Metric and Power Analysis for a Biomonitoring Program in Banff National Park, Alberta.

by

Jessica Penno

A Thesis Submitted to the Faculty of Social and Applied Sciences in Partial Fulfillment of the Requirements for the Degree of

Master Of Science

In

Environment And Management

Royal Roads University
Victoria, British Columbia, Canada

Supervisor: Dr. Mark Taylor
January, 2016
COMMITTEE APPROVAL

The members of Jessica Penno’s Thesis Committee certify that they have read the thesis titled Metric and Power Analysis for a Biomonitoring Program in Banff National Park, Alberta and recommend that it be accepted as fulfilling the thesis requirements for the Degree of Masters of Science in Environment and Management:

Dr. Audrey Dallimore [signature on file]

Final approval and acceptance of this thesis is contingent upon submission of the final copy of the thesis to Royal Roads University. The thesis supervisor confirms to have read this thesis and recommends that it be accepted as fulfilling the thesis requirements:

Dr. Mark Taylor [signature on file]
Creative Commons Statement

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 Canada License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/2.5/ca/.

Some material in this work is not being made available under the terms of this license:

- Third-Party material that is being used under fair dealing or with permission.
- Any photographs where individuals are easily identifiable.
ABSTRACT

The purpose of this thesis is to inform a biomonitoring study in Banff National Park, Alberta, that will use benthic macroinvertebrates as a biological indicator of aquatic ecosystem condition in response to the re-introduction of bison. To inform the biomonitoring program, macroinvertebrate metrics were evaluated for biological relevance to the study design. Existing data from a surrogate flooding disturbance in Banff National Park was used to evaluate the effects of different sampling scenarios on the power to detect effects from a bison reintroduction and prescribed burn program. Power analysis highlighted that four metrics would provide >80% chance to detect an effect within two to five years of the beginning of the monitoring program. The metrics were richness, true diversity, % shredders and % shredders all of which have both biological importance to the productivity of fishery resources and are easily communicated to the general public.

Keywords: disturbance, benthic macroinvertebrates, biomonitoring, power
# TABLE OF CONTENTS

ABSTRACT .................................................................................................................................4  

LIST OF TABLES..........................................................................................................................6  

LIST OF FIGURES............................................................................................................................7  

ACKNOWLEDGEMENTS ..............................................................................................................8  

CHAPTER 1 — INTRODUCTION ..................................................................................................9  

CHAPTER 2 — METHODS ...........................................................................................................17  
  Metrics....................................................................................................................................18  
  Power....................................................................................................................................26  

CHAPTER 3 — RESULTS ...........................................................................................................32  
  Metrics....................................................................................................................................32  
  Power....................................................................................................................................46  

CHAPTER 3 — DISCUSSION .....................................................................................................51  
  Metrics....................................................................................................................................53  
  Power....................................................................................................................................58  

CHAPTER 4 — CONCLUSION/RECOMMENDATIONS .................................................................63  

REFERENCES ...........................................................................................................................66  

APPENDIX .................................................................................................................................73  
  A. RStudio OutPut File ...........................................................................................................73
LIST OF TABLES

Table 1 Annual CABIN Stations In 2011 And 2012, Banff National Park, AB…………..19
Table 2 Summary Of Benthic Macroinvertebrate Metrics (Raw Mean ± Standard Error)..33
Table 3 BACI Contrast Estimates For Benthic Macroinvertebrate Metrics…………………35
Table 4 Results Of The BACI Power Analysis…………………………………………..48
LIST OF FIGURES

Figure 1. Banff National Park, Alberta, Canada ................................................................. 15
Figure 2. Bison and fire study watersheds in Banff National Park, AB .............................. 20
Figure 3. Annual CABIN sampling stations in 2011 and 2012, Banff National Park, AB ....... 18
Figure 4. Profile plots of invertebrate abundance and richness ......................................... 36
Figure 5. Profile plots of Shannon diversity and true diversity ......................................... 38
Figure 6. Profile plots of EPT abundance and richness ....................................................... 40
Figure 7. Profile plots of chironomid abundance and percent composition ....................... 42
Figure 8. Profile plots of relative composition of functional feeding groups ....................... 44
Figure 9. Relative composition of major taxonomic groups based on abundance ............... 45
Figure 10. Relative composition of functional feeding groups (FFGs) based on abundance ..... 46
Figure 11. Power curves for selected metrics ................................................................. 50
ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor and friend, Mark Taylor, for working with me to find a project that aligns with my current research interests, and his engagement throughout the learning process of this master’s thesis. The support he provided through Parks Canada from data to funding is deeply appreciated. It is always a pleasure working with him and I hope that the future brings more opportunity for partnership. Furthermore, I would like to thank my old colleagues, Lucie Sliva and Mike VanderMeulen. Lucie was instrumental in my transition from fish to invertebrates, and her mentorship gave me a solid foundation to build upon; Mike was also an excellent mentor, always trying to help me understand the complex world of statistics and available for interesting discussions. Finally, I would like to thank my family and friends; I could not have finished this without their continuous support and encouragement.
CHAPTER 1—INTRODUCTION

The National Park Act (2013) mandates the maintenance of ecological integrity, which means “with respect to a park, a condition that is determined to be characteristic of its natural region and likely to persist, including abiotic components and the composition and abundance of native species and biological communities, rates of change and supporting processes”. One of the indicators of ecological integrity is aquatic ecosystem health. In addition to routine baseline monitoring of park waters, Parks Canada is diligent in monitoring the potential effects of their own management and restoration activities. For example, Parks Canada is proposing to re-introduce a herd of bison into the remote Panther and Dormer river valleys of Banff National Park’s (BNP) backcountry in 2017 and 2018. Historically, bison played a key role in the ecosystem by grazing and causing physical disturbance of vegetation, which helps maintain a patchwork of meadows and grasslands that other animals depend on. Reintroducing bison is also an opportunity to restore cultural connections that have been lost for more than a century. While the project is expected to have these conservation gains, there are some concerns about the impacts of bison on aquatic ecosystems because they can alter habitat by widening and trampling stream channels which may lead to sediment and nutrient suspension (Larson et al. 2013). In addition to the bison reintroduction, Parks Canada will be conducting prescribed burns to maintain natural grasslands for bison to graze. The combination of a new large grazer on the landscape and prescribed burning prompted Parks Canada to monitor these potential disturbances to aquatic ecosystems.
Disturbance has major influence on the structure and dynamics of communities as well as the life history strategies of species. Sousa (1984) provided a general review of disturbance in natural communities but noted that there was very little research on disturbance in freshwater environments, especially streams. The first definition of disturbance applicable to the lotic environment was provided by Resh et al. (1988), and was defined as “any discrete event in time that is characterized by a frequency, intensity, and severity outside a predictable range, and that disrupts ecosystem, community, or population structure and changes resources or the physical environment” (p. 433). The criterion of “predictable range” recognizes that aquatic ecosystems are spatially and temporally variable, and that the inhabitants of lotic environments are adapted to predictable seasonal fluctuations such as discharge, temperature, dissolved oxygen, etc. (Resh et al., 1988). With respect to freshwater streams, disturbance can be integral to the structure and function of these systems.

It is common to use biological communities to assess the impacts of disturbance to aquatic ecosystems. Biomonitoring is the use of living organisms or their responses to characterize the quality of the aquatic resources (Merrit, Cummins, & Berg, 2008). Traditional water quality monitoring programs in the aquatic environment provide only a “snapshot” of conditions that existed at the time of sampling, whereas biological indicators can provide information on exposure to past conditions (Environment Canada, 2002). Therefore, an advantage to using biomonitoring is the ability to detect past disturbances, which may have happened weeks or months before the collection of samples in the field.
Species used for biomonitoring in freshwater systems can include fish, periphyton, and other groups of organisms, but benthic macroinvertebrates are the most common (Bonada, Prat, Resh, & Statzner, 2006; Fore & Karr, 1996; Merrit et al., 2008). As per the Canadian Aquatic Biomonitoring Integrated Network (CABIN) wadeable streams protocol, benthic macroinvertebrates are defined as organisms with no backbone that inhabit the bottom substrates of streams and are sampled using 200-500 micrometer screens (Environment Canada, 2012).

CABIN is a national biomonitoring program that evaluates aquatic ecosystem condition using standardized sampling protocols and the Reference Condition Approach (RCA) study design (Environment Canada, 2002). Since the bison reintroduction and prescribed burn study (subsequently referred to as the bison and burn study) investigated in this thesis employs CABIN sampling protocols, there is potential for using RCA in the future (i.e., a multivariate approach). However, a Before-After-Control-Impact study design was preferred since this approach has the benefit of comparison to a pre-disturbance state.

Benthic macroinvertebrates are frequently used in assessments of aquatic ecosystem health in freshwater systems because of their ubiquitous occurrence, diversity and abundance, long life cycles, sedentary nature, response to a wide variety of stressors, and importance to the aquatic food web (Bonada et al., 2006; Environment Canada, 2012). In addition, there are standard protocols for the collection of benthic macroinvertebrates (i.e., CABIN) and the sampling equipment is relatively inexpensive and simple to use. With respect to data analysis, many biotic and diversity metrics have been developed for the assessment of benthic communities and the responses of many common species to various types of pollution have been
documented (Rosenberg and Resh, 1993). Finally, benthic macroinvertebrate data are suitable for many experimental design approaches that are generally used in biomonitoring.

Two study design approaches routinely used in biomonitoring are multimetric and multivariate. A multimetric approach analyzes a number of metrics that represent the structure, function or other features of the biological community to assess impacts that are related to disturbance (Bonada et al., 2006). The calculation of multiple metrics to evaluate a potential impact reduces the risk of an incorrect assessment. Multivariate approaches examine the predicted patterns at reference sites so that predicted patterns at a test site (i.e., potentially impacted) can be compared. If the biota found at a test site is different from what is expected to be present at a reference site, there is evidence of disturbance. For the purpose of this thesis, a multimetric approach was used because the resulting univariate analysis supports the BACI study design in the bison and burn study.

The inherent variability of natural systems should be carefully examined when designing environmental studies that assess disturbance (Benedetti-Cecchi, 2001). A BACI study design is not only based on comparing before and after a disturbance, but also comparing control and impact sites to help account for natural variation that influences the sites in the same way (Stewart-Oaten & Bence, 2001; Smith, 2002). By measuring the controls before and after a disturbance, the temporal change that occurs in the absence of the disturbance can be measured, and then the differential change in the difference over time between control and impact sites is the evidence of an environmental disturbance (Schwartz, 2014).
The overall purpose of this thesis is to inform the study design of an aquatic ecosystem monitoring program that assesses the impacts of a bison reintroduction and associated prescribed burn program in Banff National Park, Alberta (Figure 1). This purpose were achieved using the following objectives: 1) evaluate various benthic macroinvertebrate metrics based on the biology of the organisms, their functional groups, and their relevance to potential impacts that may be expected; and 2) conduct a power analysis on data collected from another aquatic ecosystem disturbance study. The data that was analyzed for this thesis was from 2011 and 2012 and was related to a flood that occurred on one river in the Park. The physical effects of a flood were considered a suitable surrogate for the physical effects of bison and burning for the following reasons. The increased flows as a result of flooding have the ability to cause slope failures, streambank erosion, substrate instability and scouring of the substrate surface, and loss of habitat and biota (Resh et al. 1988; Reice, 1985; Rempel, Richardson, & Healy, 1999). The effects that would be similar to the disturbance of bison and burning would include the direct trampling of stream organisms, decreased riparian vegetation, and increased soil compaction and erosion (Fritz, Dodds, & Pontius, 1999; Larson, Grudzinski, Dodds, Daniels, Skibbe, & Joern, 2013; Meadows, 2011). While flooding and bison and fire may not be equal, they are comparable because both would result in benthic macroinvertebrate drift, loss of habitat and direct mortality. In addition, the flood was approximately 10 times higher than the annual base flows so larger effects such as the movement of large boulders and slope failure were not observed.

