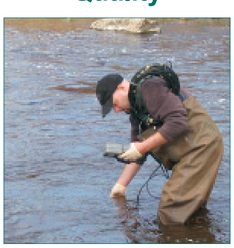
RAMP

Regional Aquatics Monitoring Program

Climate and Hydrology

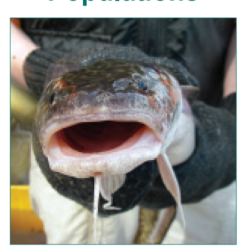
Water Quality



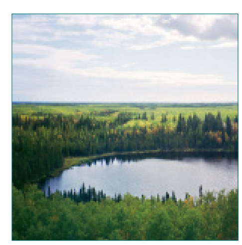
Sediment Quality

Benthic Invertebrates

Fish Populations



Acid-Sensitive Lakes



Scientific Peer Review of RAMP: Response to Panel Comments and Recommendations

June 2011

Regional Aquatics Monitoring Program

SCIENTIFIC PEER REVIEW OF RAMP:

RESPONSE TO PANEL COMMENTS AND RECOMMENDATIONS

FINAL

Prepared for:

RAMP SCIENTIFIC PEER REVIEW PANEL

Prepared by:

RAMP TECHNICAL PROGRAM COMMITTEE
AND IMPLEMENTATION TEAM

JUNE 2011

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

LIST OF ACRONYMS

ADL	Analytical Detection Limit
AENV	Alberta Environment
ALS	ALS Environmental
ALPAC	Alberta-Pacific Forest Industries Inc.
ANC	Acid Neutralizing Capacity
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
AOSERP	Alberta Oil Sands Environmental Research Program
AITF	Alberta Innovates Technology Futures
ASL	Acid-Sensitive Lakes
ASRD	Alberta Sustainable Resource Development
ATI	Assemblage Tolerance Index
CCME	Canadian Council of Ministers of the Environment
CEA	Cumulative Effects Assessment
CEMA	Cumulative Environmental Management Association
CL	Critical Load
CONRAD	Canadian Oilsands Network for Research and Development
CPUE	Catch-Per-Unit-Effort
CV	Coefficient of Variation
DFO	Department of Fisheries and Oceans
DNA	Deoxyribonucleic Acid
DOC	Dissolved Organic Carbon
EC	Environment Canada
EEM	Environmental Effects Monitoring
EDA	Exploratory Data Analysis
EIA	Environmental Impact Assessment
EMAP	Environmental Monitoring and Assessment Program
EPEA	Environmental Protection and Enhancement Act
EPT	Ephemeroptera, Plecoptera and Trichoptera
EROD	Ethoxyresorufin-O-deethylase
FAM	Fish Assemblage Monitoring

FSA	Focus Study Area
FT-IRMS	Fourier Transform Infrared Microscopy
FWMIS	Fish and Wildlife Management Information System
GLM	General Linear Model
GOA	Government of Alberta
GSI	Gonadosomatic Index
HRI	Hydrologic Regime Indicator
HSPF	Hydrological Simulation Program—Fortran
IBI	Index of Biotic Integrity
IMB	Isotopic Mass Balance
ISQG	Interim Sediment Quality Guideline
KIR	Key Indicator Resource
LARP	Lower Athabasca Regional Plan
LSI	Liversomatic Index
LTRN	Long-Term Regional Network
MDL	Method Detection Limit
MFO	Mixed-function Oxygenase
NO_xSO_x	Emissions of nitrogen oxides (NO_x) and sulphur oxides (SO_x)
OSPW	Oil Sands Process Water
OSRIN	Oil Sands Research and Information Network
OSTWAEO	Oil Sands Tailings Water Acid-Extractable Organics
PAH	Polycyclic Aromatic Hydrocarbon
PAI	Potential Acid Input
QA	Quality Assurance
QC	Quality Control
RAMP	Regional Aquatics Monitoring Program
RCA	Reference Condition Approach
RMCC	Research and Monitoring Coordinating Committee
SOP	Standard Operating Procedure
SSWC	Steady-State Water Chemistry
SSWQO	Site-Specific Water Quality Objectives
TDS	Total Dissolved Solids
TEK	Traditional Ecological Knowledge

TIE	Toxicity Identification Evaluation
TKN	Total Kjeldahl Nitrogen
TN	Total Nitrogen
TP	Total Phosphorus
TOC	Total Organic Carbon
TRH	Total Recoverable Hydrocarbon
TSS	Total Suspended Solids
USEPA	United States Environmental Protection Agency
VMV	Variable Method Values
WBEA	Wood Buffalo Environmental Association
WQI	Water Quality Index
WSC	Water Survey of Canada
WY	Water Year
YOY	Young-of-Year

1.0 INTRODUCTION

1.1 BACKGROUND

The Regional Aquatics Monitoring Program (RAMP) was initiated in 1997 in association with mining development in the Athabasca oil sands region near Fort McMurray, Alberta. RAMP is a multi-stakeholder, industry-funded initiative that undertakes long-term monitoring of rivers and lakes in the region.

The focus of monitoring is on surface waters with specific emphasis on key components of boreal aquatic ecosystems, including:

- Climate and hydrology – monitors changes in the water level of selected lakes and in the quantity of water flowing through rivers and creeks in the Athabasca oil sands area;

- Water quality in rivers, lakes and the delta – reflects potential exposure of fish and invertebrates to organic and inorganic chemicals;

- Benthic invertebrate communities and sediment quality in rivers, lakes and the delta – reflect habitat quality, serve as biological indicators, and are important components of fish habitat;

- Fish populations in rivers and lakes – biological indicators of ecosystem integrity and are a highly valued resource in the region; and

- Water quality in regional lakes sensitive to acidification – early warning indicator of potential effects related to acid deposition.

The mandate of RAMP is to determine, evaluate and communicate the state of the aquatic environment and any changes that may result from cumulative resource development within the Regional Municipality of Wood Buffalo. The information collected by RAMP is, in turn, provided to the environmental regulatory authorities in support of ongoing assessment and the provision of various monitoring requirements stipulated in Alberta EPEA approvals for oil sands projects.

Given the size, scope and importance of RAMP, a scientific peer review of the program is conducted periodically to solicit external feedback on the program, including recommended adjustments necessary for the program's success and ongoing improvement. The first peer review was completed in 2004, which resulted in numerous refinements to the monitoring program. Since that time, the program has continued to grow in response to rapidly increasing oil sands development and a second peer review was undertaken in 2010.

1.2 2010 SCIENTIFIC PEER REVIEW

The overall goal of the 2010 scientific peer review of RAMP was to evaluate whether the program successfully answered the following questions (AITF 2011):

1. Can the present program detect changes if they occur?

2. Can the source of any potential changes be identified by the present program?

3. Are the appropriate questions being asked by RAMP and are the appropriate criteria being monitored to answer those questions?

An independent coordinator was hired to manage the process and select the review panel. The review panel included scientists from academia and government based on their expertise in the monitoring disciplines of RAMP (Table 1).

Table 1 Scientists selected to undertake the 2010 peer review of RAMP.

RAMP Monitoring Discipline	Reviewer	Organization
Climate and Hydrology	Dr. Donald Burn Dr. John Gibson	University of Waterloo University of Victoria
Water Quality	Dr. George Dixon Dr. Monique Dubé	University of Waterloo University of Saskatchewan
Benthic Invertebrate Communities and Sediment Quality	Dr. Kelly Munkittrick Dr. Joseph Flotemersch	University of New Brunswick (Canadian Rivers Institute) US EPA
Fish Populations	Dr. John Post Dr. William Franzin	University of Calgary Laughing Water Arts & Science Inc. (formerly with DFO)
Acid-Sensitive Lakes	Dr. Shaun Watmough Dr. John Gibson Dr. George Dixon	Trent University University of Victoria University of Waterloo

Reviewers were asked to conduct a detailed review within their area of expertise, as well as comment on the monitoring program as a whole. The review focused on the following documents:

- RAMP 2009 Technical Report (RAMP 2010); and
- RAMP Technical Design and Rationale document (RAMP 2009b).

