the 30-year period (Figure 40A). Furthermore, the period 2017-2019 represented a prolonged period of low 30-day maximum flows relative to previous years. When median monthly flows for September were plotted (relevant to the time of sampling for the Hay River), flows were generally low from 2012 to 2019, while 2020 represented the second highest median September flows over the 30-year record (Figure 40B). The increase in flows was also faster than in previous years, as the rise rate plotted from 1990-2020 indicated that 2020 had the fastest rise rate in the 30-year record (Figure 40C).

Hydrologic characterization of the Hay River emphasized that although the high flows observed in 2020 were similar to levels achieved in past years, the high rate and magnitude of increase, following a period of lower flows, appeared to be unusual. However, given that sampling was not completed in 2020, this does not address the question of whether data from 2017-2019 can be separated based on flow. Interestingly, though water levels at the time of sampling in 2019 were extremely high, peak flows in 2019 were lower than in the previous two years, and the high water levels were the result of a delayed discharge peak in 2019 (Figure 39C), emphasizing the importance of the timing of peak flows. Discharge was most similar between 2017 and 2018, though 2018 had higher peak discharge and there was a period of extreme low flows prior to the spring freshet in 2018. Development of flow-specific normal range and CES could therefore include 2017 and 2018 as typical non-flood years. The inclusion of 2019 may not be advisable, given the delayed timing of peak flows and its likely effect on assemblage composition.

3.4.1.2. Flow-ecology relationships

The TITAN2 analysis of Hay River BMI in relation to velocity indicated a community response threshold that separated *z*- responders and *z*+ responders (Figure 41). Out of 41 possible taxa, there were 6 taxa identified as pure and reliable negative responders and 9 taxa identified as pure and reliable positive responders. Both *z*- responders and *z*+ responders were distributed across a wide range of velocities, and there was some overlap of their density probability plots, but community response peaks were clearly different (Figure 41). The community change threshold for *z*- responders was a velocity of 0.127 m/s (5^{th} quantile = 0.080, 95^{th} quantile = 0.163; Figure 41) and the community change threshold for *z*+ responders was 0.310 m/s (5^{th} quantile = 0.194, 95^{th} quantile = 0.443; Figure 41).

The strongest *z*- responder was the mayfly Metretopodidae (cleftfooted minnow mayfly), which is known to inhabit slow-water habitats and river banks of large rivers during high water years

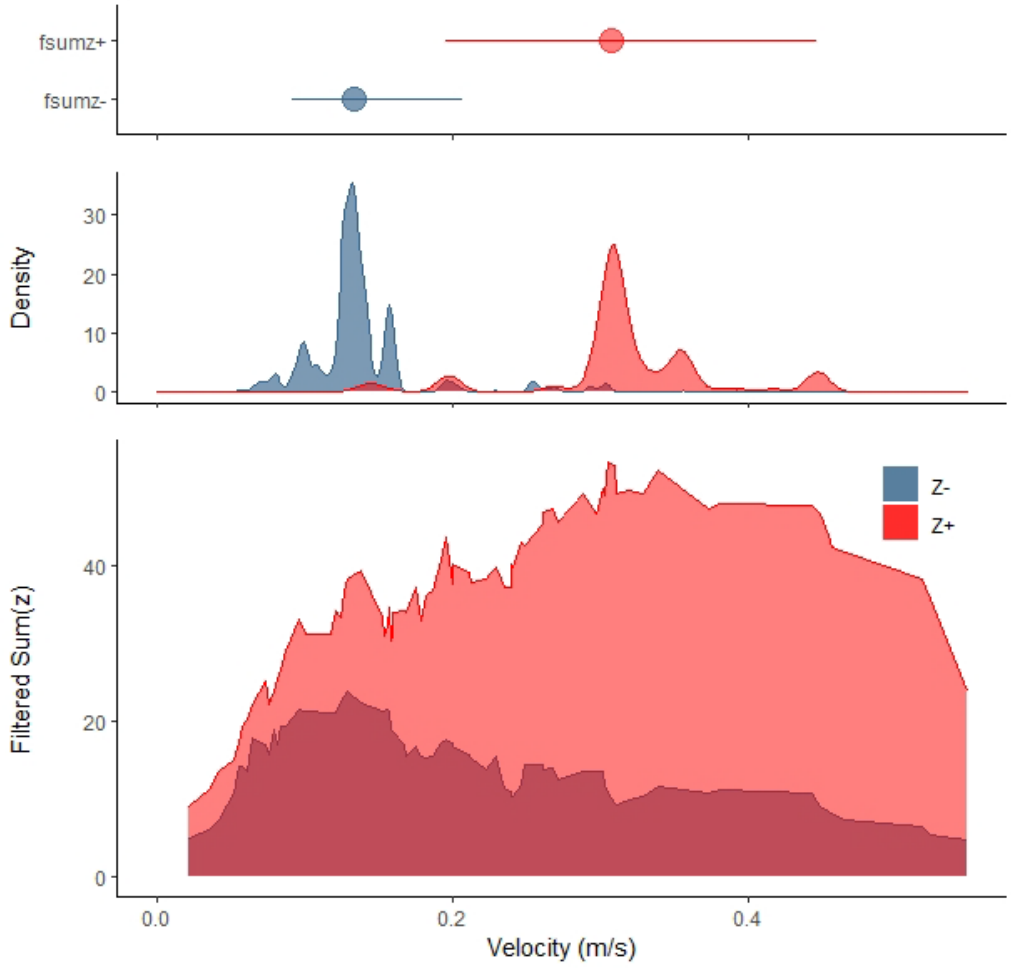

Figure 41 Community response plots from Threshold Indicator Taxa ANalysis (TITAN2) for the Hay River, including (top) the filtered sum of negative responders (z-) and positive responders (z+), with the maxima along the velocity gradient plotted as circles with 95th percentiles as horizontal lines; (middle) taxa change points plotted as a probability density function; and (bottom) the magnitude of change in negative responder taxa and positive responder taxa along the velocity gradient. Taxa that are more abundant at higher velocities are z+ responders and plotted in red. Taxa that are more abundant at lower velocities are z- responders and plotted in blue.