It is important to identify appropriate biological endpoints when using benthic macroinvertebrates as biological indicators (Clements & Kiffney, 1995). Multiple metrics can be
calculated that represent changes to the structure, function, or other relevant characteristics of the biological assemblage (Bonada et al., 2006). Frequently used metrics include species abundance and composition, taxonomic richness, diversity, pollution tolerance/intolerance and functional feeding groups. Selecting the appropriate metrics is challenging because of the inherent variability of benthic macroinvertebrate communities attributed to their spatial and temporal heterogeneity (Carlisle & Clements, 1999). Each metric should be evaluated for its ability to separate natural variation from environmental disturbance (Clements & Kiffney, 1995). My first objective was to select the most relevant metrics that will detect an impact of bison or prescribed burning to the aquatic ecosystem. Both the biological and statistical considerations amongst the different metrics will be discussed.
Another consideration to the design of this biomonitoring program was statistical power. Power is associated with Type II error, which is falsely concluding that there is no effect of environmental disturbance when there is in fact an effect. In environmental management, a Type
II error can mean unnecessary environmental degradation when it could have been prevented. Two ways that statistically sufficient power can improve an environmental monitoring program are: 1) allowing the most sensitive tests to be chosen among those applicable to the data; and 2) a \textit{a priori} power analysis can indicate the sample size or number of monitoring years necessary to detect a biologically important effect of a disturbance (Fairweather, 1991). A study with high power helps to reduce the chance of Type II error. My second objective was thus to conduct a power analysis on the benthic macroinvertebrate data from before and after a major flood on the Spray River, Banff National Park, AB. The power of tests (i.e., ability to detect disturbance) will be intrinsically different for different metrics (Underwood, 1991); therefore, evaluating which metrics give the most power within the study budget is valuable to the design of the biomonitoring program. The results of the power analysis will help guide which metrics will be most effective at detecting disturbance in the bison and fire study.

The use of sound science ensures that the public will know that the ecological integrity of their parks is being well protected. An aquatic disturbance has the potential to impact both the physical habitat of a stream and its biological community; therefore, a biomonitoring program will help to detect an effect if one exists (i.e., type II error). The methods developed for this thesis will help shape a scientifically defensible biomonitoring program that can be also be communicated to the general public. This communication is especially important from a conservation biology perspective, as it helps develop the public and political support that is needed for the conservation of park ecosystems.
CHAPTER 2 — METHODS

For the bison and burn study, watersheds were assessed in the eastern slopes of the Rocky Mountains, within Banff National Park, that include: 1) Clearwater River - no bison or prescribed burn, 2) Red Deer River - no bison or prescribed burn, and 3) Panther River - bison and prescribed burn (Figure 2). Each watershed was chosen for a specific reason. All watersheds are similar in that they are in the foothills ecoregion, have east-west orientation, and all three rivers flow eastward into the prairies. Bison will be introduced into the Panther watershed because it is known historic bison habitat and has the best habitat suitability (e.g. grasslands for grazing). The Clearwater and Red Deer watersheds are control watersheds where no deliberate fires will be set and bison are not expected to roam there. Sampling stations were established equal distance apart at 1 km in the mainstem of each watercourse in Fall 2014. The stations were located in the lower portions of each watershed since the lower portion of the Panther River watershed provides the most suitable habitat for bison and is where they are expected to congregate for the majority of the year. Prescribed burning has occurred in Fall 2015 in the lower portion of the Panther watershed in order to improve and maintain sufficient meadow grass habitat for the bison. In areas where meadow features are being invaded by trees and shrubs, prescribed burning is a tool that helps to maintain and restore the herbaceous ecosystem. Regions that have sufficient moisture to support trees are maintained as meadow ecosystems by a regular fire disturbance regime, which enables herbs and grasses to thrive and eliminates or reduces the woody plants (Larson, Grudzinski, Dodds, Daniels, Skibbe, & Joern, 2013). These meadows will provide critical year round habitat and foraging areas for a variety of animals, especially bison.
Benthic macroinvertebrate data used for metric analysis was from the annual CABIN sampling program conducted in Banff National Park (Figure 3). The benthic community was analyzed as a classic spatial design rather than the reference area approach suggested by CABIN. Nine of the 17 CABIN stations on various creeks and rivers were sampled in both 2011 and 2012 for the annual CABIN program (Table 1). The general objective of the CABIN program is to add enough sampling stations to build a data set for the reference condition model; therefore, not all
stations were sampled in both years. Table 1 indicates the control and impact sites that were used in the BACI analysis.

Table 1

*Annual CABIN Sampling Stations in 2011 And 2012, Banff National Park, AB.*

<table>
<thead>
<tr>
<th>Station</th>
<th>River</th>
<th>Site Class</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-B-11-1</td>
<td>Bow</td>
<td>Control</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>B-B-11-2</td>
<td>Bow</td>
<td>Control</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BAltrude</td>
<td>Altrude</td>
<td>Control</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>BBrewster</td>
<td>Brewster</td>
<td>Control</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>BCarrot</td>
<td>Carrot</td>
<td>Control</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>BCascadeLow</td>
<td>Cascade</td>
<td>Control</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BFlintsPk</td>
<td>Flints Park</td>
<td>Control</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>BFortyMile</td>
<td>Forty Mile</td>
<td>Control</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>BHealy</td>
<td>Healy</td>
<td>Control</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>BJohnson</td>
<td>Johnson</td>
<td>Control</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>BJohnston</td>
<td>Johnston</td>
<td>Control</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>BSpray01</td>
<td>Spray</td>
<td>Impact</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BSpray02</td>
<td>Spray</td>
<td>Impact</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BSpray03</td>
<td>Spray</td>
<td>Impact</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BSpray04</td>
<td>Spray</td>
<td>Impact</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BSpray05</td>
<td>Spray</td>
<td>Impact</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BSpray06</td>
<td>Spray</td>
<td>Impact</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Figure 3. Annual CABIN sampling stations in 2011 and 2012, Banff National Park, AB.

Six stations on the Spray River were sampled in both years and were considered the impacted sites in the BACI analysis because of the flooding that occurred in 2011 (Table 1, Figure 2). The Spray River is part of a Parks Canada longitudinal study that uses CABIN to monitor the impacts of the Canyon dam that is located on the River. The flood that occurred in August 2011 stemmed from Trans Alta plant operations, which uses the water from the Spray Lakes reservoir for the generation of electrical power. TransAlta released water from the reservoir in an effort to reduce water levels in the reservoir which resulted in a discharge rate of non-polluted water that was approximately 10X higher than August base flows (M. Taylor,
personal communication, November 27, 2015). This hydrological event was limited to the Spray River; therefore, the Spray River was the only river that experienced a disturbance in 2011 and was considered the after period in the BACI analysis. CABIN data from 2012 was considered the before period based on the assumption that the benthic macroinvertebrate community in the Spray River had recovered to normal conditions. This assumption seemed reasonable based on the following discussion.

There have been numerous studies on flooding and it has been found that flooding can deplete the biota of streams but recovery can be rapid (Lake, 2004). A study conducted by Reice (1985) tumbled patches of cobbles, which would be a realistic simulation of substrate movement during a flood, and monitored the benthic macroinvertebrate recovery. The majority of taxa had reductions in population abundance immediately following the disturbance, but recovered to near normal population levels within four weeks of the disturbance (Reice, 1985). Scrimgeour, Davidson, and Davidson (1988) also found that benthic macroinvertebrate abundance, biomass, and richness was severely reduced after major flooding but recovered to pre-flood levels within 132 days. Subsequently, Quinn and Hickey (1990) investigated the threshold levels at which physical factors, including flooding, have a strong influence on benthic macroinvertebrate community characteristics on 80 rivers with gravel/cobble substrates in New Zealand. In general, they found that flood flows would have to exceed 20X the median flow rate to have significant effects on benthic macroinvertebrate abundance and richness (Quinn & Hickey, 1990). Finally, a study conducted by Rempel et al., (1999) showed that the majority of benthic
Macroinvertebrates shift to the near shore zone in the Fraser River, British Columbia, where there is more stable substrate and lower hydraulic stress during flooding. Essentially, factors such as the presence of flow refugia (i.e., near shore and hyphoeic zones with gravel/cobble substrates), and drift from upstream habitats could be major contributors to the short recovery time of the benthic community in headwater streams.

As previously mentioned, the Spray River flood was used as a surrogate disturbance for the reintroduction of bison and prescribed burning since both are large-scale and occur at the landscape level. A flood disturbance is likely a realistic surrogate because both flooding and bison are expected to change channel morphology and possibly cause sedimentation. While the impacts of bison are likely more chronic and long-term compared to the acute impacts of a flood, the magnitude of each disturbance is likely similar (i.e., medium-scale disturbance compared to a large-scale disturbance such as a toxic spill or significant watershed development). Although the disturbance of a flood is not a perfect match for bison and burning, it was the best available data for this analysis. A pilot study would be a superior option but the reality is that management agencies typically don’t have funding for pilot studies and have to maximize the resources and data that are available for scientific research.

The data set used was unbalanced meaning there were stations that were not sampled in both the before and after period. The missing data may result in a reduction of power; however, the sites sampled in only one of the periods still contribute valuable information to the variance components that were integrated into the analysis, and are needed when the formal hypothesis
tests are conducted. The statistical analysis used did account for this unbalanced study design. The caveat is that sophisticated analytical approaches such as those found in R be used because of the complex algorithms required.