In addition, other RAMP-related information was made available as required, including technical reports from other years, the 2004 peer review report (Ayles *et al.*, 2004), a copy of the RAMP terms of reference, a review of the fish monitoring program by Hughes and Whittier (2008), as well as access to the RAMP database (http://www.ramp-alberta.org/ramp/data.aspx).

Following the completion of their individual reviews of the program, and a meeting to compare and synthesize results, the reviewers and the coordinator developed their final report, which was released on January 31, 2011 (AITF 2011). A key element of the report was the provision of comments and recommendations by the panel members related to the scientific design and implementation of an aquatic monitoring program for the oil sands region.

1.3 SCOPE OF RAMP'S RESPONSE

The objective of the current report is to document RAMP's response and follow-up action specific to comments and recommendations provided by the peer review panel. RAMP's ability to make large-scale changes to the program at this time is strongly influenced by initiatives currently underway by the federal and provincial governments to evaluate and redesign environmental monitoring in the Athabasca oil sands region. Specifically, the federal government is developing a surface water quality monitoring plan for the lower Athabasca River and tributaries, with the intent to expand the plan to other environmental media and ultimately ensure the media-specific plans are integrated into a single, holistic, ecosystem-based approach. In addition, the provincial government has set up a provincial environmental monitoring panel to provide recommendations on developing a world-class environmental monitoring, evaluation, and reporting system for Alberta's oil sands (http://www.environment.alberta.ca/03289.html). Given these ongoing activities, RAMP was advised by Alberta Environment to defer plans to make major changes to the program until after the above initiatives were largely complete (see Appendix A1). Alberta Environment members of RAMP have further suggested that RAMP focus on recommendations specific to refining the scientific elements of the program, and refrain from addressing issues of governance and structure (including modifications to the current mandate and scope of the program) at this time. Accordingly, responses to the peer review comments and recommendations provided in this document were developed within these boundaries.

2.0 RESPONSES TO PANEL COMMENTS AND RECOMMENDATIONS

Comments and recommendations from the members of the peer review panel and associated responses have been organized in this document to address general comments provided in the main body of the peer review report, as well as detailed comments specific to each monitoring discipline and reviewer as outlined in the report and supporting appendices.

Reviewer's comments and recommendations are shown in italics, along with the page and paragraph number they were found in the peer review report, followed by the response by RAMP. Responses were developed by the RAMP Technical Program Committee (RAMP Tech) and the Implementation Team during a workshop held on April 6 and 7, 2011. Where possible, RAMP Tech indicated when a recommendation or comment was already being addressed by RAMP or could be implemented in the program. Some panel requests were addressed in the 2010 Technical Report released on May 4, 2011 (RAMP 2011, e.g., suggested statistical analyses, additional written context, etc.). In other cases, recommendations and comments are being incorporated into the ongoing 2011 program or planned for the 2012 program. Table 2 and Table 3 provide a list of the recommendations that were addressed in the 2010 report (RAMP 2011) and the recommendations that will be addressed during the 2011 monitoring program, respectively. Additional recommendations will be implemented at a later date, subject to further review or addressed based on direction received from the government monitoring initiatives. Those recommendations subject to further review or dependent on the direction provided by the government monitoring initiatives are summarized in Table 4.

Table 2 Components of the RAMP peer review that were addressed in the 2010 Technical Report (RAMP 2011).

Comment/Recommendation	Reference	Response
Use individual fish analysis rather than means in analyses of metals in fish tissues.	Sec. 3.4	p. 5-397 to 5-404, 5-462 to 5-464
Include all available historical pre-RAMP [fish] data.	Sec. 3.4	p. 5-78 to 5-96
Use appropriate statistical methods to both decrease variability in sampling results and associated noise in the data (benthos and water quality].	Sec. 3.3	p. 6-8 to 6-24, 6-32 to 6-41
Clarify the method and application for use of the Water Quality Index.	Sec. 3.2	p. 3-75
Use of the steady state critical load component [for acid-sensitive lakes component] needs to be reevaluated.	Sec. 3.5	p. 3-95
Development of a clearly stated sampling design strategy [Acid-Sensitive Lakes component].	Sec. 3.5	p. 3-53
Regionally distributed climate data are limited, and data that are available are not always available through the online data access point.	Sec. 3.1	Data for all RAMP monitoring activities are now publicly available online through RAMP's website (www.ramp-alberta.org).
Clarification of test site locations with respect to developments and land disturbances.	Sec. 3.2	Land change development is assessed and documented each monitoring year using satellite imagery. The verification and classification of stations as baseline or test is completed using the satellite imagery and any new land change from the previous year. See p. 2-10 to 2-16 and Appendix A.
Include naphthenic acids (NAs) as part of the monitoring program for water quality.	Sec. 3.2	Low level naphthenic acids are already being measured in water using three different analytical methods. See p. 6-1 to 6-7 and Appendix A2 of this report.

Table 2 (Cont'd.)

Comment/Recommendation	Reference	Response
Consider studies on seasonal variability in water quality in relation to results obtained from the annual fall sampling program.	Sec. 3.2	Seasonal sampling occurs at stations for three years prior to development. Seasonal sampling occurs every year on the Athabasca River at the upstream and downstream stations. AENV also conducts monthly sampling for three stations on the lower the Athabasca River.
Return to lethal sampling of sentinel [fish] species.	Sec. 3.4	Lethal sampling for the sentinel species was re-introduced on a 3-year rotation in 2010. See p. 5-25 to 5-29.
Harmonization and integration of both [benthos and sediment quality] RAMP components.	Sec. 3.3	The harmonization of the benthos and sediment quality sampling has been completed for each reach (i.e., when benthos is collected from a depositional reach, sediment samples are also collected from this reach). Benthos samples are collected at 10 sites within the reach, however, sediment samples for analytical analyses are only taken from a single location located at the downstream boundary of the reach. Sediment is collected to provide supporting data for benthic invertebrate community structure and the scale of sampling is consistent with water quality sampling. See watershed results in Chapter 5.
The integration or harmonization of the hydrologic, chemical and biotic components is seen as integral in the understanding of impact significance.	Sec. 3.3	From a sampling location perspective, a great deal of harmonization among the water quality, sediment quality and benthic invertebrate (benthos) components has occurred. See watershed results in Chapter 5.

Table 3 Components of the RAMP peer review that will be addressed in the 2011 monitoring program.

Comment/Recommendation	Reference	Response
Focus on long-term trend analysis by maintaining a high number of monitoring stations. The addition of climate stations, specifically in the region south of Fort McMurray.	Sec. 3.1	A climate station will be installed south of Fort McMurray in 2011. A field survey to establish new baseline stations will occur in summer 2011 with plans to begin monitoring in 2012 (or fall 2011, if possible).
Increase [water quality] monitoring of lakes and other surface water features, especially baseline sites.	Sec. 3.2	A new baseline lake was added to the monitoring program in 2011. Field surveys will be conducted to find additional baseline lakes.
Consider studies on seasonal variability in water quality in relation to results obtained from the annual fall sampling program.	Sec. 3.2	A pilot study will be conducted in 2012 to look at seasonal variability at a subset of RAMP water quality stations.
Include polycyclic aromatic hydrocarbons (PAHs) as part of the monitoring program in the water column.	Sec. 3.2	PAHs will be sampled in water for the 2011 program. It was discontinued in 2005 given all results were below detection limit. With improved detection limits, RAMP has decided to re-introduce this analysis.
Increase in sampling [of benthic communities] along the mainstem Athabasca River.	Sec. 3.3	A pilot study to determine the best gear to sample benthos in the shifting-sand environment of the mainstem Athabasca River will be conducted in 2011.

Table 4 Components of the RAMP Peer Review that will be re-examined once more direction is received from government monitoring initiatives.