(McCafferty 1998). Other *z-* responders included Corixidae, which are known to migrate between wetlands and large rivers in large numbers (Srayko et al. 2022), freshwater clams (Pisidiidae) and the mayfly Leptophlebiidae, both of which tend to be associated with slow-flowing water and pools (McCafferty 1998), and two Chironomidae subfamilies (Figure 42). In contrast, *z+* responders included several taxa that are known to be associated with higher flows, including black flies (Simuliidae), the stonefly Perlodidae, and the filter-feeding caddisfly Hydropsychidae (Figure 42). Other *z+* responders included the family Oceanidae, which is a family of hydra that was not observed in the Slave River, segmented worms (Enchytraeidae) and several true fly taxa, including crane flies, dance flies, and a Chironomidae subfamily. Multi-modal organism responses for *z-* or *z+* responders reflected a degree of variability in the flow conditions in which those taxa were found (Figure 42).

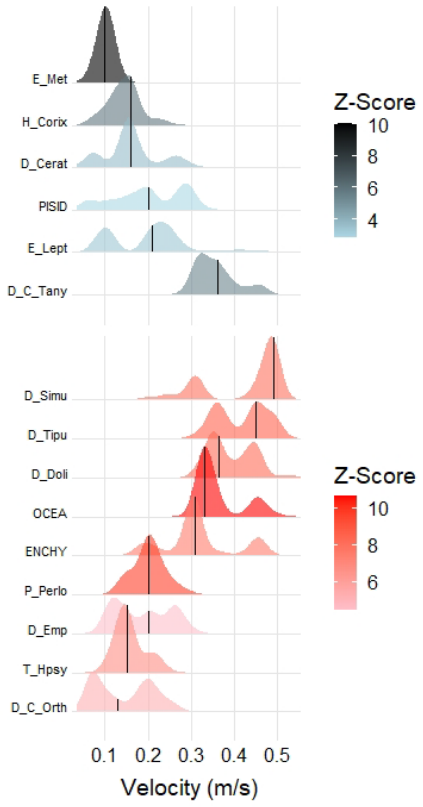

Figure 42 Individual response plots from Threshold Indicator Taxa ANalysis (TITAN2) for Hay River taxa with a pure and reliable response to the velocity gradient. Taxa that were negative responders along the gradient are indicated in blue, whereas positive responders are shown in red. Taxa change points are plotted as probability density functions, with shading representing the z-score, and thus magnitude of response.

The results of the TITAN2 analysis for the Hay River provide useful information about the velocity at which there is a change in the assemblage (designated by the community change thresholds for *z*- and *z*+ responders), and there are particular taxa associated with each end of the velocity gradient in Hay River samples. With additional sampling under different flow conditions, it may be possible to begin to more accurately characterize the BMI assemblages of the river in response to flow conditions based on the relative dominance of *z*- and *z*+ responders in a given year. Furthermore, the relative abundance of some key *z*- or *z*+ responders could be considered as additional metrics with which that status of the BMI assemblage can be assessed over time.

3.4.2. Slave River

3.4.2.1. Characterizing flow conditions

Similar to the Hay River, the hydrograph for the full record of Slave River data (1960-2020) indicated more variable flows with higher peak discharge in the earlier part of the data record, and the high flows in 2020 were characterized as a small flood in the context of the larger record (Figure 43A). In particular, the hydrograph showed lower peak discharge and less extreme low flows following the construction of the WAC Bennett Dam on the Peace River (completed in 1967) and the beginning of operations (1972). The dam is known to have caused changes to flows in the Peace River, with less extreme peak flows in