Field sampling was conducted in accordance with the CABIN protocols (Environment Canada, 2002) for both the annual CABIN program and bison and burn study. Parks Canada staff and volunteers from their citizen scientist program executed the field sampling in 2011 and 2012 (for annual CABIN program) and 2014 (for bison and burn study). CABIN protocols included sample collection, quality assurance/quality control procedures, equipment requirements, and field notes/observations. Sampling occurred in the fall when: 1) most taxa of the benthic community are in an aquatic life stage; 2) taxa are in the stage of their life cycle that allows them to be identified at levels of taxonomic resolution required by CABIN; and 3) water levels are low, which helps sampling efficiency and indicates the wetted portions of stream channel that are stable habitats (Environment Canada, 2002). Benthic macroinvertebrate samples were collected from the erosional zone (riffle or straight run) using a 400 µm kick net since these areas generally support a greater diversity of organisms than depositional (pool) habitats (Environment Canada, 2002). At each sample station the benthic macroinvertebrate sample was collected as follows:

1. Place the kick net downstream with the flat side resting on the substrate.
2. Walk backwards upstream, dragging the net along the bottom and kicking the substrate to a depth of 5 cm to 10 cm.
3. Zig zag from bank to bank while continuing upstream for a total of three minutes.
4. Working over a pail or tray, remove the cap attached to the collection cup at the end of the net and transfer contents to sample container.

5. Wash any material remaining in the cup or net into the container (use tweezers to remove any cling insects) ensuring to leave room for ethanol.

6. Carefully rinse and remove any large organic debris or stones.

7. Label the container (inside and outside) with unique station ID, date, time, sampler name, and number of containers.

8. Preserve with ethanol ensuring that any water in the container does not dilute the preservative below 20%.

The river reach was defined as six times the bankfull width representing a complete pool:riffle sequence (Environment Canada, 2002). At each sample location, flow rate (velocity and depth) and channel measurements were also taken. Channel measurements included: wetted width, bankfull width, bankfull wetted width, and slope. Substrate characterization accomplished by conducting a 100-pebble count and associated embededness as well as an assessment of the surrounding material. Benthic macroinvertebrate samples were collected concurrently with the recommended surface water quality samples and appropriate in situ parameters.

Benthic macroinvertebrate samples were identified to at least the genus level. Taxonomic identification and enumeration was conducted by a qualified professional who is certified by the North American Benthological Society according to the protocols described in Environment Canada (2011).
For each benthic replicate sample, several metrics of community composition that are commonly used were calculated that assess the structure and function of the benthic macroinvertebrate community. These metrics include:

1. **Abundance**, the total number of individuals.

2. **Taxon richness S**, where S is the number of distinct taxa per sample.

   \[ H = -\sum_{i=1}^{S} p_i \times \log p_i \]  

   where \( p_i \) is the proportion of S made up by the \( i^{th} \) species. The Shannon Index is commonly used to characterize species diversity in a community, and accounts for both abundance and evenness of the species present. The resulting product is multiplied by -1 to ensure a positive index value. The index ranges from 1-4 and a higher index value indicates a more diverse community (Jost, 2006).

3. **True Diversity** (\( e^H \)), which is the effective number of species. True diversity attempts to address an inherent weakness in diversity indices such as Shannon by determining the number of equally-common species required to give a particular value of an index (such as the Shannon in the formula). Converting indices to True Diversities (i.e. effective numbers of species) gives them a set of common behaviors and properties. After conversion, diversity is always measured in units of number of species, no matter the index used, which allows for easier comparison and interpretation. It also helps to avoid the serious misinterpretations generated by the nonlinearity of most diversity indices (Jost, 2006).

The abundance and relative percentage of EPT taxa: Ephemeroptera (mayflies),
Plecoptera (stoneflies) and Tricoptera (caddisflies) were calculated for each station. Since the aquatic larvae or nymphs of these insect orders are pollution sensitive, EPT is commonly used as an ecosystem integrity index for lotic environments (Mandaville, 2002). Alternatively, the highly adaptable Chironomidae (midge larvae) were used to calculate the abundance and relative percentage of tolerant organisms per station. These organisms are generally tolerant of pollution and their numbers are thought to increase in degraded conditions.

In addition to the above, metrics used to interpret community function were calculated based on feeding strategy of benthic macroinvertebrates (Merritt et al., 2008). The five general feeding strategies that are common in benthic macroinvertebrates are: collecting filterers, gathering collectors, predators, scrapers, and shredders. Based on their ecological importance within the food web, benthic macroinvertebrate functional feeding group (FFG) analysis has been shown to reveal certain ecological values (Fairchild, Faulds, & Saunders, 1999). For example, in lotic ecosystems, specialized feeders such as scrapers and shredders are presumed to be more sensitive to disturbance, whereas generalists such as filterer collectors and gathering collectors are more tolerant to disturbance that might alter the availability of a certain food (Rawer-Jost, Bohmer, Blank, & Rahmann, 2000). Predators have also been observed to increase in disturbed stream systems (Kearns and Karr, 1994; Rawer-Jost et al., 2000).

**Power**

The general methods used for statistical analysis were based on the course notes provided by Schwartz (2014) and all tests were conducted in RStudio (R Core Team, 2014). A key assumption in a BACI study design is that the system in question is in equilibrium before and
after the disturbance. Hypothesis testing must be combined with an estimate of effect size (i.e., BACI interaction) and measurement of precision. This was conducted using a two-way analysis of variance (ANOVA) on the following general model:

\[ Y = SiteClass + Period + SiteClass*Period \]

where \( Y \) represents the metric being tested, \( SiteClass \) represents the site effect (i.e., control vs. impact) that cannot be measured or controlled, \( Period \) represents the time effect which is the difference in the mean before and after the disturbance, and \( SiteClass*Period \) interaction represents the BACI interaction which is the differential change evident of an environmental disturbance. The overall hypotheses tested were:

\( H_0: \) there is no BACI interaction (i.e., no interaction between temporal change and siteclass)

\( H_A: \) there is a BACI interaction (i.e., an interaction between temporal change and siteclass)

These tests were conducted with the Car Package (Fox & Weisberg, 2011) first using the \texttt{lm()} function to fit the model. Since the design is unbalanced (i.e., missing data), a Type III test for sum of squares (SS) was used for the effects test in the \texttt{ANOVA()} function as opposed to the RStudio (R Core Team, 2014) default of Type I SS. Type I SS evaluates each effect in the model sequentially (evaluation of first effect entered, then second effect entered, etc.) whereas the Type III SS evaluates all effects by taking into consideration all the other effects in the model (not just the effects entered before; Field, Miles, & Field, 2012). Essentially, the treatment contrasts from
the ANOVA() function were set “sum-to-zero” rather than the default “treatment” prior to fitting the lm() function. The estimate of the BACI contrast was calculated using least square means with the following equation:

\[
BACI = \mu_{CA} - \mu_{CB} - (\mu_{TA} - \mu_{TB}) = \mu_{CA} - \mu_{CB} - \mu_{TA} + \mu_{TB}
\]

Where \(\mu_{CA}\) is control-after, \(\mu_{CB}\) is control-before, \(\mu_{TA}\) is treatment after, and \(\mu_{TB}\) is treatment-before. Least square means were estimated from the linear model as opposed to the means that are averages of the raw data. The use of both Type III ANOVAs and least square means were required for analyzing the unbalanced study design in RStudio (R Core Team, 2014).

Although ANOVA was required to get the BACI contrast, an improved method for analyzing benthic macroinvertebrate data uses linear mixed-models. By adding random effects, the model can account for pseudoreplication, which is inherent in these types of studies. The benthic macroinvertebrate stations were clustered in watersheds so were considered pseudoreplicates, which violates the assumption of independence required for linear models. In addition, stations were repeatedly sampled over time, which violates the assumption of independence a second time. Using a more refined analysis such as a mixed-model provides estimates of both station-station and station-year variation. The following mixed model was analyzed with the lmerTest Package (Kuznetsova, Brockhoff, & Christensen, 2013) from the lmer() function:
\[ Y = \text{SiteClass} + \text{Period} + \text{SiteClass}*\text{Period} + \text{Station} (R) + \text{Station}*\text{Period}(R) \]

where Station(R) represents the random effect of station, and Station*Period(R) represents the inconsistency of the stations over time. Three variance components were calculated from the mixed-model: 1) station variation signifying not all stations will have similar responses in the mean for each year. Although somewhat controlled by stratifying each reach into riffle habitat as per CABIN methods, some stations may naturally have better habitat and higher metric values (i.e., abundance) than other stations; 2) station-period interaction variation indicating that station mean responses may not be the same for the same year specific factors. A station with better habitat could deal with a drier year so a change in abundance may not appear as great as opposed to stations with poorer habitat; and 3) residual variation where not all stations have the same response at a specific location in a year.

Power analysis is a useful tool to measure the ability of a study design to detect an impact if one existed (Quinn & Keough, 2000). A BACI power analysis was conducted a priori using data from 2011 and 2012 annual CABIN sampling program in Banff National Park, Alberta in attempt to evaluate and inform the statistical power of the bison and burn analysis with respect to the number of years and the number of replicates necessary to detect any effects and responses related to the bison reintroduction and prescribed burning. The information required to calculate power \((1 - \beta)\), where \(\beta\) is the Type II error) includes: \(n = \) number of samples (n), \(\alpha = \) significance level (Type I error rate), \(\sigma = \) estimate of variation in the data, and the effect size.
A BACI power function created by Schwartz (2014) was used to conduct the power analysis. This method generates fake data with no variability, using only the means from the treatment applied to the observation. Data are generated that correspond to the two years of data (before and after), two sites (control and impact), and multiple benthic macroinvertebrate stations. The assessed metric values take on one of the four means corresponding to the treatment-before impact, treatment-after impact, control-before impact, and control-after impact. The BACI interaction, which is the differential change between the temporal change and siteclass in the statistical model, represents the biologically important difference. The mean data were analyzed in the BACI power model using the variance components estimated from the linear mixed-model conducted on the 2011 and 2012 CABIN data. A function of the F-statistic provided the estimated power for the study design. Since the objective was to inform the bison and burn program, the specific details known about this program were used in the power analysis. This includes a sample size of eight replicates per river, and two control and one treatment watershed. The BACI contrast was estimated from the ANOVA and the three aforementioned variance components were used from the mixed-model analysis. Power was assessed for 1, 2, 5, 10 and 15 years of after monitoring data using only one year of before data. It is known that two or three years of pre-disturbance data will be collected for the bison and burn program; however, the data analyzed had only one year of data and that was considered the before period (i.e., 2012 data was used for the before period and 2011 data was used for the after period) so the year effect cannot be separated from just noise. Essentially, until there are multiple years of before data collected, the year-to-year effect is integrated into the residual
variation. Since this effect cannot be separated out, power was reduced for the current analysis; however, the fact that multiple years of pre-disturbance data will be collected for the bison and burn study will definitely result in higher power.
CHAPTER 3 — RESULTS

Metrics

Average values of benthic invertebrate metrics of the four BACI groups are shown in Table 2 and include standard errors (SE) for each. Metrics are also summarized in Figures 4 through 10. Results from statistical estimates of the BACI contrast for these metrics between groups are shown in Table 3. The results of the power analysis are provided in Table 4 and Figure 11. An example of data output for the analysis of abundance in RStudio ((R Core Team, 2014) is provided in Appendix 1.