Comment/Recommendation	Reference	Response
1. There should be more integration with the airshed and groundwater monitoring programs to accurately assess and characterize the baseline sites.	2.0	The integration among monitoring programs of different media in the region is currently beyond the scope of RAMP; however, this is a topic being evaluated by the government monitoring initiatives.
3. The RAMP Review Panel recommends that the compliance monitoring be integrated into a broader monitoring strategy that includes RAMP.	2.0	Currently, the RAMP design has a strong focus on project-specific compliance needs, but also incorporates core regional stations that are baseline or test in nature. The core test stations are typically located downstream of multiple operations and are used in understanding potential cumulative effects (e.g., Athabasca River downstream of development and the Athabasca Delta). RAMP is unable to make changes to the current scope or mandate of the program until the government monitoring initiatives are complete and RAMP is provided more direction.
4. All monitoring should be integrated into a broader monitoring strategy that includes RAMP.	2.0	The integration of monitoring programs is currently being evaluated by the government monitoring initiatives.
7. An external Science Advisory Panel should be created to provide continuous, hands-on oversight. This external panel should work concurrently with the RAMP Technical Committee.	2.0	RAMP's original intent was to establish a long-term science panel from the scientists involved in the Peer Review process. This initiative is currently on hold pending results of the government monitoring initiatives.
8. The 5-year RAMP Scientific Peer Review should be continued using a review panel composed of experts that are separate from of the External Science Advisory Panel. The review process should ensure that the integration across components is addressed before delivery of a final report.	2.0	As stated in the response to Recommendation #7, RAMP's original intent was to establish a long-term science panel from the scientists involved in the Peer Review process. The formation of a review panel is also possible. These initiatives will be re-examined once the government monitoring initiatives are complete and RAMP is provided more direction.
10. Incorporate groundwater modeling as a formal part of RAMP and/or utilize groundwater data that are currently collected in the study area by others.	Appendix B	This is beyond the current scope of RAMP, but will be re-examined once government monitoring initiatives are complete and RAMP is provided more direction.
17. Addition of complementary water table monitoring stations in the major terrain units (i.e. low-lying areas, mixed deciduous, jack pine and open land (wetland) / lake) would be of great value for hydrological modeling. This may be better addressed by sideline research projects rather than by expansion of core monitoring.	Appendix B	This is beyond the current scope of RAMP but will be re-examined once government monitoring initiatives are complete and RAMP is provided more direction.
18. Due to the importance of groundwater data for use in assessing possible causes of changes in water quality data it is recommended that some of this information be made available via RAMP.	Appendix B	This is beyond the current scope of RAMP but will be re-examined once government monitoring initiatives are complete and RAMP is provided more direction.

Table 4 (Cont'd.)

Comment/Recommendation	Reference	Response
25. New baseline streamflow and lake, wetland and soil water level stations need to be added to maintain baseline/test ratio and to capture storage changes on the watershed. New stations should ideally not be slated for development for 20+ years. Selection of additional baseline stations outside the Regional Municipality of Wood Buffalo will be required. These new stations should also target acid sensitive lake watersheds to provide cross-linkage and additional records for calibration of isotope mass balance.	Appendix B	This comment relates to the Acid-Sensitive Lakes component. Adding these elements to sampling would greatly increase the scope of the ASL component, which is currently sampled by AENV, with data provided to RAMP for analysis. For example, installation of continuous level meters would require equipment servicing at all lakes every month. Because it would represent a substantial redesign of the ASL component, the group deferred this item to broader discussions of the redesign of RAMP following direction from the government monitoring initiatives.
34. RAMP could consider the establishment of a standing external review committee to comment on the activities of the Technical Program Committee.	Appendix C	RAMP agrees that there is value in having a review committee for all regional monitoring that would ensure broader program linkages. This is a governance related issue, which will be addressed by the government monitoring initiatives.
43. For both river systems and lakes, efforts should be made to assess the number of baseline sites that would be appropriate given the current number of test sites and increase the number of reference sites to that level.	Appendix C	RAMP will continue to focus on maintaining and increasing the number of baseline stations in the program. However, the appropriate ratio of test:baseline stations must be based on the questions being asked and the analytical design adopted for RAMP. As such, beyond continuing to add baseline stations in 2011 and 2012 as previously planned, this is a larger redesign issue that will be re-examined once the government monitoring initiatives are complete and RAMP is provided more direction.
66. The overall initiative needs to be tied together in a more transparent and public fashion. Monitoring or Surveillance? Components need to be tied together and need a range of initiatives that are linked with similar and overlapping components. There needs to be sufficient linkages to overlap the programs (i.e. so that baseline assessments are tied to indicators useful for EEM and performance, as well as are important in EIA and CEA evaluations).	Appendix E	This will be considered in the review of the overall scope of RAMP and other monitoring in the region that is currently being undertaken by government monitoring initiatives.
68. Limited opportunity for finding reference sites in large rivers – focusing on smaller sites will (at minimum) double the number of sites available for each decrease in the order of the river under study; these could also be standardized between watersheds and add another layer to the analysis (1st order vs. 1st order, 2nd order vs. 2nd order….etc.).	Appendix E	A focus on smaller tributaries is a good idea, and will be re-examined once the government monitoring initiatives have provided further direction to RAMP. A field survey to establish new baseline stations on small tributaries to the Athabasca River will be conducted in 2011.

Table 4 (Cont'd.)

Comment/Recommendation	Reference	Response
88. Fish Fence - Continue the Muskeg River fish fence spawning survey in all years with sufficiently low spring discharge. Also extend the spawning fish fence program to other trap-able tributaries. Further investigation should be completed on spawning habitat, egg survival, fry survival, rearing habitat and toxicological assessments on early life history stages.	Appendix G	The Department of Fisheries and Oceans (DFO) required three years of fish fence monitoring on the Muskeg River. This was completed in 2003, 2006, and 2009. An expanded fish fence monitoring program is currently outside the scope of RAMP, although we agree this information would be useful in obtaining more biological information on fish species that use the tributaries for spawning. If additional monitoring of spawning fish populations and spawning habitat becomes an activity under RAMP, alternative gear types (i.e., underwater camera, hoop nets, etc.) will be investigated given the difficulty of installing fish fences in all hydrologic conditions. Alternative gear types would allow for multiple tributaries to be monitored each year. It is recommended that the use of fish fences or investigations/monitoring into spawning fish populations in tributaries be re-examined once the government monitoring initiatives are complete and RAMP is provided more direction. Directed studies on early life history stages of large-bodied species would obtain important biological information, which would provide context for data currently collected by RAMP.
90. Sampling Design - The second philosophical shift that I will argue for is one that is front and centre in the Whittier and Hughes review (Hughes and Whittier 2008). This works needs to be done at a large number of clearly stratified and random sites, not at a small number of fixed sites.	Appendix G	The development of a probabilistic design for monitoring is a major change in scope for all components of RAMP. There are concerns about whether a probabilistic design will still meet the site-specific sampling requirements given the random selection of sampling sites each year using this approach. Certainly there is currently a strong need to incorporate site-specific monitoring in support of monitoring requirements stipulated in EPEA approvals for individual oil sands projects. RAMP can re-examine this recommendation once government monitoring initiatives are complete and RAMP is provided more direction.

Table 4 (Cont'd.)

Comment/Recommendation	Reference	Response
91. A more informative approach involves development of mechanistic models of physical, hydrological and biological processes that control success of various species followed by application to putative impact sites to examine deviations in success. This requires several philosophical changes. First develop these models within the Athabasca basin at un-impacted sites. Second recognize that the various components of the ecosystem are linked in some cases strongly and in some weakly and coordinate sampling of all components including hydrology, chemistry, and biota, both spatially and temporally. Additional data on benthic prey abundance, assessments of reproductive effort and success of rearing juveniles in pristine sites would provide the models to assess impacts of development on success. Of course this is best done in the context of the fish community analysis (which I discuss below) and the sentinel species program should be imbedded within it.	Appendix G	To date, RAMP has not considered developing specific mechanistic models for the purpose of long-term monitoring. However, it is recognized that RAMP is collecting data that could be used for this purpose. There are other initiatives undertaken by some industry members through their approval conditions focusing on developing a better understanding between fish population size/biomass in relation to key habitat variables/thresholds; however, these are site-specific in nature and not focused on impact scenarios. Given the use of mechanistic models does represent a shift in monitoring philosophy, RAMP will re-examine this recommendation once the government monitoring initiatives are complete and RAMP is provided more direction.
92. Sampling Design - I recommend strongly that a whole watershed design with random (or at least regular) sampling along all waterways from low order streams to the mainstem Athabasca River be implemented for the hydrology, chemistry, benthos and fish components in an integrated design. A spatial data base such as this could indentify "hot spots" of concern in various measures, provide time series of whole basin measures and facilitate assessments of spatial and temporal cumulative effects.	Appendix G	The transition from a site-specific, control-impact design (with some regional stations) to a completely randomized design is a large change in scope of the program (with different monitoring objectives) not only for the Fish Populations component but for all monitoring components. RAMP will re-examine this recommendation once the government monitoring initiatives are complete and more direction is provided.
97. All Fish Populations Activities - The RAMP program has provided key data on which to develop a rigorous monitoring program but now needs to focus on stratified random sampling to appropriately characterize spatial and temporal variability in the Athabasca watershed. In fact, the focus needs to shift from the idea of variability in data to variability in processes.	Appendix G	As stated in the response to Recommendation #92, the transition from a site-specific, control-impact design (with some regional stations) to a completely randomized design is a large change in scope of the program (with different monitoring objectives) not only for the Fish Populations component but for all monitoring components. RAMP will re-examine this recommendation once the government monitoring initiatives are complete and more direction is provided.