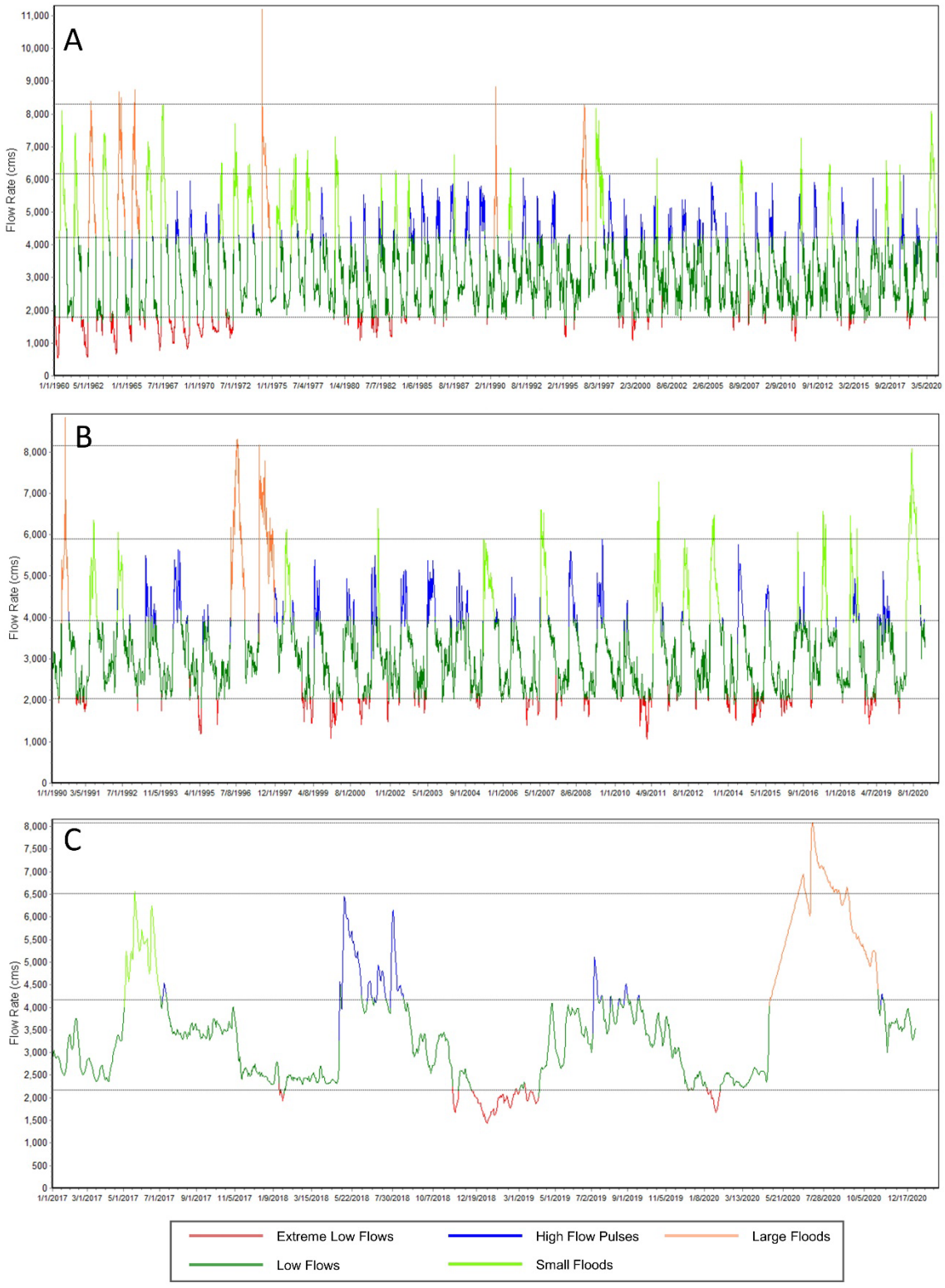

Figure 43 Environmental flow components of the Slave River hydrograph, calculated and plotted over three time periods: (A) 1963-2020, (B) 1990-2020, and (C) 2017-2020. Colours indicate periods of extreme low flows, low flows, high flow pulses, small floods, and large floods (see methods for definitions). Data from station 07NB001 (Slave River at Ft. Fitzgerald); all data from wateroffice.ec.gc.ca.