Analyses of variance (ANOVA) was used to test the BACI contrast (i.e., differential change between control and impact stations) in the benthic macroinvertebrate metrics to explore the suitability of each for detecting disturbance to the aquatic ecosystem. The following assumptions for ANOVA were met: 1) the response variables (i.e., metrics) were interval scale; 2) the standard deviations at each site-year combination were approximately equal (i.e., ratio of smallest to largest standard deviation was less than 5:1); 3) the observations were considered functionally independent. Stations were sufficiently far enough from each other to be spatially independent and sampling of the same stations each year allowed for some blocking of temporal effects (it is noted that issues of pseudoreplication are inherent aquatic ecosystem studies but will be examined in the discussion); and 4) the data were examined for normality and did not deviate considerably (Schwartz, 2014). Of all the above assumptions, it is more important that standard deviations are equal and the observations are independent since the relative robustness of ANOVA can deal with minor deviations from normality in the data (Schwartz, 2014).
Table 2

Summary of Benthic Macroinvertebrate Metrics (Raw Mean ± Standard Error).

<table>
<thead>
<tr>
<th>Metric</th>
<th>Control Before</th>
<th>Control After</th>
<th>Impact Before</th>
<th>Impact After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abundance</td>
<td>298 ± 12</td>
<td>315 ± 8</td>
<td>341 ± 18</td>
<td>320 ± 22</td>
</tr>
<tr>
<td>Abundance</td>
<td>298 ± 12</td>
<td>315 ± 8</td>
<td>341 ± 18</td>
<td>297 ± 4</td>
</tr>
<tr>
<td>Richness</td>
<td>31 ± 4</td>
<td>33 ± 2</td>
<td>30 ± 2</td>
<td>34 ± 2</td>
</tr>
<tr>
<td>Shannon Diversity</td>
<td>2.50 ± 0.17</td>
<td>2.55 ± 0.07</td>
<td>2.27 ± 0.14</td>
<td>2.41 ± 0.09</td>
</tr>
<tr>
<td>Shannon Evenness</td>
<td>0.73 ± 0.03</td>
<td>0.73 ± 0.01</td>
<td>0.67 ± 0.04</td>
<td>0.69 ± 0.02</td>
</tr>
<tr>
<td>True Diversity</td>
<td>9 ± 1</td>
<td>13 ± 1</td>
<td>10 ± 1</td>
<td>11 ± 1</td>
</tr>
<tr>
<td>EPT Abundance</td>
<td>240 ± 24</td>
<td>242 ± 17</td>
<td>230 ± 33</td>
<td>267 ± 33</td>
</tr>
<tr>
<td>EPT Richness</td>
<td>18 ± 1</td>
<td>22 ± 1</td>
<td>18 ± 1</td>
<td>22 ± 2</td>
</tr>
<tr>
<td>% EPT</td>
<td>80 ± 7</td>
<td>77 ± 6</td>
<td>69 ± 11</td>
<td>83 ± 5</td>
</tr>
<tr>
<td>Chironomid Abundance</td>
<td>44 ± 16</td>
<td>59 ± 20</td>
<td>104 ± 43</td>
<td>44 ± 14</td>
</tr>
<tr>
<td>% Chironomid</td>
<td>15 ± 6</td>
<td>18 ± 6</td>
<td>29 ± 11</td>
<td>15 ± 5</td>
</tr>
<tr>
<td>% Gathering Collectors</td>
<td>49 ± 8</td>
<td>32 ± 8</td>
<td>47 ± 7</td>
<td>37 ± 10</td>
</tr>
<tr>
<td>% Filtering Collectors</td>
<td>4 ± 2</td>
<td>2 ± 1</td>
<td>2 ± 1</td>
<td>1 ± 1</td>
</tr>
<tr>
<td>% Predators</td>
<td>12 ± 2</td>
<td>10 ± 1</td>
<td>11 ± 3</td>
<td>14 ± 2</td>
</tr>
<tr>
<td>% Scrapers</td>
<td>22 ± 10</td>
<td>47 ± 7</td>
<td>35 ± 9</td>
<td>30 ± 12</td>
</tr>
<tr>
<td>% Shredders</td>
<td>13 ± 2</td>
<td>10 ± 2</td>
<td>5 ± 1</td>
<td>18 ± 7</td>
</tr>
</tbody>
</table>

Notes: 1 n=6; 2 n=8; 3 n=6; 4 n=6; 5 BSpray06 After was removed from analysis; 6 BAltrude and BCascadeLow Before were removed from analysis; 7 BAltrude Before was removed from analysis.

The average abundance was higher in both the before and after impact groups when compared to control groups (Table 2). The profile plot of average abundance indicates an overall increase for the control group as opposed to a decrease in abundance for the impact group (Figure 5a). When assessing individual stations, the above trend was shown with the exception of
one impact station which had an increase between the before and after period (Figure 5b). There was no evidence of a difference in the means detected for the BACI contrast ($t(22) = -1.265, p = 0.219$; Table 3). The data were analyzed a second time with the BSpray06-after data point removed since it was substantially larger than the other stations in the after period and was therefore considered an outlier. This resulted in a larger BACI contrast and there was evidence in a difference in the means ($t(21) = -2.544, p = 0.02$; Table 3). Even though there were slight differences noted in abundance, species richness was similar amongst all groups as illustrated by the similar slopes between the before and after period (Figure 5c and 5d). There was no evidence of a difference in the means detected for the BACI contrast ($t(22) = 0.301, p = 0.77$; Table 3).
Table 3

*BACI Contrast Estimates for Benthic Macroinvertebrate Metrics.*

<table>
<thead>
<tr>
<th>Metric</th>
<th>BACI Contrast</th>
<th>Standard Error</th>
<th>Degrees of Freedom</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abundance</td>
<td>-39</td>
<td>31</td>
<td>22</td>
<td>0.22</td>
</tr>
<tr>
<td>Abundance(^1)</td>
<td>-61</td>
<td>24</td>
<td>21</td>
<td>0.02</td>
</tr>
<tr>
<td>Richness</td>
<td>1</td>
<td>4.84</td>
<td>22</td>
<td>0.77</td>
</tr>
<tr>
<td>Shannon Diversity</td>
<td>0.098</td>
<td>0.24</td>
<td>22</td>
<td>0.68</td>
</tr>
<tr>
<td>Shannon Eveness</td>
<td>0.025</td>
<td>0.05</td>
<td>22</td>
<td>0.61</td>
</tr>
<tr>
<td>True Diversity(^2)</td>
<td>-2</td>
<td>2</td>
<td>20</td>
<td>0.30</td>
</tr>
<tr>
<td>EPT Abundance</td>
<td>35</td>
<td>53</td>
<td>22</td>
<td>0.52</td>
</tr>
<tr>
<td>EPT Richness(^3)</td>
<td>1</td>
<td>3</td>
<td>21</td>
<td>0.82</td>
</tr>
<tr>
<td>% EPT</td>
<td>16</td>
<td>15</td>
<td>22</td>
<td>0.29</td>
</tr>
<tr>
<td>Chironomid Abundance</td>
<td>-75</td>
<td>51</td>
<td>22</td>
<td>0.16</td>
</tr>
<tr>
<td>% Chironomid</td>
<td>-17</td>
<td>14</td>
<td>22</td>
<td>0.24</td>
</tr>
<tr>
<td>% Gathering Collectors</td>
<td>8</td>
<td>17</td>
<td>22</td>
<td>0.65</td>
</tr>
<tr>
<td>% Filtering Collectors</td>
<td>2</td>
<td>2</td>
<td>22</td>
<td>0.42</td>
</tr>
<tr>
<td>% Predators</td>
<td>4</td>
<td>4</td>
<td>22</td>
<td>0.37</td>
</tr>
<tr>
<td>% Scraper</td>
<td>-30</td>
<td>19</td>
<td>22</td>
<td>0.13</td>
</tr>
<tr>
<td>% Shredder</td>
<td>16</td>
<td>8</td>
<td>22</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Notes: \(^1\) BSpray06 After was removed from analysis; \(^2\) BAltrude and BCascadeLow Before were removed from analysis; \(^3\) BAltrude Before was removed from analysis; Significant at the p < 0.05 level.
Large variability was observed for measurements of diversity. Average Shannon diversity was higher in the control stations when compared to the impacted stations, but both groups only showed slight increases from the before and after periods (Table 2, Figure 6a). Individual station analysis exhibited mixed results with some stations increasing and some station decreasing regardless of whether they were control or impact (Figure 6b). Shannon evenness (data not shown) was similar to Shannon diversity and the BACI contrast for both metrics did not show
evidence of a difference in the means ($t(22) = 0.417, p = 0.68$ and $t(22) = 0.522, p = 0.61$, respectively; Table 3). True Diversity, which helps to reduce misinterpretations that arise from the non-linearity of other diversity indices, measures the effective number of species by giving a common set of behaviors and properties to the metric. Two stations were removed as outliers from the analysis of true diversity, BAltrude and CascadeLow from the control-before group. A more prominent increase in average true diversity was observed for the control group whereas there was only a slight increase in the impact group (Figure 6c). Similar to the other diversity metrics, individual stations for true diversity showed mixed results (Figure 6d). There was no evidence of a difference in the means detected for the BACI contrast which is expected because of the large associated error ($t(20) = -1.071, p = 0.30$; Table 3).
Figure 5. Profile plots of Shannon diversity and true diversity.

The abundance and richness of benthic macroinvertebrates belonging to sensitive taxa for all groups were high and typical of fast-flowing mountainous streams. With respect to EPT abundance, there was no distinct change between the before and after period for the control group and only a minor increase was observed for the impact group (Table 2, Figure 7a). The general trend for individual stations was a slight increase between periods with one control and one impact station exhibiting a major decrease (Figure 7b). Although only a slight increase in EPT abundance was noted, the distinct number of EPT taxonomic groups identified increased for
both the control and impact groups (Table 2, Figure 7c). This increase was observed for both the average and individual station analysis with the exception of one control station where EPT richness decreased (Figure 7c and 7d). Note that BAltrude before was removed from the analysis of EPT richness as it was determined an outlier. Similar to the diversity metrics, the large standard error associated with the BACI contrast for ETP abundance and richness did not result in evidence of a difference in the means ($t(22) = 0.65, p = 0.52$ and $t(21) = 0.224, p = 0.82$, respectively; Table 3).
Figure 6. Profile plots of EPT abundance and richness.

The chironomid family, which is generally found to be more tolerant to disturbance, was examined for abundance and percent composition. Both metrics, when averaged, followed similar trends where the control stations increased slightly between periods whereas the impact stations showed a distinct decrease (Table 2, Figure 8a and 8c). Individual stations had variable results which is similar to the majority of metrics evaluated (Figure 8b and 8d). There were no evidence of a difference in the means detected for chironomid abundance or percent composition ($t(22) = -1.457$, $p = 0.16$ and $t(22) = -1.202$, $p = 0.24$, respectively; Table 3). In general,
chironomids were less prevalent in the benthic community for all groups when compared to EPT taxa. The low proportion of tolerant taxa is expected because of their habitat requirements. These organisms generally are associated with finer substrates (e.g., silt and clay) associated with depositional zones whereas the mountain streams sampled consists largely of coarser substrate (e.g., sand and cobbles) associated with erosional zones (e.g., riffle habitat).
Figure 7. Profile plots of chironomid abundance and percent composition.