Table 4 (Cont'd.)

Comment/Recommendation	Reference	Response
104. Fish Fence – recommend using mobile gears such as hoop nets to monitor spawning runs in tributaries instead so that spawning runs can be monitored in any hydrologic conditions and capture all variability.	Appendix H	RAMP agrees that it is important to assess spawning fish populations in tributaries to determine which tributaries are being used and evaluate the variability in the strength/richness of spawning runs over time. Unfortunately, fish fences have not proven to be a reliable method to monitor spawning fish populations (cannot be used in all hydrologic conditions, labour intensive, costly). Accordingly, hoop nets or other types of gear could be considered to ensure successful deployment during high water years. Alternate approaches may prove to be less labour intensive (and more cost-effective) such that a greater number of tributaries could be monitored. Once direction from the government monitoring initiatives is provided, this topic will be re-examined by RAMP as it is outside the current scope.
120. Fish Inventory - the Summer inventory on the Clearwater should not continue as in 2008-09 but rather be rolled into a broader Athabasca River assemblage monitoring program as was done on the smaller rivers in 2009.	Appendix H	This recommendation is suggesting that the Athabasca and Clearwater fish inventories should be conducted using a probabilistic sampling design by randomly sampling reaches from the upstream end of the Clearwater to the Athabasca Delta. We agree that it would be worthwhile to look into a randomized design for these rivers given their size and the influence of many tributaries; however, this is a major change from the current scope of RAMP and will be re-examined once the government monitoring initiatives are complete and RAMP is provided more direction.

2.1 GENERAL COMMENTS

The following general comments and recommendations were made by all reviewers with respect to RAMP. These were presented as part of the panel's overall synthesis in the main body of the review report. Following each comment or recommendation, a response has been provided by the RAMP Technical Program Committee.

1. *There should be more integration with the airshed and groundwater monitoring programs to accurately assess and characterize the baseline sites* (Sec 2.0, p. 3, para 3).

The integration among monitoring programs of different media in the region is currently beyond the scope of RAMP; however, this is a topic being evaluated by the government monitoring initiatives.

2. *There should be an assessment of which predictions have the potential to be validated with the existing RAMP monitoring program* (Sec 2.0, p. 3, para 4).

A summary of potential effects and related variables was summarized in the RAMP Design and Rationale document (RAMP 2009b). This assessment formed, in part, the basis for selecting specific monitoring endpoints and effect sizes in RAMP. Accordingly, data collected by RAMP can already be more easily used by regulatory agencies to evaluate EIA predictions and approval requirements.

3. *The RAMP Review Panel recommends that the compliance monitoring be integrated into a broader monitoring strategy that includes RAMP* (Sec 2.0, p. 4, para 1).

Currently, the RAMP design has a strong focus on project-specific compliance needs, but also incorporates core regional stations that are baseline or test in nature. The core test stations are typically located downstream of multiple operations and are used in understanding potential cumulative effects (e.g., Athabasca River downstream of development and the Athabasca Delta). RAMP is unable to make changes to the current scope or mandate of the program until the government monitoring initiatives are complete and RAMP is provided more direction.

4. *All monitoring should be integrated into a broader monitoring strategy that includes RAMP* (Sec 2.0, p. 4, para 2).

The integration of monitoring programs is currently being evaluated by the government monitoring initiatives.

5. *Traditional Ecological Knowledge (TEK) should be integrated into a broader monitoring strategy that includes RAMP* (Sec 2.0, p. 4, para 3).

This is an ongoing challenge that is of interest to RAMP and its stakeholders. It is agreed that more TEK is required, where appropriate. In October 2009, RAMP began to work with Elders from Fort McKay to develop a better understanding of their needs and interests. Information sharing between the Elders and RAMP was initiated and will hopefully develop over time. Should this approach prove successful, then there will be an opportunity to apply it to other Aboriginal communities in the region.

6. *An improved communications strategy for the release of data and reports is required* (Sec 2.0, p. 4, para 4).

RAMP does have a Communications Committee and strategy (not requested by the panel for review) and continues to work on improving this element of the program.

7. *An external Science Advisory Panel should be created to provide continuous, hands-on oversight. This external panel should work concurrently with the RAMP Technical Committee* (Sec 2.0, p. 4, para 4).

RAMP's original intent was to establish a long-term science panel from the scientists involved in the Peer Review process. This initiative is currently on hold pending results of the government monitoring initiatives.

8. *The 5-year RAMP Scientific Peer Review should be continued using a review panel composed of experts that are separate from of the External Science Advisory Panel. The review process should ensure that the integration across components is addressed before delivery of a final report* (Sec 2.0, p. 5, para 1).

As stated in the response to Recommendation #7, RAMP's original intent was to establish a long-term science panel from the scientists involved in the Peer Review process. The formation of a review panel is also possible. These initiatives will be re-examined once the government monitoring initiatives are complete and RAMP is provided more direction.

2.2 CLIMATE AND HYDROLOGY

2.2.1 Comments and Recommendations by Dr. Donald Burn

The following comments and recommendations were made by Donald Burn to improve the Climate and Hydrology component of RAMP.

9. *Reconsider the use of the current water balance model as the basis for estimating baseline measurement endpoints. If the water balance model is to be the basis for the evaluation of impacts, it is important that the modeling approach be as accurate as possible [numerous assumptions in the WB model that reduce estimate accuracy]* (Ap A, p. 10, p. 4, para 3).

The use of the water balance model was discussed in the peer review meetings held in October 2010. The most critical assumption from past water balance analyses was the apportionment of annual and/or monthly industrial data to daily data. It was agreed that it is important that, whenever possible, industrial withdrawal and discharge data used in the model needs to be provided on a daily time-step rather than apportioning data from a monthly (or annual) time-step to meet the daily time-step requirements of the model. While data for some withdrawals has been typically provided using a daily time-step, in previous years, many have been provided on an aggregated monthly and, in some cases, aggregated annual basis. Withdrawal and discharge data reported by industry on a daily basis has significantly increased for 2010. For the 2010 WY analyses all but one industrial data set was provided in daily values thereby improving model accuracy.

The current water balance assumptions regarding the runoff affects caused by clearing, are based on the results of long-term studies regarding land-use change in similar topographic conditions (AENV 2000). A sensitivity analysis is underway to assess the model assumptions around runoff distribution. The preliminary results of the sensitivity analysis support the use of the model assumptions regarding runoff apportionment when applied under the current RAMP development scenario. The assumptions are applicable given that land-use change affects a relatively small portion of the total watershed area for most tributaries being assessed. In the future, the proportion of development will dictate the need for analytical adjustments, and given that this would be in the future, it is assumed that the hydrometric time-series will be sufficiently long to support alternative approaches for an increased number of stations, at a greater degree of accuracy. The water balance model, particularly when using a daily time-step for model inputs, is considered a useful approach during the current period when record-length is a limiting factor for analytical alternatives and land-use change is generally affecting a small proportion of the total watershed area. There are several stations where record-length is becoming sufficiently long to consider alternative approaches to supplement the water balance model. This includes the Muskeg, MacKay, Firebag, Steepbank, Clearwater and Christina watersheds as discussed below.