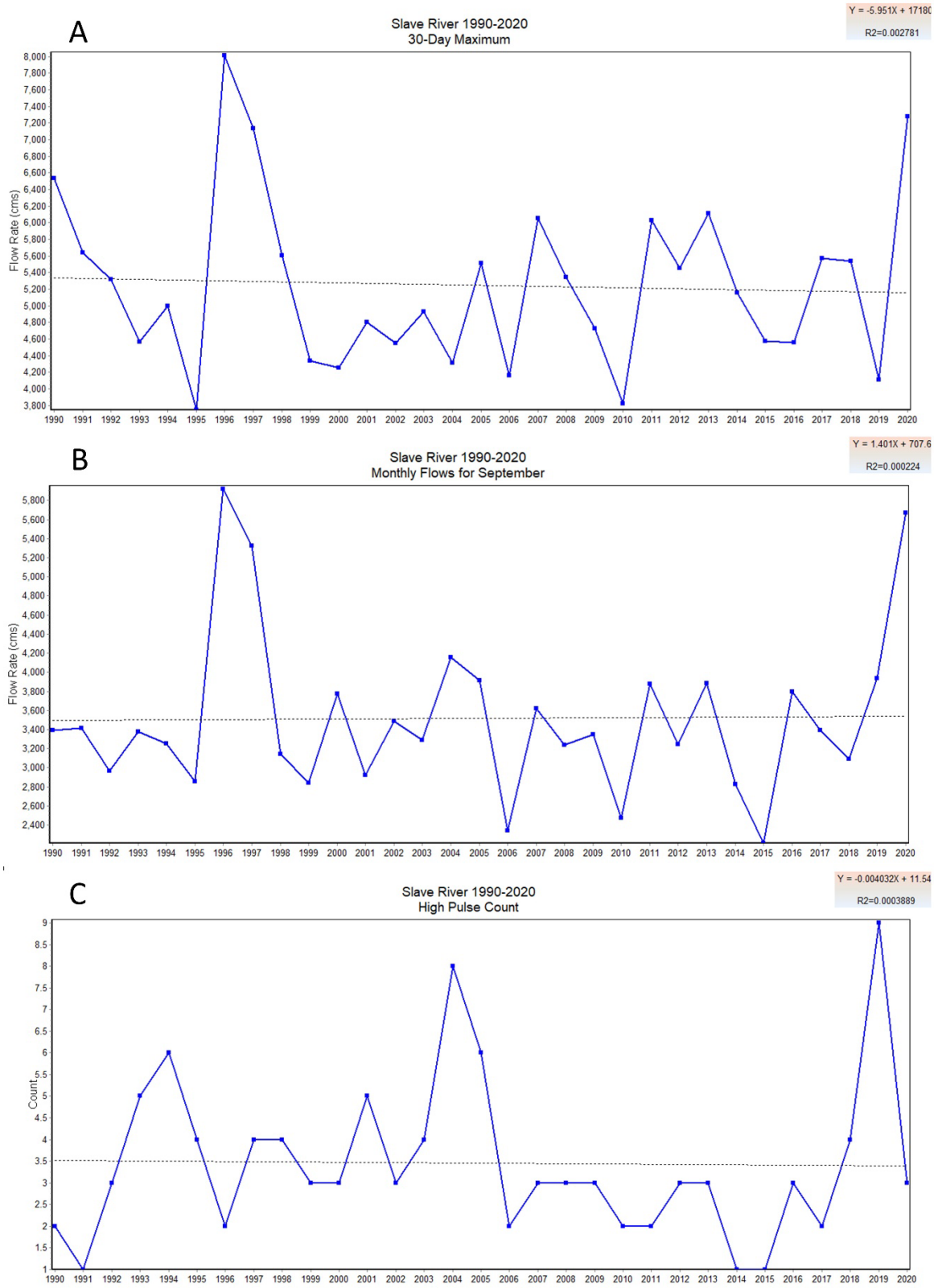

Figure 44 Hydro-ecological variables plotted for the Slave River for the period 1990-2020, including (A) 30-day maximum flows, (B) median monthly flows for September, and (C) high pulse count. Dashed line is a non-parametric trend line fit to the data. Data for station 07NB001 (Slave River at Ft Fitzgerald) from wateroffice.ec.gc.ca.

the spring freshet and higher low flows, a pattern that is consistent with the flattening of the hydrograph that is typical of regulated rivers (Peters et al. 2014). Such changes to the flow regime reduce the seasonality of the hydrograph, with potential impacts on timing of ecological processes (Bunn and Arthington 2002, Peters et al. 2014). Though less extreme in the Slave River than in the upstream Peace River basin, the change to the Slave River hydrograph is apparent (Figure 43A).

When viewed in the context of the hydrograph from 1990-2020, the high flow event in 2020 was among the highest peaks in discharge, just below the level of a large flood (Figure 43B). Peak discharge in 2020 was the highest observed since 1997 in the Slave River, though discharge was also high in 2011 (Figure 43B). Peak flows in 2017 and 2018 were also classified as small floods, though the peak had a smaller magnitude. In contrast, the peak in 2019 was classified as a high flow pulse as a result of the lower peak discharge in that year (Figure 43B).

Considering only the four years of sampling, the hydrograph was similar to that of the Hay River, with similar peaks in 2017 and 2018, a lower and flatter peak in 2019, and a sharp increase in 2020 (Figure 43C). The higher variability of flows in 2018 was apparent due to the high rise and fall rate of reversals during the peak flow period, but the hydrograph for 2019 also appeared quite variable, with a number of pulses throughout the peak period (Figure 43C).

Hydro-ecological variables generally confirmed the patterns that were evident in the hydrographs. From the early 2000s, the 30-day maximum flow for the Slave River fluctuated, but was fairly steady. But in 2020 it increased to the highest value since 1997 (Figure 44A). Median monthly flows for September similarly showed steady fluctuations from the early 2000s followed in 2020 by the largest median monthly flow since 1997 (Figure 44A). While peak flow was low in 2019, Figure Figure 44C shows the large number of high pulse counts recorded that year, further emphasizing the variability of flows in 2019.

Overall, the hydrographs and hydro-ecological variables indicated similar peak flows in the Slave River in 2017-2018, variable flows in both 2018 and 2019, and low peak flows in 2019 followed by high flows in 2020. Though the magnitude of peak flows in 2020 was not unusual for the long-term hydrograph, such high flows have not been observed in the river for over two decades.

Though this assessment supports the conclusion that flows have been variable across the period of sampling, it does not provide insight into the strong changes in assemblage composition observed between 2017 and 2018 in Slave River samples. The hydrograph did differ somewhat in 2017, with a plateau in the fall rather than the typical decline in flow, but peak flow was similar between 2017 and 2018, and flows were fairly regular in the river to that point. Variability in flow in 2018 may have contributed to the shift in assemblage composition, and the hydrograph in 2019 was similarly variable, but more years of data would be required to assess whether flow variability contributed to the change. Partitioning data based on flow magnitude alone would suggest that 2017 and 2018 should be used to create a normal range and CES for typical flow conditions in the river, but these two years differed greatly in assemblage composition. It is possible, therefore, that flow-based CES will not be possible for this river without additional data and additional examples of assemblages that are typical for low and high flow years.