Analysis of the relative composition of FFGs showed that gathering collectors decreased for both groups between the before and after period but the decrease was more distinct in the impact group (Figure 9a). Increased predators and shredders were observed for impact groups whereas there was little change in the control groups between periods (Figure 9b and 9c). On the other hand, scrapers had a distinct increase in the control but percent composition was relatively similar for the impact group between periods (Figure 9d). Individual stations exhibited similar directional change between periods with all but one station in each the control and impact groups.
contradicting that change (data not shown). Although differences between groups were observed, there was no evidence of a difference in the means found for the BACI contrast estimates for all FFGs (gathering collects: $t(22) = 0.461, p = 0.65$; predators: $t(22) = 0.916, p = 0.37$; scrapers: $t(22) = -1.57, p = 0.13$) with the exception of scrapers which was at the alpha level ($t(22) = 2.121, p = 0.05$; Table 3). The relative composition of filtering collectors was minimal so was not included in the above analysis.
Figure 8. Profile plots of relative composition of functional feeding groups.

The average relative composition of benthic macroinvertebrate communities between the BACI groups was fairly similar with few notable differences (Figure 10). The benthic communities were dominated by ephemeropterans (mayflies) that contributed, on average, approximately 50% to the community total in all groups. In the control group, there was a considerable loss of 10% of the plecopteran (stoneflies) community between the before and after period which was replaced by slight increases in chironomids, and trichopterans (caddisflies). On the other hand, the impact group lost half of its chironomid community between the before
(29%) and after period (14%) which was replaced by increases in ephemeropterans and plecopterans. In general, community composition analysis reveals a diverse community of sensitive EPT taxa, followed by smaller contribution by chironomids and other taxa.

Figure 9. Relative composition of major taxonomic groups based on abundance.

The average contribution from FFGs to the community total and comparisons of their relative abundance is shown in Figure 11. For the control group, gathering collectors dominated the community at 49% followed by scrapers which contributed 22% to the community total in the before period. The opposite was observed in the after period with scrapers dominating the community at 46% followed by gathering collectors which contributed 32% to the community total. For the impact group, the before period had similar contributions from gathering collectors
(45%) and scrapers (38%) with decreases in both of these FFGs in the after period. The decrease in contributions from these two FFGs was replaced by a 13% increase in the shredder community between periods in the impact group. There was little variation in the predator community for all groups and minute contributions to the community total by filtering collectors. Overall, the fairly large presence of shredders and scrapers in all groups support the metric analysis that habitat condition is generally good in watercourses assessed.

![Figure 10](image.png)

**Figure 10.** Relative composition of functional feeding groups (FFGs) based on abundance.

**Power**

It was appropriate to use ANOVA to determine the BACI contrast and test for differences in the means in the above metric analysis. However, the large variability observed, apparent from
the profile plots for the average BACI groups and individual stations of each metric required a more complex analysis to help account for the station effect. This was conducted using a linear mixed model that included station as a random effect and by including the random effect, the linear mixed model provided the variance components (i.e., \( \sigma \)) required for the power analysis.

Power analysis was conducted on 15 metrics to determine how many monitoring years after the reintroduction of bison would be needed to get sufficient power to detect the BACI interaction (i.e., measure of differential change between before and after impact), from the 2011 and 2012 data analysis which used the flooding of the Spray River in 2012 as the disturbance. A metric was considered to have sufficient power at 0.8 (80% power to detect an effect) when \( \alpha = 0.05 \) which is a common choice when conducting a power analysis (Schwartz, 2014). Note that the missing values in the CABIN data set will reduce the precision of both the BACI estimate and power. However, the testing of each metric will provide a general understanding of their performance to aid the selection of the most powerful ones.
Table 4

Results Of The BACI Power Analysis.

<table>
<thead>
<tr>
<th>Metric</th>
<th>BACI Contrast</th>
<th># Years Before</th>
<th># Years After</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Abundance</td>
<td>-39</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>Abundance¹</td>
<td>-61</td>
<td>1</td>
<td>0.14</td>
</tr>
<tr>
<td>Richness</td>
<td>-1</td>
<td>1</td>
<td>0.20</td>
</tr>
<tr>
<td>Shannon Diversity</td>
<td>0.098</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>Shannon Eveness</td>
<td>0.025</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>True Diversity²</td>
<td>-2</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>EPT Abundance</td>
<td>35</td>
<td>1</td>
<td>0.06</td>
</tr>
<tr>
<td>EPT Richness³</td>
<td>1</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>% EPT</td>
<td>16</td>
<td>1</td>
<td>0.06</td>
</tr>
<tr>
<td>Chironomid Abundance</td>
<td>-75</td>
<td>1</td>
<td>0.08</td>
</tr>
<tr>
<td>% Chironomid</td>
<td>-17</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>% Gathering Collectors</td>
<td>8</td>
<td>1</td>
<td>0.11</td>
</tr>
<tr>
<td>% Gathering Collectors</td>
<td>16⁴</td>
<td>1</td>
<td>0.14</td>
</tr>
<tr>
<td>% Filtering Collectors</td>
<td>2</td>
<td>1</td>
<td>0.09</td>
</tr>
<tr>
<td>% Predators</td>
<td>4</td>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>% Scrapers</td>
<td>-30</td>
<td>1</td>
<td>0.21</td>
</tr>
<tr>
<td>% Shredders</td>
<td>16</td>
<td>1</td>
<td>0.10</td>
</tr>
<tr>
<td>% Shredders</td>
<td>32⁴</td>
<td>1</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Notes: ¹ BSpray06 After was removed from analysis; ² BAltrude and BCascadeLow Before were removed from analysis; ³ BAltrude Before was removed from analysis; ⁴ two times the BACI contrast; Bold indicates that power > 0.80.

Out of the 15 metrics evaluated for power, only six cases achieved power of greater than 0.8 in the BACI power analysis (Table 4, Figure 12). With the removal of BSpray06 in the analysis of abundance, power would be realized for the larger BACI contrast estimate at 10 years.
of monitoring. Richness and % scrapers showed high power in five monitoring years at 97%.
This was also the case for % shredders when the power analysis was conducted to detect two
times the BACI contrast. True diversity would attain sufficient power to detect an effect in 10
years of monitoring whereas % gathering collectors would require 15 years of monitoring to
detect two times the BACI contrast. For the other metrics that were assessed, power was below
38% after the collection of 15 years of monitoring data (data not shown). With respect to
Shannon diversity, Shannon eveness and EPT abundance, 15 years of monitoring data did not
have power that exceeded 10%. At this time, the most powerful metrics are richness, % scrapers
and % shredders.
Figure 11. Power curves for selected metrics. Note that the x-axis is the number of monitoring years after disturbance and the red lines indicate a power level of 80%.
Overall, the benthic macroinvertebrate community was characteristic of fast-flowing mountainous streams that have high quality habitat. This was supported by the dominance of sensitive EPT taxa, especially ephemeropterans that contributed approximately 50% to the community composition. When classified based on FFGs, there was a large presence of specialized feeders such as shredders and scrapers, which are also representative of good aquatic ecosystem health. The majority of metrics did not show evidence of a difference in the mean metric values, which may be attributed to the large variability observed between the BACI groups assessed. With respect to power, the most powerful metrics were richness, % scrapers and % shredders, which would detect an effect in just over two years of monitoring data after the reintroduction of bison. Abundance, true diversity and % gathering collectors also achieved power within 5 to 10 years of monitoring data. It is important to note that design improvements to the bison and burn study will result in increased sensitivity in both the metric and power analysis.

When designing a biomonitoring study in the aquatic environment, it is difficult not to have pseudoreplication creep into the study design. Pseudoreplication, first described by Hurlbert (1984) is the use of inferential statistics where treatments are not replicated or the replicates are not statistically independent. In the studies considered for this thesis, the treatment (i.e., flooding or bison and burning) is not replicated. This is typical in ecology as it is generally not possible to provoke a flood in a watercourse just like it is not feasible to introduce bison into multiple watersheds for the sake of replication when it is unknown if they will even survive the
reintroduction. The collection of multiple samples in a single river could also be viewed as pseudoreplication if the scope of inference extends beyond the scope of the system(s) sampled. Since rivers are linear systems, what is influencing stations upstream could also be influencing the stations downstream, and so some of the variables sampled along a stream are arguably not independent. The use of controls in a different watercourse than the treatment river does improve the study design. Finally, there is temporal pseudoreplication. Time cannot be randomized meaning that measurements taken closer together in time are more correlated than measurements taken farther apart in time, which is termed autocorrelation. Essentially, the effects of pseudoreplication are the improper computation of p-values and the standard errors of effect sizes, which can lead to an increased chance of type I error (Schwartz, 2014).

The statistical analyses used in this thesis helps to account for pseudoreplication. For the tests of significance using ANOVA, the replicates were considered functionally independent by placing replicate stations far enough apart to reduce similar influences between stations, but what is “far enough” could have different interpretations by different ecologists, and respecting different taxa. Temporal pseudoreplication may not be that big of an issue for this study which only assessed one year of before and after data; however, this will need to be integrated into the analysis as the number of monitoring years increases for the bison and burn study. The linear mixed model used to extract the variance components for the power analysis included station as a random effect. As the number of monitoring years increases in the bison and burn study, time could also be added as a random effect. The use of linear mixed models helps to recognize and
adjust for pseudoreplication. This will allow for generalizations to be applied to the entire population as opposed to just the individuals being studied (Schwartz, 2014). Although pseudoreplication has been recognized and addressed where possible in the study design of this thesis, the issue may not be of critical importance since the purpose is to inform the bison and burn study and approximations should be acceptable at this stage.

**Metrics**

The purpose of biomonitoring is to characterize the status of aquatic resources and monitor trends in the condition of biological communities that are associated with disturbance (Resh, Norris & Barbour, 1995). The use of multiple metrics is a common approach to these assessments since a range of metrics can be defined that measure the diverse biological attributes of the community that are thought to signify ecological features of interest. Metrics such as abundance or taxonomic composition represent the structural attributes of benthic macroinvertebrate communities. On the other hand, metrics that define functional attributes of the aquatic ecosystem such as FFG composition are thought to represent food sources and community feeding types, as well as microhabitats and community microhabitat preferences (Feld & Hering, 2007). An advantage to study designs that use a multimetric approach is that statistical tests of significance can be applied so that the resulting decisions about the impacts of disturbance are based on probability rather than solely based on subjective interpretation. This, in combination with selecting metrics that are sensitive enough to distinguish natural variability
from the impacts of disturbance, results in study designs that are robust and scientifically defensible.