The development of a longer dataset at some stations provides the increasing opportunity to also consider other approaches (statistical, HRI, etc). The development of these approaches has been underway since 2009 and can only be applied to stations with sufficient record length. It was discussed with Donald Burn at the RAMP Tech meeting if he would agree that continued use of the water balance model (using daily industrial data) coupled with HRI/statistical approaches at select long-term stations would be an acceptable forward direction. He agreed.

As the monitoring program measures the discharge at key locations downstream of development, these measurement points measure the cumulative effect of all influences on the upstream hydrology. The water balance model "removes" the influence of industrial activity alone and provides an opportunity to consider this potential impact without considering climate change and/or other factors that contribute to changes in the hydrologic regime.

While RAMP is a monitoring program and field measurement of impact is the most effective approach to determine "what is happening", the data collected by RAMP is also very useful for calibrating predictive models to support EIA predictions of "what if". In this regard it is important to continue the dialogue with Industry and regulatory representatives to support data collection that will provide valuable inputs to the predictive models used in EIAs such as the Hydrological Simulation Program—Fortran (HSPF) and others.

10. *Incorporate groundwater modeling as a formal part of RAMP and/or utilize groundwater data that are currently collected in the study area by others.* (Ap B, p. 5, para 1 and Ap A, p. 10, p. 8, para 3).

This is beyond the current scope of RAMP, but will be re-examined once government monitoring initiatives are complete and RAMP is provided more direction.

11. *Develop a more proactive approach to data collection network design. [current network design process is reactive]* (Ap A, p. 10).

The annual planning meetings of the RAMP Climate and Hydrology Subgroup provide the opportunity to assess development plans and station recommendations. This group proactively identifies monitoring needs in conjunction with available knowledge pertaining to anticipated development scenarios. As a result of this proactive approach, most RAMP stations have greater than three years of baseline data collection in advance of development.

In 2010, the RAMP Climate and Hydrology network included monitoring at 14 baseline hydrometric stations, six stations measuring flow from watersheds with less than 5% oil sands-related land-use change, and 16 stations with greater than 5% oil sands-related land-use change.

RAMP members have decided to implement a Lower Athabasca River monitoring location that will be downstream of all currently anticipated oil sands development. Planning for this station, located approximately six kilometres upstream of the confluence of the Embarras River with the Athabasca River, began more than four years ago with station reconnaissance work completed in 2009 and station installation planned post-ice-breakup (when ice-free conditions exist) in 2011.

RAMP maintains stations that support a long-term record length. Of the 40 stations monitored under the Climate and Hydrology component, since 2005, 11 stations have been added and only two stations have been discontinued. In terms of hydrometric stations, as of 2010, there are seven stations with more than ten years of record; 17 stations with five to ten years; and 13 stations with less than five years. The 2010 RAMP network represents a density of 5.8 stations/1000 km^2 compared to the national WSC program station density of 2.9 stations/1000 km^2. The number of stations (and the density) is continuing to grow every year as more stations are added to the network.

RAMP is continuing to identify areas that are baseline and stations that have potential to remain baseline into the foreseeable future (i.e., upper Tar River). This discussion occurs on an annual basis at RAMP Technical meetings.

12. *Where practical, monitor streamflow gauging stations on a year round basis. Streamflow gauging stations need to be monitored on a year round basis rather than on a seasonal basis. Seasonal monitoring misses important components of runoff regime* (Ap A, p. 10, p. 5, para 4).

RAMP supports winter measurement at WSC stations to provide year-round data at these stations. RAMP stations are selected for year-round monitoring based on locations where winter flow is expected, i.e. where the watercourse is not expected to freeze to depth in winter. A pilot program will be conducted in January 2012 to re-confirm that seasonal stations (not monitored in winter) do freeze to depth and evaluate if additional winter stations could be added. Clarification regarding the freeze-to-depth issue was provided to Donald Burn at the RAMP Tech meeting.

13. *Add additional measurement end points that reflect the timing of the hydrological response. The calculation of additional measurement endpoints would be beneficial to provide a more complete understanding of the hydrological regime, and changes to the regime, in the study area. It may be necessary to rethink the overall strategy of determining differences between baseline and test conditions in order to obtain a more comprehensive set of measurement endpoints* (Ap A, p. 10, p. 8, para 4, p. 6, para 3).

RAMP currently assesses changes in hydrologic response using EDA (exploratory data analyses) and, as the time-series record becomes sufficiently long, additional approaches are being developed. Initial assessment of potential approaches occurred in 2009 with further development of techniques in 2010/2011 that consider biologically relevant measurement endpoints using a Hydrologic Regime Indicator (HRI) approach. This approach is being developed to support identification and understanding of trends and pre/post shifts in hydrologic regime characteristics and considers changes in hydrologic regime including seasonality and other factors to support the understanding of potential linkages with others components. This approach requires sufficient record length and is; therefore, applicable to a sub-set of RAMP locations until sufficient record length is obtained at other stations.

14. *Reconsider the use of trend analysis for record lengths that are very short. A better approach to address changes in variables when the record length is so short is to summarize the slope values (calculate using a robust estimate of the slope) and highlight changes that are, and are not, consistent with the hypothesis of interest* (Ap A, p. 10, p. 7, para 1).

The group discussed the increasing use of trend analysis in RAMP generally (as the number of years of data increase), and tasked the Implementation Team with assessing the strengths and weaknesses of different trend analysis approaches and which ones were most appropriate for the datasets collected by RAMP. The Implementation Team will report its findings at the next RAMP Tech workshop in fall 2011.

The ASL component is already doing this in one sense. The ANOVA using the general linear model regresses measurement endpoints against year in each individual lake and determines the significance over all lakes collectively. The analysis lists all 50 individual regression coefficients. We have also tried plotting these regression coefficients against the potential acid input (PAI) to determine whether the observed trends in each lake were related to acid deposition.

15. *More on Proactive rather than Reactive design: The network design process needs to anticipate the development of oils sands properties and locate gauging stations both upstream and downstream of potential development sites to ensure that : 1) baseline conditions are continually monitored for the undisturbed portion of the watershed; and 2) downstream baseline record lengths for the period prior to the development are sufficiently lengthy to form a strong basis of comparison with test conditions measured after the development of oil sands in the watershed. Consideration should also be given to the concept of developing a "regional" network of gauging stations, consisting of both stations in the study area and stations close to the study area. The latter should be stations that can be considered to be hydrologically similar to the stations within the study area* (S. 3.1, p. 6, para 3).

As stated in the response to Recommendation #11, the annual planning meetings of the RAMP Climate and Hydrology Subgroup provide the opportunity to assess development plans and station recommendations. This group proactively identifies monitoring needs in conjunction with available knowledge pertaining to anticipated development scenarios. As a result of this proactive approach, most RAMP stations have greater than three years of baseline data collection in advance of development.

In 2010, the RAMP Climate and Hydrology network included monitoring at 14 baseline hydrometric stations, six stations measuring flow from watersheds with less than 5% oil sands-related land-use change, and 16 stations with greater than 5% oil sands-related land-use change.

RAMP members have been working toward the addition of a Lower Athabasca River monitoring location that will be downstream of all currently anticipated oil sands development. Planning for this station began more than four years ago with station reconnaissance work completed in 2009 and station installation planned post-ice-breakup (when ice-free conditions exist) in 2011.

RAMP maintains stations that support a long-term record length. Of the 40 stations monitored under the Climate and Hydrology component since 2005, 11 stations have been added and only two stations have been discontinued. In terms of hydrometric stations, as of 2010, there are seven stations with more than ten years of record; 17 stations with five to ten years; and 13 stations with less than five years. The 2010 RAMP network represents a density of 5.8 stations/1,000 km^2 compared to the national WSC program station density of 2.9 stations/1,000 km^2. The number of stations (and the density) is continuing to grow every year as more stations are added to the network.

RAMP is continuing to identify areas that are baseline and stations that have potential to remain baseline into the foreseeable future (i.e., upper Tar River). This discussion occurs on an annual basis at RAMP Tech meetings.

16. *More could be done with the climate data, particularly if the Climate and Hydrology component were to move beyond the simple water balance model to the use of a hydrological model to better represent the response of watersheds and the effects of the oils sands development on the watersheds. [See 5. re hydrologic response above]* (Ap A, p. 9, para 2).