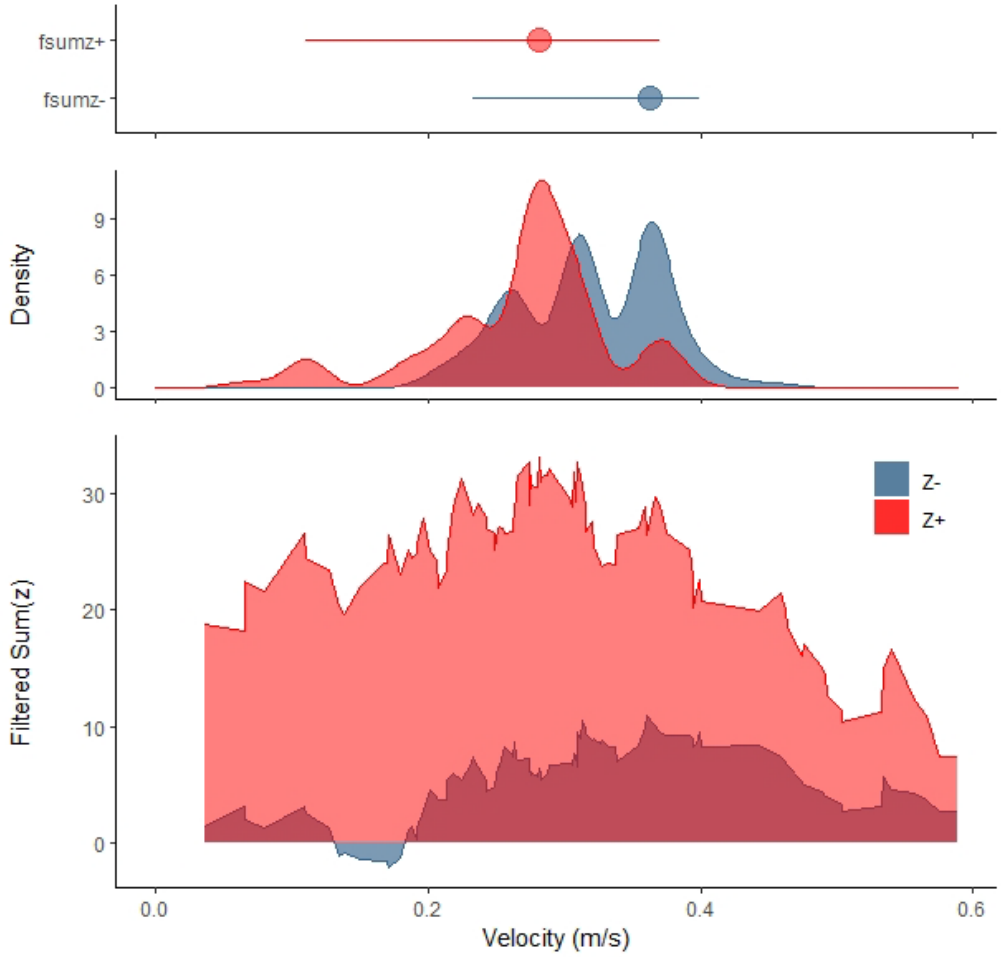

Figure 45 Community response plots from Threshold Indicator Taxa ANalysis (TITAN2) for the Slave River, including (top) the filtered sum of negative responders (z-) and positive responders (z+), with the maxima along the velocity gradient plotted as circles with 95th percentiles as horizontal lines; (middle) taxa change points plotted as a probability density function; and (bottom) the magnitude of change in negative responder taxa and positive responder taxa along the velocity gradient. Taxa that are more abundant at higher velocities are z+ responders and plotted in red. Taxa that are more abundant at lower velocities are z- responders and plotted in blue.

3.4.2.2. Flow-ecology relationships

The TITAN2 analysis of Slave River BMI in relation to velocity did not indicate a strong community threshold response to velocity (Figure 45). There was a strong overlap of *z*- responders and *z*+ responders, and the maxima were reversed (higher velocity maxima for *z*- responders) because of the high degree of overlap in the probability density functions (Figure 45). There was somewhat greater separation between the two groups when data from 2020 were excluded, but the lack of distinction along the gradient remained. Out of 38 possible taxa, there were 4 taxa identified as pure and reliable negative responders and 9 taxa identified as pure and reliable positive responders. The community change threshold for *z*- responders was a velocity of 0.363 m/s (5^{th} quantile = 0.262, 95^{th} quantile = 0.538;) and the community change threshold for *z*+ responders was 0.282 m/s (5^{th} quantile = 0.099, 95^{th}

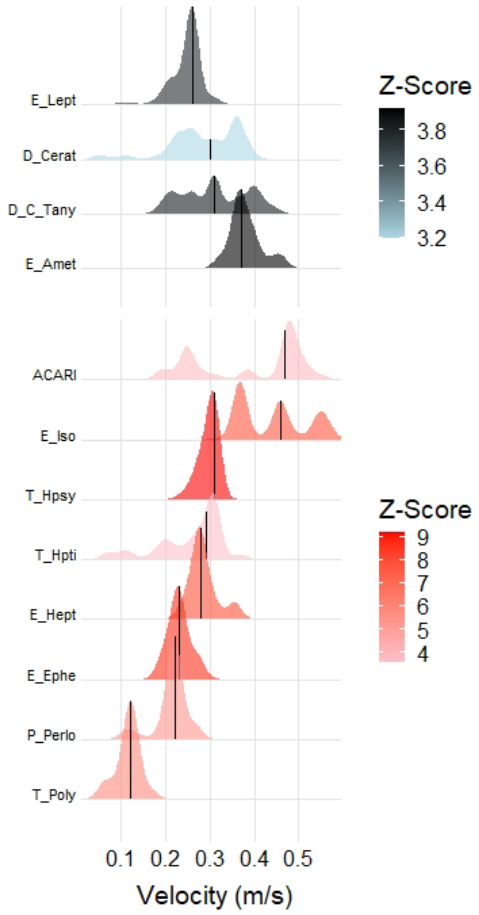

Figure 46 Individual response plots from Threshold Indicator Taxa ANalysis (TITAN2) for Slave River taxa with a pure and reliable response to the velocity gradient. Taxa that were negative responders along the gradient are indicated in blue, whereas positive responders are shown in red. Taxa change points are plotted as probability density functions, with shading representing the z-score, and thus magnitude of response.

quantile = 0.370; Figure 45). The strong overlap and higher *z*- changepoint indicated that the results of this analysis were not meaningful.

The *z*- responders had low individual *z*-scores, which likely contributed to the lack of distinction along the velocity gradient (Figure 46). The *z*- responders included a Chironomidae subfamily, biting midges, and segmented worms (Naididae). The *z*+ responders included some taxa typical of faster flow conditions, such as the caddisfly Hydropsychidae, the mayflies Isonychiidae, Heptageniidae, and Ephemerellidae, the latter of which is often found associated with areas with wave action (McCafferty 1998), and the stonefly Perlodidae (Figure 46).