There are numerous metrics that have been defined to detect the impacts of disturbance in aquatic ecosystems using benthic macroinvertebrates as biological indicators. Rapid bioassessment techniques proposed by Resh et al. (1995) identified metrics that were accurate in distinguishing unimpaired versus impaired conditions such as taxonomic richness, EPT richness, % scrapers and % shredders (note that this list is not inclusive but highlights the metrics that were also evaluated for this study). Other studies have reported that abundance is one of the more variable components of the biological assemblage and that taxonomic richness and relative abundances are generally more sensitive to ecological disturbance (Carlisle & Clements, 1999; Fore et al., 1996). A more recent review of biomonitoring approaches conducted by Bonada et al. (2006) supported a multiple metric approach but cautioned that structural metrics are not predictive *a priori* whereas functional metrics are. Merrit & Cummins (2008) critiqued the use of diversity indices, focusing on the poor understanding of their biological and ecological meaning and other deficiencies in their application. Finally, Feld and Hering (2007) indicated that sensitive taxa such as EPT decreased with environmental stress whereas many tolerant taxa did not. Since tolerant taxa are generally present even under natural conditions, degrading environments will cause the loss of sensitive taxa rather than a community shift from sensitive to tolerant species. Overall, there is both support and opposition to the use of any metric that could
be used. It is important to select multiple metrics considering both their biological and statistical attributes and how they will discriminate natural variability from the impacts of disturbance.

Since the dataset used to evaluate metrics used flooding as a surrogate disturbance for bison and fire, it is important to understand what the impacts of these disturbances are in benthic macroinvertebrate communities. Flooding is one of the most researched disturbances (Resh et al., 1995) because of its regular occurrence in seasonal fluctuations such as spring freshet and high precipitation events. In general, species abundance, biomass and richness are commonly used to detect impacts of flooding on aquatic ecosystems (Bunn & Arthington, 2002; Reice, 1985; Resh et al., 1995; Stanley, Powers, & Lottif, 2010). During flooding, organisms can be eliminated by substrate mobility or displaced downstream; however, a more diverse substrate will have a variety of microhabitats that can provide refugia during high flow events. Flow refugia help to retain species composition as well as provide recolonization opportunities for areas that were more severely impacted by the effects of a flood. In addition, the short generation times of many benthic macroinvertebrate taxa may contribute to a higher resilience within the community. The aforementioned study conducted by Reice (1985) found that benthic macroinvertebrate communities recovered to near normal population levels within four weeks after flood disturbance.

With respect to bison, there are some localized impacts expected to occur that relate to sediment composition and loss of riparian vegetation. Bison are known to have a high fidelity to specific stream crossing sites resulting in the above impacts at these locations. There is also
widening of the stream channel, which can influence abiotic factors such as velocity, temperature and nutrients. In the two studies reviewed, both found taxonomic richness to be similar between trampled and untrampled areas (Fritz et al., 1999; Meadows, 2001). Meadows (2001) also found differences in diversity and evenness although did not indicate which measure of diversity was calculated. Overall, the general finding was changes to species composition of the benthic macroinvertebrate community as opposed to a reduction in taxonomic richness (Meadows, 2011). For the impacts of fire, it is expected there will be a larger reduction in riparian vegetation, which may result in increased solar inputs, water temperatures and nutrient enrichment. There is also the potential for increased sediment mobilization and associated contaminants to the aquatic ecosystem. However, the study area is in a National Park and there are no known large anthropogenic disturbances upstream; therefore contaminants would be more related to natural background conditions than anthropogenic sources of pollution. In general, bison will have chronic, localized impacts whereas fire will have shorter, temporal impacts.

The only metric that had evidence of a differential change between periods was detected for abundance; however, the benthic invertebrate metric results indicated that benthic macroinvertebrate communities at both the control and impact stations appear to be experiencing similar high quality habitat conditions. Richness, evenness, diversity, EPT taxa as well as the various other metrics calculated were indicative of good environmental conditions. It is probable that the minor differences between periods were a result of natural inter-annual variability in benthic macroinvertebrate communities and highlights the importance of comparing samples
from potentially impacted stations to suitable reference stations. Also, CABIN sampling occurred in the third week of September whereas the flooding ended in August. There is the possibility that the benthic macroinvertebrate community recovered which would be supported by the study conducted by Reice (1985). Finally, CABIN sampling targets riffle habitat because of diverse substrate composition, which create complex microhabitats that support a diverse community of benthic macroinvertebrates. These microhabitats create flow refugia, which provide areas of protection and allows for recolonization of organisms after a flood. This could explain the reduction in abundance at the impact stations but similarity in taxonomic richness between periods.

The benthic macroinvertebrate community was diverse with large contributions of sensitive EPT taxa detected at all stations sampled which is characteristic of fast flowing headwater streams. EPT abundance was similar between the before and after periods, but there was a distinct increase in the number of EPT taxonomic groups in the after period. Chironomids were the next largest contributors for all groups with a large reduction (approximately half) in the chironomid community in the impact stations in the after period. This could be due to the flooding that occurred at the impact sites, which could have flushed finer sediments out of the substrate that are generally more suitable habitat for chironomids. The decrease in fines could result in a greater diversity of microhabitats, which would be colonized by other benthic macroinvertebrates as noted by the general increase in EPT taxa in the impact-after group.
The analysis of the relative composition of FFGs based on five feeding strategies also revealed a diverse community. For the control group, an increase in scrapers and decrease in gathering collectors was observed. This could be related to an increase in algal growth, which would provide food for scrapers but limit the amount of microhabitats for gathering collectors. For the impact group there were two notable differences between the before and after periods. First, the relative composition of scrapers and gathering collectors decreased. This may be because of the flushing and scouring effects of a flood, which would reduce both the fine particulate organic matter and the algal population. Second, there was an increase in the relative composition of shredders, which was also at the significance level for the BACI contrast. It would be expected that the coarse particulate organic matter (CPOM) would also be flushed during a flood event; however, sampling was conducted in the fall when the leaves are falling off the trees. The fact that sample collection was three weeks post-flooding and the timing of sampling in the fall, may have given the shredder community a boost in population because of large inputs of CPOM.

**Power**

The ability to detect an environmental disturbance depends very much on the techniques and procedures adopted (Resh et al, 1995). Although extremely important, power is often overlooked in the study design of biomonitoring programs. Power is inversely related to the probability of making a type II error that is concluding that there is no impact of a disturbance when in fact there is (Fariweather, 1991; Stewart-Oaten & Bence, 2001; Underwood, 1991). This
is critical in environmental management to ensure that unnecessary degradation of the environment does not occur. Power is dependent on the type I error rate, sample size, the effect size to be detected, and the characteristic variability of the data. The integration of power analysis into study design allows environmental managers to critically assess the feasibility of a monitoring program. It is not desirable to proceed with a program with low power whereas the most efficient use of resources, such as the number of samples to be collected, can be determined for a design that has high statistical power, or for which statistical power has been quantified.

Since the bison and burn program has the unique opportunity to collect pre-disturbance data, a BACI study design is being used which measures both control and impact stations before and after a disturbance to determine whether there is an effect. One of the most difficult aspects of power analysis is the determination of effect size. This is especially relevant to benthic macroinvertebrates when used as biological indicators, because of the variability that is inherent in these communities, resulting in difficulties distinguishing between the effects of an impact and background noise. The statistical methods used here calculated the BACI contrast, which helps to define the important biological effect to be detected, by calculating the differential change between the before and impact groups for the 2011 and 2012 CABIN monitoring data. In addition, the BACI contrast provides the direction of change for the metric being assessed. It is generally thought that the effects of disturbance will cause a reduction in the majority of metrics such as abundance, richness and diversity (Resh et al., 1995). This is not always the case as found in the metrics assessed for this study (see Table 3). This reinforces the need for both the
biological and statistical considerations when determining how metrics might change as a result of a disturbance, and which suite of metrics to choose.

The role of power analysis in study design is to determine which response variables (i.e., metrics) will be the most sensitive in detecting an environmental disturbance. Power does not necessarily provide any additional understanding about nature, but allows for some quality control over the inferences derived from environmental monitoring (Osenberg, Schmitt, Holbrook, Abu-Saba, & Flegal, 1994). The final objective of this thesis was to evaluate the power of the bison and burn study using CABIN monitoring data from 2011 and 2012. Since the first year of data was collected in fall of 2014, the specific attributes of the bison and fire study design was incorporated into the power analysis such as number of replicate samples and control/impact watersheds. The use of a BACI approach helped to deal with the lack of spatial and temporal replication as well as lack of randomization (Osenberg et al., 1994). This was partially accounted for in the linear mixed models, which provided the estimates of variance. Overall, this thesis helped to develop the R programming code (following the methods provided by Schwartz (2014)), which can be used for the analysis of the bison and burn study. Power was just one of the tools used in evaluating the study design for the bison and fire program.

The results of the power analysis showed that many of the metrics did not have the sensitivity of detecting the impacts of disturbance after 15 monitoring years of data collection (e.g. Shannon diversity, Shannon evenness, true diversity, EPT metrics, chironomid metrics, % filtering collectors and % predators). Richness, % scrapers, and % shredders had the highest
power and would be close to 80% after two years. Other metrics that achieved relatively high power were abundance, true diversity and % gathering collectors.

The biological importance of abundance is that it is representative of the amount of fish food available, which relates to the productivity of a fishery. The abundance of fish food is much easier for stakeholders to understand compared to a metric such as Shannon diversity or evenness. Abundance is also important from a communications perspective because the general public can understand the implications of a decline in fish food. Part of Parks Canada’s mandate is to communicate these findings to park visitors and the general public. In addition, the metric richness is relevant as it simply represents the total number of species present in an ecological community. It is generally understood that diversity is linked to ecosystem function (Bonada et al., 2006) and therefore, there is value to society in conserving maximal numbers of species. The other metrics that demonstrated the power to detect an impact (i.e. %gathering collectors, %scrapers, %shredders) would allow a conservation practitioner to examine the change in ecological function following a disturbance based on functional feeding groups. For example, a change in %scraper could be correlated to a change in algal abundance. This may provide justification to add a measure of algal abundance to the biomonitoring program.

In general, the results of the power analysis were similar with the metrics identified by Resh et al. (1995) for detecting impaired conditions using rapid bioassessment techniques. Their recommended metrics included species richness, EPT richness, the biotic index (not considered here), and percentage of scrapers functional group (Resh et al., 1995). It is interesting to note that
there is general support from all the literature reviewed for the power of richness metrics to detect the impacts of disturbance.

For the low performing metrics (e.g. Shannon diversity, Shannon eveness, true diversity, EPT metrics, chironomid metrics, % filtering collectors and % predators), low power may be attributable to large variability in the benthic macroinvertebrate community, which is inherent to these types of biological datasets. It is generally recognized that the dataset used to conduct the power analysis was not perfect and that improvements in the bison and fire study design will help to increase power. It may be prudent to continue to analyze Shannon diversity and eveness in the bison and burn study since the research conducted by Meadows (2011) found significant differences between bison trampled and untrampled sites. In addition, Parks Canada is mandated to manage for ecological integrity, which includes the diversity of ecosystems. This results in the importance of communicating on measures of diversity for aquatic ecosystem management.