Climate data collected by RAMP are available for use in predictive models required for EIA purposes. These data can be used to support calibration of models used for this purpose. The climate data also provide a regional context for assessing flow conditions in response to climate. The EDA (exploratory data analysis) is included in the RAMP technical reports.

2.2.2 Comments and Recommendations by Dr. John Gibson

The following comments and recommendations were made by John Gibson to improve the Climate and Hydrology component of RAMP.

17. *Addition of complementary water table monitoring stations in the major terrain units (i.e. low-lying areas, mixed deciduous, jack pine and open land (wetland) / lake) would be of great value for hydrological modeling. This may be better addressed by sideline research projects rather than by expansion of core monitoring* (Ap B, p. 4, para 4).

This is beyond the current scope of RAMP but will be re-examined once government monitoring initiatives are complete and RAMP is provided more direction.

18. *Due to the importance of groundwater data for use in assessing possible causes of changes in water quality data it is recommended that some of this information be made available via RAMP* (Ap B, p. 6, para 1).

This is beyond the current scope of RAMP but will be re-examined once government monitoring initiatives are complete and RAMP is provided more direction.

19. *Time-series and spatial coverage of streamflow gauging for both baseline and test stations needs to be continued at all possible stations to provide representative records of average flows, return flows, low flows, and high flows as well as to enable statistical trend analysis. Due to continual reduction in baseline stations, special attention needs to be given to maintaining as many stations as possible over the next four decades to monitor anticipated cumulative changes in regional runoff response* (Ap B, Rec #1).

Agreed. As stated in the response to Recommendation #15, of the 40 stations monitored under the Climate and Hydrology component since 2005, 11 stations have been added and only two stations have been discontinued. In terms of hydrometric stations, as of 2010, there are seven stations with more than ten years of record; 17 stations with five to ten years; and 13 stations with less than five years.

20. *Add PAH and heavy metals to routine snowpack sampling. Kelly-Schindler research suggests that these constituents require further monitoring in snowpack* (Ap B, Rec #2).

The current snow survey is not designed to look at aerial deposition of chemicals in snow. This has been discussed by RAMP and is an integration question between WBEA and RAMP once further direction has been received from the government monitoring initiatives.

Two snowpack PAH/metals surveys are currently being undertaken by AENV and Environment Canada. While the RAMP Tech committee agreed that monitoring of snowpack chemistry could be added to regional surveillance monitoring (by RAMP, WBEA or others) in future, the committee agreed that the results of these AENV/Environment Canada studies should be reviewed first, to assess if and how this monitoring component should be added to RAMP.

21. *Groundwater level and water quality data in 20-30 existing wells needs to be made accessible via RAMP for the evaluation of storage effects and water quality modeling. Operators and/or AENV may be able to provide required data* (Ap B, Rec #3).

This is beyond the current scope of RAMP but will be re-examined once the government monitoring initiatives are complete and RAMP is provided more direction.

In all cases, for QA/QC purposes, data should be accessed directly from the original source. If considered beneficial, RAMP could provide a link to these data if it is publicly available.

22. *EM terrain conductivity surveys should be used to map position of major seepage inputs to the tributaries and rivers. Inputs should be sampled and characterized for isotopic and geochemical characteristics to identify source formation(s) and influence on water quality* (Ap B, Rec #4).

AENV has funded EM surveys on the mainstem Athabasca in the past. This information would be useful to provide information for the design of a monitoring program and is a research-related question rather than a continuous monitoring task. Moreover, we understand that Natural Resources Canada is currently leading a major study to address these questions.

23. *Use of more comprehensive hydrological models should be considered* (Ap B, Rec #5).

As stated in the response to Recommendation #9, the use of the water balance model was discussed in the peer review meeting held in October 2010. The most critical assumption from past water balance analyses was the apportionment of annual and/or monthly industrial data to daily data. It was agreed that it is important that, whenever possible, industrial withdrawal and discharge data used in the model needs to be provided on a daily time-step rather than apportioning data from a monthly (or annual) time-step to meet the daily time-step requirements of the model. While data for some withdrawals has been typically provided using a daily time-step, in previous years, many have been provided on an aggregated monthly and, in some cases, aggregated annual basis. Withdrawal and discharge data reported by industry on a daily basis has significantly increased for 2010. For the 2010 water year (WY) analyses all but one industrial data set was provided in daily values thereby improving model accuracy.

The current water balance assumptions regarding the runoff effects caused by clearing are based on the results of long-term studies regarding land-use change in similar topographic conditions (AENV 2000). A sensitivity analysis is underway to assess the model assumptions around runoff distribution. The preliminary results of the sensitivity analysis support the use of the model assumptions regarding runoff apportionment when applied under the current RAMP development scenario. The assumptions are applicable given that land-use change affects a relatively small portion of the total watershed area for most tributaries being assessed. In the future, the proportion of development will dictate the need for analytical adjustments, and given that this would be in the future, it is assumed that the hydrometric time-series will be sufficiently long to support alternative approaches for an increased number of stations, at a greater degree of accuracy. The water balance model, particularly when using a daily time-step for model inputs, is considered a useful approach during the current period when record-length is a limiting factor for analytical alternatives and land-use change is generally affecting a small proportion of the total watershed area. There are several stations where record-length is becoming sufficiently long to consider alternative approaches to supplement the water balance model. This includes the Muskeg, MacKay, Firebag, Steepbank, Clearwater and Christina watersheds as discussed below.

The development of a longer data set at some stations provides the increasing opportunity to also consider other approaches (statistical, HRI, etc). The development of these approaches has been underway since 2009 and can only be applied to stations with sufficient record length. It was discussed with Donald Burn at the October RAMP meeting if he would agree that continued use of the water balance model (using daily industrial data) coupled with HRI/statistical approaches at select long-term stations would be an acceptable forward direction. He agreed.

As the monitoring program measures the discharge at key locations downstream of development, these measurement points measure the cumulative effect of all influences on the upstream hydrology. The water balance model "removes" the influence of industrial activity alone and provides an opportunity to consider this potential impact without considering climate change and/or other factors that contribute to changes in the hydrologic regime.

While RAMP is a monitoring program and field measurement of impact is the most effective approach to determine "what is happening", the data collected by RAMP is also very useful for calibrating predictive models to support EIA predictions of "what if". In this regard it is important to continue the dialogue with Industry and regulatory representatives to support data collection that will provide valuable inputs to the predictive models used in EIAs such as HSPF and others.

24. *Water quality monitoring should consider using FT-IRCMS to scan for natural organic compounds as a compliment to naphthenic acid partitioning for fingerprinting natural anthropogenic sources* (Ap B, Rec #6).

This comment relates to the Water Quality component. RAMP supported high-resolution MS-based studies (including FT-IRMS) studies of regional water quality in 2010 through collection and provision of additional samples to multiple laboratories for naphthenic acids/acid-extractable organics analysis (including the University of Alberta, AITF and ALS). Given the exploratory state of the science around this and broader studies of organic constituents of water ("petroleomics" as described by Dr. Gibson), this work is more appropriate as primary research than for use in an operational surveillance monitoring program such as RAMP. The group agreed that RAMP should collect an additional sample for acid-extractable organics analysis in 2011, which will be archived and could be provided to researchers for their use if desired. RAMP members will also alert the Oil Sands Research and Information Network (OSRIN) and/or the Canadian Oil Sands Network for Research and Development (CONRAD) about this research question, to support possible research in this area.

25. *New baseline streamflow and lake, wetland and soil water level stations need to be added to maintain baseline/test ratio and to capture storage changes on the watershed. New stations should ideally not be slated for development for 20+ years. Selection of additional baseline stations outside the Regional Municipality of Wood Buffalo will be required. These new stations should also target acid sensitive lake watersheds to provide cross-linkage and additional records for calibration of isotope mass balance* (Ap B, Rec #7 and p. 7, para 4, pg 8, para 1).