Overall, the results of the TITAN2 analysis were not useful for the Slave River, as it was not possible to derive community thresholds for velocity using the samples collected from 2017 to 2020. In part, this may indicate that spot velocity measurements were not a reliable proxy for variability in discharge between years, as no clear patterns emerged. However, as noted earlier, the large change in composition from 2017 to 2018 does not appear to be associated with changes in flow conditions, which

suggests that variability in assemblage structure may have responded to a different factor. Additional data will be needed in order to determine whether future changes can be associated with flow or a different environmental driver.

4. Recommendations and Conclusions

2020 represented an extreme year for flows in both the Hay River and the Slave River, leading to reduced sampling in the Slave River and no sampling in the Hay River. In the context of long-term flow patterns, peak flow magnitude in 2020 was not unusual for either river, but it followed a period of lower flows in the Hay River, and it was the highest peak in 20 years in the Slave River. The effects on Hay River BMI assemblages are as yet unknown, but Slave River assemblages in 2020 were dominated by *Hydra* in most reaches, with abundance of *Hydra* in some reaches leading to highly elevated total abundance values compared to previous years. *Hydra* are not well studied, but they are highly adaptable, which may have contributed to their ability to thrive in the high flow conditions of 2020. Patterns in *Hydra* abundance and concurrent declines in relative abundance and richness of other taxa dominated the benthic assessment of the Slave River in 2020.

Despite the strong influence of *Hydra* on the 2020 data, there was some evidence of consistent patterns across the 2017-2020 period. Of the metrics tested for the Slave River, Chironomidae abundance and EPT richness appeared to have the greatest initial potential for developing monitoring and management triggers. Both metrics were relatively invariable across years (Chironomidae more so if 2017 was excluded), with a narrow normal range and CES boundaries. Abundance metrics (e.g., EPT abundance, Chironomidae abundance) and richness metrics generally appeared to be more effective for developing normal range criteria than relative abundance metrics, as the latter were highly influenced by *Hydra* in 2020.

Some of the variability in normal range estimates is due to inter-annual differences in assemblages, which could result from differences in flow conditions, temperature, or other environmental drivers from one year to the next, and this variability can be accounted for by developing different normal range criteria for different environmental conditions. The multi-year PCA with 95% normal probability ellipses provides a potential approach to identify the years that should be grouped together in the development of more precise biotic metric normal range and CES. For example, the PCA for Hay River sites indicated a strong similarity of 2017 and 2018, and indicated that composition differed in 2019. This suggests that the first two years of sampling could be grouped to create CES, whereas 2019 could be considered separately, once data are collected under similar conditions. The PCA for the Slave River indicated that 2018-2020 were most similar, while composition differed in 2017. While grouping 2018-2020 and excluding 2017 might not lead to more narrow normal range estimates for all metrics (for example, those that were affected by high abundance of *Hydra* in 2020), there was a strong similarity among data from 2018-2020 for a selection of biotic metrics that included Chironomidae and Diptera + Oligochaeta abundance, relative abundance, and richness, as well as other taxonomic richness metrics. Earlier assessment of temporal patterns in the Slave River focused on the decline in Chironomidae abundance and diversity between 2017 and 2018. However, there have now been three years with low abundance of Chironomidae in the Slave River. Furthermore, the change in composition could not be linked to changing flow conditions with the data available (river-scale discharge data, rather than site- or reach-scale). It is possible that normal range estimates for biotic metrics would be more precise if data

from 2017 were excluded, given that abundance and diversity of Chironomidae have not rebounded to previous levels. If higher abundance and diversity of Chironomidae are observed in future sampling years, it would be logical to compare those data with 2017 data to develop normal range criteria; however, it is unknown whether such a shift will occur, and it will be necessary to try to discern the cause if such a shift does take place. The relationship between multivariate probability ellipses and biotic metric CES should continue to be examined as more data are collected, as the utility and composition of such groupings will likely evolve as more years of data are added.

Procrustes residuals should also continue to be explored as more data are collected. With additional years of data, it will be possible to continue to assess the degree of temporal stability in the composition of sites relative to one another, and potentially identify any sites that change composition substantially relative to previous years. This provides an additional tool to detect possible impairment based on the full assemblage, rather than a particular biotic metric.

Hydrologic characterization of the Hay River and Slave River emphasized that although the high flows observed in 2020 were similar to high discharge events in past years, other conditions related to the timing and magnitude of flows prior to 2020 contributed to the extreme conditions. For example, the high rate and magnitude of increase in discharge in the Hay River, which followed a period of lower flows in 2019, appeared to be unusual. Development of flow-specific normal range and CES for the Hay River could therefore include 2017 and 2018 as typical non-flood years (consistent with the grouping suggested by the multivariate analysis). The inclusion of data from 2019 may not be advisable, given the delayed timing of peak flows and its likely effect on assemblage composition. For the Slave River, 2020 represented the highest peak flow since the late 1990s. But the hydrograph did not highlight strong differences between 2017 and 2018 that might have contributed to differences in BMI assemblage composition. Furthermore, partitioning data based on flow magnitude alone would suggest that 2017 and 2018 should be used to create a normal range and CES for typical conditions in the river, despite the fact that the BMI assemblage in 2017 appeared to differ from all other sampling years. Assessment of flow-ecology relationships also failed to identify strong associations with velocity for the Slave River, though this may simply indicate that site-scale spot velocity measurements are a poor proxy for inter-annual variability in discharge. It is possible, therefore, that flow-based CES will not be possible for the Slave River without additional data and additional examples of assemblages that are typical for low and high flow years.