The metrics were assessed for their sensitivity to detect the impacts of disturbance, and then analyzed for power to determine the number of monitoring years it would take to detect an effect. The value in using a multi-metric approach is that different metrics will vary in their sensitivity to the impacts of a disturbance. A benefit to studying disturbance in a National Park is that an effect may be easier to detect in regions that have natural systems rather than in regions that are significantly altered or have multiple anthropogenic influences. For example, control sites are relatively easy to identify in a National Park.
CHAPTER 4 — CONCLUSION/RECOMMENDATIONS

It was no surprise that the variability in the metric results was as variable as the benthic macroinvertebrate communities themselves. Overall, the power analysis reveals three metrics that will be the most sensitive to detecting disturbance (abundance, richness and % shredder). These metrics have the possibility of detecting an effect in two to five years of monitoring, which is desirable to ensure that the appropriate management actions can be implemented if impacts in the benthic macroinvertebrate community are apparent. In addition, these metrics have been proven to be biologically important to fisheries and general aquatic ecosystem integrity.

For the 2011 and 2012 annual CABIN monitoring data, a simple BACI approach was used since there was only one year of before and after data. Some limitations to the analysis for this dataset were the differences between the two periods could be the result of year effects (i.e, before = wet year, after = dry year), the use of 2012 data as the before period and the 2011 data as the after period in the BACI analysis, and there was missing values in the dataset. In addition, an assumption in the power analysis is that the standard deviation will remain in effect as the study progresses. This results in increased variability in the analysis and not enough data to determine if the general trends for impact and control sites between the before and after periods will be consistent for the future. The concerns noted above are common realities in the field of environmental management. It would be rare to have a perfect data set to inform a power analysis; however, the results of this power analysis provided a rough approximation of what
metrics would be useful in detecting the effect of a disturbance and is also necessary step when conducting good research.

Although there were limitations in the dataset analyzed for this thesis, there are aspects of the Parks Canada bison and burn study design that will help to reduce variability, for example the number of samples (eight samples per watershed), resampling in the same watersheds, two to three pre-disturbance data, and long term monitoring. This thesis only evaluated the benthic macroinvertebrate community data but there are numerous environmental attributes that are collected during the CABI\(\text{N}\) assessments. For the bison and burn program, it is recommended that some of the most relevant habitat variables be included in the overall analysis. This would include flow and substrate composition, which are known to have a large influence on benthic macroinvertebrate communities. The addition of these factors as covariates in an analysis of covariance or random effects in a linear mixed model should help to account for the variability that was observed in the 2011 and 2012 CABI\(\text{N}\) monitoring data and therefore lead to higher power in an analysis that includes habitat covariates.

Since the data are being collected using CABI\(\text{N}\) protocols, it is also recommended that the CABI\(\text{N}\) data be analyzed using the multivariate approach RCA. RCA is a nationally accepted program headed by Environment Canada and used by Parks Canada for their annual CABI\(\text{N}\) programs. Essentially, CABI\(\text{N}\) assesses the biological and environmental attributes collected at each sampling station (i.e., test site), and uses predictive modeling to determine a station’s biological condition (i.e. multivariate approach; Environment Canada, 2012). The relationship
between environmental characteristics and invertebrate communities expected at reference sites is used to predict communities expected at test sites integrating the environmental characteristics of those sites. The advantage of RCA is that annual CABIN sampling is conducted in all of the Mountain National Parks and uploaded to the CABIN database each year. Effort would be needed to create the reference condition model and it would have to be updated somewhat frequently (e.g., 5 years); however, once built the model would be available for all biomonitoring programs in the mountain parks.
REFERENCES


http://www.ces.iisc.ernet.in/energy/water/paper/cistup_TR1/DEVELOPMENTS%20IN%20AQUATIC%20INSECT%20BIOMONITORING.pdf


http://www.esajournals.org.ezproxy.royalroads.ca/doi/pdf/10.1890/1051-0761%282001%29011%5B0783%3ABBOOES%5D2.0.CO%3B2


http://www.publish.csiro.au/paper/MF9910555


http://www.researchgate.net/profile/Christian_Feld/publication/227668579_Community_structure_or_function_effects_of_environmental_stress_on_benthic_macroinvertebrates_at_differentSpatial_scales/links/00463516c2d1e30e98000000.pdf


*Ecological Monographs, 54*(2), 187-211.


[http://dx.doi.org/10.1080/02705060.2001.9663791](http://dx.doi.org/10.1080/02705060.2001.9663791)


[http://dx.doi.org/10.1080/00288330.1990.9516433](http://dx.doi.org/10.1080/00288330.1990.9516433)


http://download.springer.com/static/pdf/718/art%253A10.1023%252FA%25253A1017078401734.pdf?auth66=1402774265_0a2d8a001c80812d05d213831e7c2e94&ext=.pdf


APPENDIX

A. RStudio Output File
The following is the output from the computer program RStudio that was used for all statistical analysis. The output includes both the code for each command whether it was graphing or hypothesis testing and then the result from RStudio. Note that all results in RStudio are preceded by ##.

abundance.R

jesspenno
Sat Aug 29 12:55:12 2015

options(useFancyQuotes=FALSE) # renders summary output corrects

library(car)
library(ggplot2)
library(lsmeans)

## Loading required package: estimability

library(plyr)
library(lmerTest)

## Loading required package: Matrix
## Loading required package: lme4
## Loading required package: Rcpp
##
## Attaching package: 'lmerTest'
##
## The following object is masked from 'package:lme4':
##
## lmer
##
## The following object is masked from 'package:lsmeans':
##
## lsmeans
##
## The following object is masked from 'package:stats':
##
## step
source('http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms/schwarz.functions.r')

##
## Attaching package: 'devtools'
##
## The following object is masked from 'package:lsmeans':
##
## test
##
## Loading required package: grid
##
## Loading required package: scales
##
## Loading required package: proto

source('http://www.stat.sfu.ca/~cschwarz/Stat-650/Notes/MyPrograms/BACI/baci-power.r')

##read in schwartz quotes and baci power function

# Read in the actual data and define the factors of interest

cat(" BACI design measuring invertbrate metrics \n\n")

## BACI design measuring invertbrate metrics

bnp <- read.table("/Users/jesspenno/Desktop/Thesis/Working R Driectory/abundance.txt", head
er=TRUE, as.is=TRUE, strip.white=TRUE)
bnp$trt <- interaction(bnp$SiteClass, bnp$Period)
bnp$SiteClass <- factor(bnp$SiteClass)
bnp$Period <- factor(bnp$Period, levels=c("Before","After"), ordered=TRUE) # sort correctly
bnp$trt <- factor(bnp$trt)
cat("Listing of part of the raw data \n")

## Listing of part of the raw data

bnp[1:10,]

## Site  Station SiteClass Period Abundance  trt
## 1  C1  B-B-11-1 Control After  332 Control.After
## 2  C1  B-B-11-2 Control After  330 Control.After
## 3  C1  BBrewsterCr Control After  282 Control.After
## 4  C1  BCarrotCr Control After  301 Control.After
## 5  C1  BCascadeLow Control After  299 Control.After
## 6  C1  BFlintsPk Control After  330 Control.After
## 7  C1  BFortyMileCr Control After  296 Control.After
## 8  C1  BHealyCr Control After  350 Control.After
BENTHIC MACROINVERTEBRATE METRICS POWER

### 9 I1 BSprayR01 Impact After 306 Impact.Aft
### 10 I1 BSprayR02 Impact After 295 Impact.Aft

# 2011 is Before, and 2012 is After impact
`xtabs(~Station+Period, data=bnp) # CJS`

<table>
<thead>
<tr>
<th>Station</th>
<th>Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-B-11-1</td>
<td>1</td>
</tr>
<tr>
<td>B-B-11-2</td>
<td>1</td>
</tr>
<tr>
<td>BAltritruedeCr</td>
<td>1</td>
</tr>
<tr>
<td>BBrewsterCr</td>
<td>0</td>
</tr>
<tr>
<td>BCarrotCr</td>
<td>1</td>
</tr>
<tr>
<td>BCascadeLow</td>
<td>1</td>
</tr>
<tr>
<td>BFlictsPk</td>
<td>0</td>
</tr>
<tr>
<td>BFortyMileCr</td>
<td>0</td>
</tr>
<tr>
<td>BHealyCr</td>
<td>0</td>
</tr>
<tr>
<td>BJohnson01</td>
<td>1</td>
</tr>
<tr>
<td>BJJohnston01</td>
<td>0</td>
</tr>
<tr>
<td>BSprayR01</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR02</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR03</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR04</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR05</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR06</td>
<td>1</td>
</tr>
</tbody>
</table>

# Preliminary plot
`prelimplot <- ggplot(data=bnp, aes(x=trt, y=Abundance))+
ggtitle("Preliminary plot to look for outliers etc")+
geom_point(position=position_jitter(w=0.1))+
geom_boxplot(alpha=0.1)`

prelimplot
# Get some simple summary statistics

```r
report <- ddply(bnp, c("Period","SiteClass"), function(x){
  res <- sf.simple.summary(x, "Abundance", crd=TRUE)
  return(res)
})
cat("\n\nSummary report \n")
```
# Draw a profile plot including the raw data

```r
profileplot <- ggplot(data=report, aes(x=Period, y=mean, group=SiteClass, color=SiteClass))+
ggtitle("Profile Plot of Invertebrate Abundance")+
ylab("Abundance with Mean and 95% CI")+
geom_point(position=position_dodge(w=0.1))+
geom_line(position=position_dodge(w=0.1)) +
geom_errorbar(aes(ymax=ucl, ymin=lcl), width=0.2, position=position_dodge(w=0.1)) +
geom_point(data=bnp, aes(y=Abundance), position=position_dodge(w=0.2))
```

```r
profileplot
```

## ymax not defined: adjusting position using y instead

### ymax not defined: adjusting position using y instead

### ymax not defined: adjusting position using y instead

```r
cat("\n\nUse the Type III tests from the Anova() function from the car package")
```

###

### Use the Type III tests from the Anova() function from the car package

```r
cat("\nbut you need to set the treatment contrasts to sum rather than treatment")
```

###

### but you need to set the treatment contrasts to sum rather than treatment

```r
cat(" \nSee http://r.789695.n4.nabble.com/Type-I-v-s-Type-III-Sum-Of-Squares-in-ANOVA-td1573657.html")
```
BENTHIC MACROINVERTEBRATE METRICS POWER

```r
library(car)
result.lm <- lm( Abundance ~ SiteClass + Period + SiteClass:Period, data=bnp, contrasts=c(SiteClass="contr.sum", Period="contr.sum")) # CJS
Anova(result.lm,type=3)

```
## Loading required package: gridExtra
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to change the smoothing method.

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : pseudoinverse used at 297.28

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : neighborhood radius 22.217

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : reciprocal condition number 1.6115e-16

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : There are other near singularities as well. 687.36

## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to change the smoothing method.