This comment relates to the Acid-Sensitive Lakes component. Adding these elements to sampling would greatly increase the scope of the ASL component, which is currently sampled by AENV, with data provided to RAMP for analysis. For example, installation of continuous level meters would require equipment servicing at all lakes every month. Because it would represent a substantial redesign of the ASL component, the group deferred this item to broader discussions of the redesign of RAMP following direction from the government monitoring initiatives.

26. *Improved coverage in water quality, acid sensitive lakes and lake level storage can take advantage of stations already monitored under special projects. Additional stations could include the Nexen lakes special project. New stations that build on previous work are preferred to sites with no background data* (Ap B, Rec #8).

As hydrology is a major driver of water chemistry in the ASL component lakes, continuous water-level monitoring could be added to the component, perhaps by installation and monitoring of data loggers that would be maintained in conjunction with the ASL component. The information obtained would then be available for calculation of isotopic runoff estimates. This information is not needed for assessing the acid-sensitivity or chemical status of these lakes, but would provide added value for regional monitoring program, generally. Because it would represent a substantial redesign of the ASL component, the group deferred this item to broader discussions of RAMP redesign following direction from government monitoring initiatives. The group agreed that greater integration between water quality and hydrology components was desirable, and that ways to harmonize water quality and hydrology monitoring would be explored in 2011 (in many cases, hydrology stations are located at different locations in watersheds than water quality stations because of specific geomorphology requirements for hydrology stations).

27. *Qualitative criteria for impact assessment apparently use different thresholds for mining and in situ projects. Clarification is required on choice of thresholds for projects/areas. Standardized impacts are needed* (Ap B, Rec #9).

There are different formats used in the various oil sands EIAs. RAMP has tried to consider thresholds used by mining and *in situ* operations when establishing interim criteria for the monitoring program (Table 5, taken from the RAMP Technical Design and Rationale document [RAMP 2009b]). However, further threshold development is expected through the ongoing work by CEMA, as well as the activities of the government monitoring initiatives.

Table 5 Measurement endpoints and criteria for determination of change used in RAMP (2011).

RAMP Component	Measurement Endpoints Used in 2010 Technical Report	Criteria for Determining Change Used in 2010 Technical Report
Climate and Hydrology	Mean open-water season discharge Mean winter discharge Annual maximum daily discharge Open-water season minimum daily discharge	Differences between observed *test* and estimated *baseline* hydrographs (i.e., the hydrograph that would have been observed had focal projects and other oil sands developments not occurred in the drainage, so that changes in water withdrawals, discharges, and diversions are accounted for) as follows: Negligible-Low: ± 5% ; Moderate: ± 15%;High: > 15%.
Water Quality	pH Total suspended solids Dissolved phosphorus Total nitrogen and nitrate-nitrite Various ions (sodium, chloride, sulphate) Total alkalinity, Total dissolved solids Dissolved organic carbon Total and dissolved aluminum Total arsenic, Total boron Total molybdenum, Total strontium Ultra-trace mercury, Naphthenic acids Overall ionic composition	Comparison to range of regional *baseline* conditions. Comparison to CCME and other water quality guidelines. Calculation of water quality index based on CCME water quality index found at http://www.ccme.ca/ourwork/water.html?category_id=102 , with water quality index scores classified as follows: 80 to 100: Negligible-Low difference from regional *baseline* conditions 60 to 80: Moderate difference from regional *baseline* conditions Less than 60: High difference from regional *baseline* conditions
Benthic Invertebrate Communities	Abundance Richness (number of taxa) Simpson's Diversity Evenness Abundance of EPT (mayflies, stoneflies, caddisflies) Axes of Correspondence Analysis ordination	Exceedance of regional range of *baseline* variability for the selected measurement endpoints based on the mean and standard deviation, with regional range defined as $\bar{X} \pm 2SD$, and statistically significant differences between measurement endpoints in *test* reaches/lakes as compared to *baseline* reaches/lakes; 1. Negligible-Low: no strong statistically significant difference in any measurement endpoint between *test* and *baseline* reaches/lakes 2. Moderate: strong statistically significant difference in one any measurement endpoint between *test* and *baseline* reaches/lakes, with low "noise" in the statistical test, but no measurement endpoint outside *baseline* range of natural variation 3. High: statistically significant difference in one any measurement endpoint between *test* and *baseline* reaches/lakes and either: (i) at least three measurement endpoints outside *baseline* range of natural variation or (ii) at least one measurement endpoint outside *baseline* range of natural variation for three consecutive years
Sediment Quality	Particle size distribution (clay, silt and sand) Total organic carbon Total hydrocarbons (CCME and Alberta Tier 1) Various PAH end-points, including: ▪ Total PAHs ▪ Total Low-Molecular Weight PAHs ▪ Total High-Molecular Weight PAHs ▪ Naphthelene, Retene ▪ Total dibenzothiophenes Predicted PAH toxicity Metals, Chronic toxicity	Comparison to CCME Interim Sediment Quality Guidelines (ISQG) and other guidelines. Calculation of sediment quality index based on CCME water quality index found at http://www.ccme.ca/ourwork/water.html?category_id=103, with sediment quality index scores classified as follows: 80 to 100: Negligible-Low difference from regional *baseline* conditions 60 to 80: Moderate difference from regional *baseline* conditions Less than 60: High difference from regional *baseline* conditions

Table 5 (Cont'd.)

RAMP Component	Measurement Endpoints Used in 2010 Technical Report	Criteria for Determining Change Used in 2010 Technical Report
Fish Populations: Fish Inventory	Relative abundance (catch per unit effort) Length-frequency Percent composition Condition factor	The RAMP fish inventory activity is generally considered to be a stakeholder-driven activity that is best suited for assessing general trends in abundance and population parameters for large-bodied species. It is not specifically designed for assessing environmental effects of focal project activities.
Fish Populations: Regional Lakes Fish Tissue	Mercury concentration in food fish muscle tissue	**Risk to Human Health** Negligible-Low: Fish tissue concentrations for mercury below USEPA and Health Canada criteria for recreational and subsistence fishers and the general consumer. High (subsistence): Fish tissue concentrations for mercury above USEPA and Health Canada criteria for subsistence fishers, but below criteria for recreational fishers and general consumers. High (general consumer): Fish tissue concentrations for mercury above USEPA and Health Canada criteria for general consumers, and recreational and subsistence fishers.
Fish Populations: Sentinel Species Monitoring	Age Growth Condition Factor Gonadosomatic Index (GSI) Liversomatic Index (LSI)	Comparison to Environment Canada's Environmental Effects Monitoring (EEM) criteria (Environment Canada 2010) where an effect is determined by a difference of ± 10% in condition, ± 25% in age, growth, GSI, and LSI of fish at the *test* reach relative to fish condition at the *baseline* reach. Negligible-Low: no exceedance greater than ± 10% in condition, ± 25% in age, growth, GSI, or LSI of fish at *test* site compared to condition of fish at *baseline* site. Moderate: exceedance greater than ± 10% in condition, ± 25% in age, growth, GSI, or LSI of fish at *test* site compared to condition of fish at *baseline* site, but not in two consecutive years of sampling including the current year. High: exceedance greater than ± 10% in condition ± 25% in age, growth, GSI, or LSI of fish at *test* site compared to condition of fish at *baseline* site, and exceedance observed in two consecutive years of sampling including the current year.
Acid-Sensitive Lakes	Critical Load of acidity pH Gran alkalinity Base cation concentrations Nitrate plus nitrite concentrations Dissolved Organic Carbon Aluminum	Exceedance of Critical Load of acidity of a particular lake by the measured or modeled value of the Potential Acid Input (PAI) to that lake. A statistically significant change in any of the measurement endpoints beyond natural variability, resulting in a reduction of lake pH, Gran alkalinity, Critical Load or base cation concentrations or an increase in nitrates or aluminum concentrations. For each lake, mean and standard deviation calculated for each of seven measurement endpoints over all the monitoring years. The number of lakes in 2010 within each subregion with endpoint values greater than two standard deviations from the mean is calculated. Negligible-Low: subregion has <2% endpoint-lake combinations exceeding ± 2SD criterion. Moderate: subregion has 2% to 10 % endpoint-lake combinations exceeding ± 2SD criterion. High: subregion has > 10% of endpoint-lake combinations exceeding ± 2SD criterion.