Though data in 2020 were limited, they offered the opportunity to examine compositional patterns more closely in the Slave River, and to explore additional tools to characterize spatial and temporal variability in sample reaches. With more data, these tools and approaches can be strengthened.

5. References

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.

Angradi, T.R., Schweiger, E.W., and Bolgrien, D.W. 2006. Inter-habitat variation in the benthos of the Upper Missouri River (North Dakota, USA): implications for Great River bioassessment. River Res. Applic. **22**(7): 755-773. doi: 10.1002/rra.932.

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.

Baker, M.E., and King, R.S. 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods in Ecology and Evolution **1**(1): 25-37. doi: 10.1111/j.2041-210x.2009.00007.x.

Baker, M.E., King, R.S., and Kahle, D. 2020. TITAN2: Threshold Indicator Taxa Analysis. (R package version 2.4.1). Available from https://CRAN.R-project.org/package=TITAN2.

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.

Di Camillo, C.G., Bavestrello, G., Cerrano, C., Gravili, C., Piraino, S., Puce, S., and Boero, F. 2017. Hydroids (Cnidaria, Hydrozoa): A Neglected Component of Animal Forests. Springer International Publishing. pp. 397-427.

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.

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.

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.

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., 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., 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. 2022. 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.

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).

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.

McCafferty, W.P. 1998. Aquatic Entomology: The Fisherman's and Ecologist's Illustrated Guide to Insects and Their Relatives. Jones & Bartlett Publishers.

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.

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.

Murphy, C.A., Johnson, S.L., Gerth, W., Pierce, T., and Taylor, G. 2021. Unintended Consequences of Selective Water Withdrawals From Reservoirs Alter Downstream Macroinvertebrate Communities. Water Resour. Res. **57**(6). doi: 10.1029/2020wr029169.

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.

Paradis, E., and Schliep, K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics **35**: 526-528.

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 2021. 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.

Srayko, S.H., Jardine, T.D., Phillips, I.D., and Chivers, D.P. 2022. Seasonal Mass Migration of Water Boatmen (Hemiptera: Corixidae) as a Wetland–River Linkage and Dietary Subsidy to Riverine Fish. Ecosystems. doi: 10.1007/s10021-021-00734-5.

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.

Tang, Y., Horikoshi, M., and Li, W. 2016. ggfortify: Unified interface to visualize statistical results of popular R packages. The R Journal **8**(2): 475-485. doi: 10.32614/RJ-2016-060.

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.

The Nature Conservancy. 2009. Indicators of Hydrologic Alteration Version 7.1 User's Manual.

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 10 Names and coordinates of all kick-sampling sites in the Hay River and Slave River. Sites sampled in 2020 are indicated with a sampling date. Two sets of coordinates differed from 2019, as indicated in the notes column.

River	Reach	Site	Latitude	Longitude	Date
HAY RIVER	REACH 1	HR-KS1-1A	59.93403	-116.95028	
		HR-KS1-2A	59.93591	-116.95175	
		HR-KS1-3A	59.93211	-116.95237	
		HR-KS1-4A	59.93135	-116.95506	
		HR-KS1-5A	59.93124	-116.95613	
	REACH 2	HR-KS2-1A	59.94548	-116.95565	
		HR-KS2-2A	59.94617	-116.95618	
		HR-KS2-3A	59.94654	-116.95647	
		HR-KS2-4A	59.94703	-116.95702	
		HR-KS2-5A	59.94759	-116.95744	
	REACH 3	HR-KS3-1A	59.98767	-116.93236	
		HR-KS3-2A	59.98827	-116.93060	
		HR-KS3-3A	59.98845	-116.93037	
		HR-KS3-4A	59.99023	-116.93049	
		HR-KS3-5A	59.99182	-116.93127	
	REACH 4	HR-KS4-1A	60.00158	-116.97036	
		HR-KS4-2A	60.00205	-116.97145	
		HR-KS4-3A	60.00261	-116.97126	
		HR-KS4-4A	60.00308	-116.97089	
		HR-KS4-5A	60.00319	-116.97009	
	REACH 5	HR-KS5-1B	60.01064	-116.92032	
		HR-KS5-2B	60.01096	-116.92088	
		HR-KS5-3B	60.01125	-116.92177	
		HR-KS5-4B	60.01138	-116.92274	
		HR-KS5-5B	60.01163	-116.92348	
	REACH 6	HR-KS6-1B	60.02772	-116.92342	
		HR-KS6-2B	60.02779	-116.92217	
		HR-KS6-3B	60.02785	-116.92155	
		HR-KS6-4B	60.02787	-116.92075	
		HR-KS6-5B	60.02802	-116.91985	