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : pseudoinverse used at 297.28

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : neighborhood radius 22.217

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : reciprocal condition number 1.6115e-16

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : There are other near singularities as well. 687.36

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : pseudoinverse used at 0.12479

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : neighborhood radius 0.041875

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : reciprocal condition number 4.7336e-17

## Warning in simpleLoess(y, x, w, span, degree, parametric, drop.square, normalize, : There are other near singularities as well. 0.0017535
BENTHIC MACROINVERTEBRATE METRICS POWER

diagplot

**Residuals vs Fitted**

- Residual vs Fitted Values
  - Residual
  - Fitted Values

**Normal Q-Q**

- Standardized Residuals
  - Theoretical Quantiles

**Scale-Location**

- Plot of Standardized Residuals
  - Fitted Values

- Plot of Standardized Residuals
  - Leverage

```r
cat("\n\nEstimated marginal means \n\n")

##
## Estimated marginal means

# Estimate the marginal means and the various effects
cat("\n\nEstimated marginal means for SiteClass \n\n")

##
## Estimated marginal means for SiteClass

result.lsmo.S <- lsmeans::lsmeans(result.lm, ~SiteClass)

## NOTE: Results may be misleading due to involvement in interactions

summary(result.lsmo.S)

## SiteClass lsmean SE df lower.CL upper.CL
## Control    306.25 10.52821  22 284.4158 328.0842
## Impact    330.25 11.25513  22 306.9083 353.5917
##
## Results are averaged over the levels of: Period
## Confidence level used: 0.95
cat("Estimated marginal means for Period effects")

##
## Estimated marginal means for Period effects

summary(result.lsmo.S)

## SiteClass lsmean    SE df lower.CL upper.CL
## Control     306.25 10.52821  22 284.4158  328.0842
## Impact      330.25 11.25513  22 306.9083  353.5917

## Results are averaged over the levels of: Period
## Confidence level used: 0.95

result.lsmo.P <- lsmeans::lsmeans(result.lm, ~Period)

## NOTE: Results may be misleading due to involvement in interactions

cat("Estimated marginal means for SitePeriod effects")

##
## Estimated marginal means for SitePeriod effects

result.lsmo.SP <- lsmeans::lsmeans(result.lm, ~SiteClass:Period)

summary(result.lsmo.SP)

## SiteClass Period lsmean    SE df lower.CL upper.CL
## Control Before   297.5 15.91716  22 264.4898  330.5102
## Impact  Before   341.0 15.91716  22 307.9898  374.0102
## Control  After  315.0 13.78467  22 286.4124  343.5876
## Impact  After   319.5 15.91716  22 286.4898  352.5102

## Confidence level used: 0.95

## Estimate the BACI contrast along with a se

cat("Estimated BACI contrast")

##
## Estimated BACI contrast

contrast(result.lsmo.SP, list(baci=c(1,-1,-1,1)))

## contrast estimate    SE df t.ratio  p.value
## baci                -39 30.82345  22  -1.265   0.2190
confint(contrast(result.lsmo.SP, list(baci=c(1,-1,-1,1))))

## contrast estimate  SE  df  lower.CL  upper.CL
## baci -39 30.82345 22 -102.9239 24.92393

## Confidence level used: 0.95

sigmaHat(result.lm) # estimated standard deviation about the regression line from car package

## [1] 38.98893

summary(result.lm)$sigma # same using standard methods

## [1] 38.98893

# estimate average standard deviations within site-year combination
mean(report$sds) # gives very similar values?

## [1] 38.32099

### Are there large station effects?

# Draw a profile plot of the stations including the raw data
tabs(~Station+Period+SiteClass, data=bnp) # CJS

###, SiteClass = Control

##
## Station  Period
## B-B-11-1  Before After
## B-B-11-2  1  1
## BALtritruCr  1  0
## BBrewsterCr  0  1
## BCarrotCr  0  1
## BCascadeLow  1  1
## BFlintsPk  0  1
## BFoargeCr  0  1
## BHealyCr  0  1
## BJohnson01  1  0
## BJohnston01  1  0
## BSpalryR01  0  0
## BSpalryR02  0  0
## BSpalryR03  0  0
## BSpalryR04  0  0
## BSpalryR05  0  0
## BSpalryR06  0  0
### SiteClass = Impact

### Period

<table>
<thead>
<tr>
<th>Station</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-B-11-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B-B-11-2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BAtritrudeCr</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BBrewsterCr</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BCarrotCr</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BCascadeLow</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BFintsPk</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BFortyMileCr</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BHealyCr</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BJohnson01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BJohnston01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BSprayR01</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR02</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR03</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR04</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR05</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>BSprayR06</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

```r
stationplot <- ggplot(data=bnp, aes(x=Period, y=Abundance, group=Station, color=SiteClass)) +
ggtitle("Profile Plot of Invertebrate Abundance at Individual Stations") +
ylab("Abundance with Mean and 95% CI") +
geom_point(position=position_dodge(w=0.1)) +
geom_line(position=position_dodge(w=0.1))
```

stationplot

### ymax not defined: adjusting position using y instead
### ymax not defined: adjusting position using y instead
library(lmerTest)
bnp$StationF <- factor(bnp$Station)
bnp$PeriodF <- factor(bnp$Period)
result.lm2 <- lmerTest::lmer(Abundance ~ SiteClass + PeriodF + SiteClass:PeriodF +
                           (1|StationF)+(1|StationF:PeriodF), data=bnp,
                           control=lmerControl(
                             check.nobs.vs.rankZ = "ignore",
                             check.nobs.vs.nlev = "ignore",
                             check.nlev.gtreq.5 = "ignore",
                             check.nlev.gtr.1   = "ignore",
                             check.nobs.vs.nRE  = "ignore"))

result.lm3 <- lme4::lmer(Abundance ~ SiteClass + PeriodF + SiteClass:PeriodF +
                           (1|StationF)+(1|StationF:PeriodF), data=bnp,
                           control=lmerControl(
                             check.nobs.vs.rankZ = "ignore",
                             check.nobs.vs.nlev = "ignore",
                             check.nlev.gtreq.5 = "ignore",
                             check.nlev.gtr.1   = "ignore",
                             check.nobs.vs.nRE  = "ignore"))

result.lm3
Linear mixed model fit by REML ['lmerMod']

Formula: Abundance ~ SiteClass + PeriodF + SiteClass:PeriodF + (1 | StationF:PeriodF)

Data: bnp

REML criterion at convergence: 232.4585

Random effects:

Groups Name Std.Dev.
StationF:PeriodF (Intercept) 25.28
StationF (Intercept) 0.00
Residual 29.68

Number of obs: 26, groups: StationF:PeriodF, 26; StationF, 17

Fixed Effects:

(Intercept) SiteClassImpact
306.25 24.00
PeriodF:L SiteClassImpact:PeriodF:LB
12.37 -27.58

anova(result.lm3, ddf='Kenward-Roger')

## Analysis of Variance Table

Df Sum Sq Mean Sq F value
SiteClass 1 1938.13 1938.13 2.2000
PeriodF 1 1.83 1.83 0.0021
SiteClass:PeriodF 1 1410.37 1410.37 1.6009

VarCorr(result.lm2)

## Groups Name Std.Dev.
StationF:PeriodF (Intercept) 25.281
StationF (Intercept) 0.000
Residual 29.681

# Power for 1,2,5,10,15 years

### Power from mixed model

```r
results <- NULL
results <- rbind(results, baci.power(n_TA=8, n_TB=8, n_CA=8, n_CB=8,
ns_T=1, ns_C=2,
ny_B=1, ny_A=1,
mu_TA=302, mu_TB=341, mu_CA=315, mu_CB=298,
sdYear=0, sdSite=0, sdSiteYear=25.281, sdResid=29.681))
```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue
### Rescale variables?

```r
results<-rbind(results,baci.power(n_TA=8, n_TB=8, n_CA=8, n_CB=8,
                                 ns_T=1, ns_C=2,
                                 ny_B=1, ny_A=2,
                                 mu_TA=302, mu_TB=341, mu_CA=315, mu_CB=298,
                                 sdYear=0, sdSite=0, sdSiteYear=25.281, sdResid=29.681))
```

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue
### Rescale variables?

```r
results<-rbind(results,baci.power(n_TA=8, n_TB=8, n_CA=8, n_CB=8,
                                 ns_T=1, ns_C=2,
                                 ny_B=1, ny_A=5,
                                 mu_TA=302, mu_TB=341, mu_CA=315, mu_CB=298,
                                 sdYear=0, sdSite=0, sdSiteYear=25.281, sdResid=29.681))
```

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue
### Rescale variables?

```r
results<-rbind(results,baci.power(n_TA=8, n_TB=8, n_CA=8, n_CB=8,
                                 ns_T=1, ns_C=2,
                                 ny_B=1, ny_A=10,
                                 mu_TA=302, mu_TB=341, mu_CA=315, mu_CB=298,
                                 sdYear=0, sdSite=0, sdSiteYear=25.281, sdResid=29.681))
```

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unidentifiable: very large eigenvalue
### Rescale variables?

```r
results<-rbind(results,baci.power(n_TA=8, n_TB=8, n_CA=8, n_CB=8,
                                 ns_T=1, ns_C=2,
                                 ny_B=1, ny_A=15,
                                 mu_TA=302, mu_TB=341, mu_CA=315, mu_CB=298,
                                 sdYear=0, sdSite=0, sdSiteYear=25.281, sdResid=29.681))
```

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model failed to converge: degenerate Hessian with 1 negative eigenvalues
### Rescale variables?

```r
results
```
---

### Benthic Macroinvertebrate Metrics Power

<table>
<thead>
<tr>
<th>alpha</th>
<th>sdSite</th>
<th>sdYear</th>
<th>sdResid</th>
<th>n_TA</th>
<th>n_TB</th>
<th>n_CA</th>
<th>n_CB</th>
<th>ns_T</th>
<th>ns_C</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>[2,]</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>[3,]</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>[4,]</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td>[5,]</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>

### ny_B ny_A mu_TA mu_TB mu_CA mu_CB baci dfdenom ncp Fcrit

| [1,] | 1    | 1    | 302  | 341  | 315  | 298  | -56  | 1    | 1.395174 | 161.4476 |
| [2,] | 1    | 2    | 302  | 341  | 315  | 298  | -56  | 2    | 1.860233 | 10.12796 |
| [3,] | 1    | 5    | 302  | 341  | 315  | 298  | -56  | 3    | 2.325291 | 5.117355 |
| [4,] | 1    | 10   | 302  | 341  | 315  | 298  | -56  | 9    | 2.536681 | 4.38075 |
| [5,] | 1    | 15   | 302  | 341  | 315  | 298  | -56  | 9    | 2.615952 | 4.182964 |

### power Tcrit os.power1 os.power2

| [1,] | 0.08115222 | 6.313752 | 0.1539334 | 0.00733773 |
| [2,] | 0.16366   | 2.353363 | 0.2821867 | 0.002782294 |
| [3,] | 0.2762802 | 1.833113 | 0.4070705 | 0.001089525 |
| [4,] | 0.3275282 | 1.729133 | 0.4565482 | 0.0007284832 |
| [5,] | 0.3462493 | 1.699127 | 0.4739557 | 0.000628731 |