River	Reach	Site	Latitude	Longitude	Date	Notes
SLAVE RIVER	REACH 1	SR-KS1-1B	59.40805	-111.46321	2020-10-05	
		SR-KS1-2B	59.40805	-111.46321	2020-10-05	
		SR-KS1-3B	59.40846	-111.46196	2020-10-05	
		SR-KS1-4B	59.40879	-111.46082	2020-10-05	
		SR-KS1-5B	59.40913	-111.45985	2020-10-05	
	REACH 2	SR-KS2-1A	59.42689	-111.46155		
		SR-KS2-2A	59.42709	-111.46199	2020-10-05	
		SR-KS2-3A	59.42761	-111.46294	2020-10-05	
		SR-KS2-4A	59.42799	-111.46361	2020-10-05	
		SR-KS2-5A	59.42858	-111.46458		
	REACH 3	SR-KS3-1B	59.53395	-111.45934	2020-10-06	
		SR-KS3-2B	59.53372	-111.45978	2020-10-06	Moved upstream of 1B
		SR-KS3-3B	59.53502	-111.45774	2020-10-06	
		SR-KS3-4B	59.53538	-111.45703	2020-10-06	
		SR-KS3-5B	59.53562	-111.45651	2020-10-06	
	REACH 4A	SR-KS4-1A	59.58906	-111.41968		
		SR-KS4-2A	59.58947	-111.4196		
		SR-KS4-3A	59.59122	-111.41951		
		SR-KS4-4A	59.59178	-111.41949		
		SR-KS4-5A	59.59225	-111.41946		
	REACH 4B	SR-KS4-1B	59.58887	-111.42283	2020-10-07	
		SR-KS4-2B	59.58975	-111.42273	2020-10-07	
		SR-KS4-3B	59.58995	-111.42268	2020-10-07	
		SR-KS4-4B	59.5909	-111.42261		
		SR-KS4-5B	59.59139	-111.42264		
	REACH 6	SR-KS6-1B	60.02772	-116.92342		
		SR-KS6-2B	60.02779	-116.92217		
		SR-KS6-3B	60.02785	-116.92155		
		SR-KS6-4B	60.02787	-116.92075		
		SR-KS6-5B	60.02802	-116.91985	2020-10-07	
	REACH 5	SR-KS5-1A	59.71284	-111.50644		
		SR-KS5-2A	59.71304	-111.50646		
		SR-KS5-3A	59.71823	-111.50577		
		SR-KS5-4A	59.71853	-111.50594		
		SR-KS5-5A	59.67804	-111.48615	2020-10-07	Coordinate change; did not have 2019 coords

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

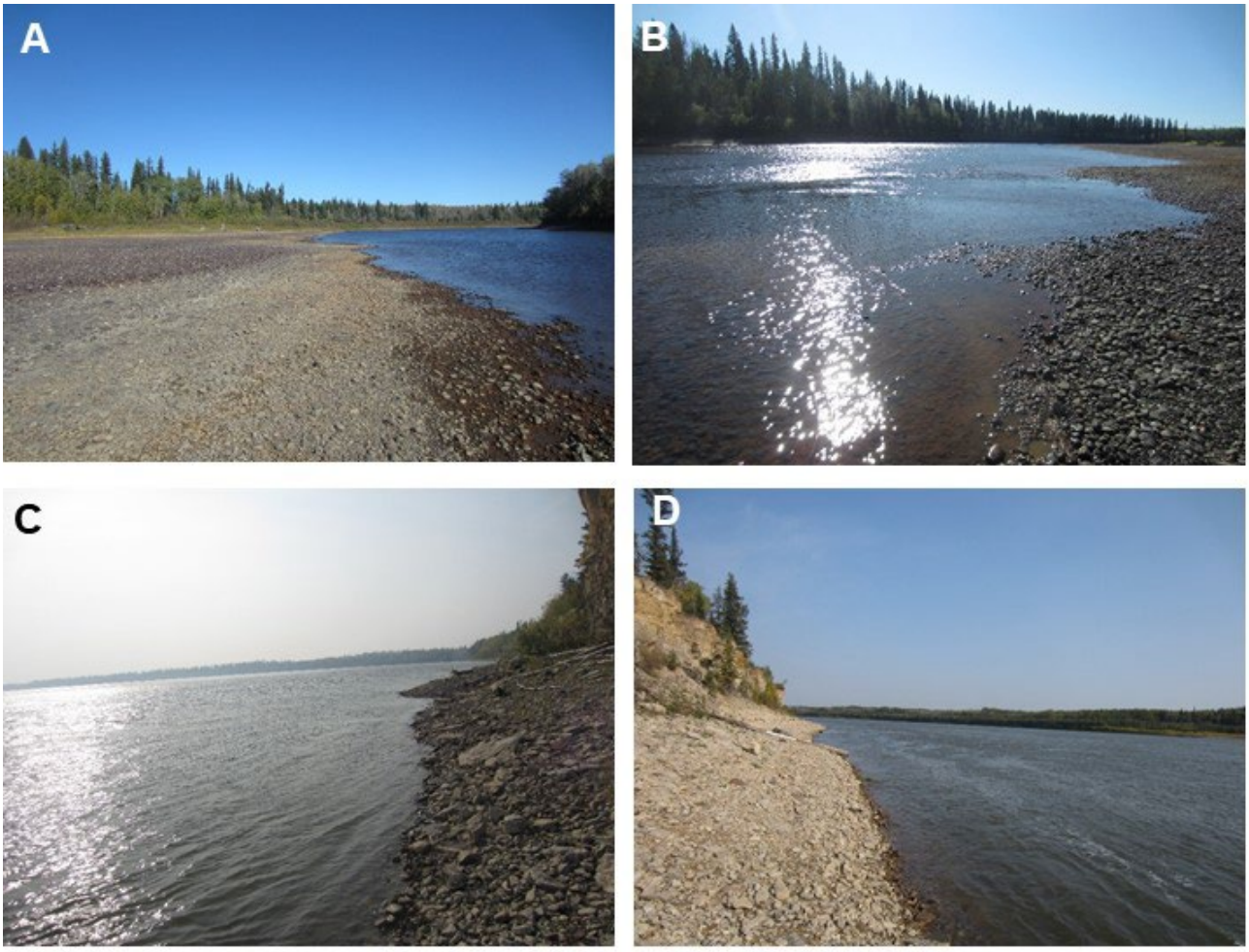

Figure 47. Pictures of sample locations, including (A) upstream view from Hay River Reach 1, (B) downstream view from Hay River Reach 1, (C) upstream view from Slave River Reach 1, and (D) downstream view from Slave River Reach 1. Photos taken in 2017.