28. *Make flow a water quality parameter and provide flow-weighted summaries of water quality data* (Ap B, Rec #10).

RAMP agrees that combined analysis of flow and water quality data could be beneficial to interpretation; this was explored to an extent in the RAMP 2010 Technical Report by comparing total suspended solids and flow in the Athabasca River (RAMP 2011, Chapter 6 and Appendix A2 of this document). In addition, this analysis will be expanded in the 2011 Technical Report where possible. Overall, the group agreed that greater integration between water quality and hydrology components was desirable, and that ways to harmonize water quality and hydrology monitoring would be explored in 2011 (in many cases, hydrology stations are located at different locations in watersheds than water quality stations because of specific geomorphology requirements for hydrology stations). Where possible, water quality and hydrology stations could be integrated in 2012 or 2013, based on direction from the government monitoring initiatives. Such harmonization would also need to consider the existing harmonization among water quality, benthos and sediment quality.

29. *Conduct one-time random stratified sampling program to characterize natural variability in runoff (water yield) and acid sensitivity (in a group of randomly selected lakes) as comparative dataset to RAMPs ASL. The WRS (2004) survey may be sufficient for this purpose if water samples are still archived and can be run for stable isotopes of water* (Ap B, Rec #11).

There are currently water quality data for about 400 lakes in the oils sands region collected in a series of lake surveys conducted from the late 1980s till the mid-2000s. These data have been compiled in WRS (2004) and more recently in CEMA (2010). The chemical data from these lakes have been included in the RAMP technical reports since 2003 in the ASL appendices in which the chemistry of the RAMP lakes is compared to the larger database. The acid sensitivities of these 400 lakes have been determined in WRS (2004) and CEMA (2010) from calculations of the critical load and their buffering capacities. Therefore, we have a good estimate of the variability in the acid sensitivity of the lakes in the region and just how typical the chemistry and morphology of the RAMP lakes are of lakes in the oil sands region. Unfortunately, the water yield to each lake was determined by standard hydrometric techniques rather than the isotopic mass balance (IMB) method. In order to obtain water yields for these lakes using the IMB method, the lakes would have to be re-sampled, as water samples were not archived. The group agrees that although obtaining yield estimates by IMB from a larger lake database than the RAMP lakes would be useful, it would require a very large, focused effort. Such a program would be a special, one-time study that is currently outside of the scope of RAMP and, perhaps, more appropriately conducted by research agencies.

30. *Poor presentation of PAI, isotope mass balance and hydrometric method comparisons. Add improved description of methodology for IMB and review of estimates based on IMB vs. hydrometric methods* (Ap B, Rec #12).

This comment refers to the Acid-Sensitive Lakes component and was addressed in the 2010 Technical Report, through expansion of the requested descriptions (RAMP 2011, page 3-95 to 3-97).

The following sections from RAMP (2011) provide support to the response to this recommendation.

The Modified Henriksen Model

The original Henriksen model was modified to account for both the buffering of weak organic anions and the lowering of ANC attributable to strong organic acids. The modified model assumed that DOC, with its associated buffering from weak organic acids (ANC_{org}) and reduction of ANC from strong organic acids (A_{-SA}), was exported from the catchment basin to each lake in the same way that we assume the export of base cations (carbonate alkalinity) to each lake. The modified Henriksen model is:

$$CL= ([BC]^*_0 + ANC_{org} - A^-_{SA} - ANC_{lim}) .Q$$

Where,

$[BC]^*_0$	is the original base cation concentration before acidification;
ANC_{lim}	is the limiting acid-neutralizing capacity of the lake required to maintain a healthy and functional aquatic ecosystem;
ANC_{org}	= 0.00680* DOC exp(0.8833*pH);
A^-_{SA}	= 6.05 *DOC +21.04; and
Q	is the runoff to each lake from the catchment and lake area.

The modifications of the Henriksen model for organic acids and the empirical relationships for developed for ANC_{org} and A_{-SA} are described in WRS (2006) and RAMP (2009b).

Calculation of Runoff (Q)

The runoff (Q) to each lake, was calculated from analysis of heavy isotopes of oxygen (^{18}O) and (^{2}H) in each lake conducted and provided by John Gibson (University of Victoria). With this technique, the natural evaporative enrichment of ^{18}O and ^{2}H in each lake is used to partition water losses between evaporation and liquid outflow and hence derive an estimate of runoff (Gibson *et al*. 2002, Gibson and Edwards 2002, and Gibson *et al*. 2010). This technique utilizes a different set of assumptions from traditional hydrometric methods, which extrapolate water yields from one or more gauged catchments to the ungauged lake catchments. Potential inaccuracies in the traditional hydrometric method, especially in low-relief catchments, have previously been recognized in lakes in the Athabasca oil sands region (WRS 2004).

Original Base Cation Concentration ($[BC]^*_0$)

During the process of acidification of a catchment, base cations are released from the soils to the lake waters. In applying the Henriksen model, it was assumed that base cations have not increased in these lakes as a result of acidic deposition; that is, the current base cation concentrations are equivalent to the original values. This simplifying assumption was adopted for the following two reasons:

1. The discrepancy between the original and the current base cation concentrations in a lake is normally calculated by an equation presented in Brakke *et al.* (1990) based on increases in sulphur concentrations in a lake resulting from aerial deposition. Calculations of [BC]*0 using the Brakke *et al.* (1990) equation indicated that there is an insignificant difference between the current and calculated original base cation concentrations in all 50 lakes.
2. A study by Whitfield *et al.* (2010b) in which the Magic Model was applied to the Athabasca oil sands region concluded that, to date, sulphate deposition levels have resulted in only a limited removal of base cations from the soil.

Choice of ANC_{lim}

The critical load concept as expressed in the Henriksen model assumes a dose-response relationship between a water quality variable and an aquatic indicator organism. In this case, the water quality variable is the acid-neutralizing capacity (alkalinity) required to maintain a healthy fish population. In applying the Henriksen model in Europe, a critical threshold ANC_{lim} of 20 μeq/L was set to protect brown trout, the most common European salmonid, and to ensure that no toxic acidic episodes occur to this species during the year.

In North America, the effects of acidification on biota have been historically related to pH rather than alkalinity or acid-neutralizing capacity. Research on pH tolerance of a wide range of aquatic organisms has shown that a pH>6 is required to maintain aquatic ecosystem functioning and protect both fish and other organisms (RMCC 1990, Environment Canada 1997, Jeffries and Lam 1993). Within a given region, lake pH has been empirically and theoretically related to alkalinity as an inverse hyberbolic sine function (Small and Sutton 1986) and this relationship has been used to equate the two variables for the purpose of critical load modelling (e.g., Jeffries and Lam 1993). The relationship between pH and alkalinity for the Athabasca oil sands region was derived from a water quality survey conducted on lakes in the ALPAC forest management area (WRS 2000). Across these lakes, a pH of 6.0 is associated with an alkalinity of ~75 μeq/L. Accordingly, this value was chosen for ANC_{lim} in the Acid Deposition Management Framework for the Athabasca oil sands region (CEMA 2004) and has been applied in numerous studies (e.g., Gibson *et al.* 2010).

Only one estimate of the runoff is presented to avoid confusion. The yearly differences in runoff were presented and related to hydrological events.

31. *Improvements in RAMP browser interface-- Significant accessibility improvements could be achieved through minor software upgrades/fixes* (Ap B, Rec #13).

This has been completed for the new publicly available database on the RAMP website. The database can be accessed via an interactive map interface (i.e., spatial access) or directly through component-specific queries allowing the user to filter by station, year, season, variable etc. The database was made available online on December 21, 2010.

32. *[difficult to locate some flow records/stations] Check data availability for all years. Check that A and B flags are included in hydrometric records. Where appropriate, add 'not available' to past records where no information is available* (Ap B, Rec #14).

This has been completed for the new publicly available database on the RAMP website (www.ramp-alberta.org/data).

33. *[website weather stn data are limited especially south of Ft McMurray] Addition of a climate station south of Fort McMurray would be advantageous to allow for river-specific weather and hydrologic information to be monitored in the southern Athabasca region. More complete data records should be made available including information required to estimate evaporation and transpiration* (Ap B, Rec #15).

Agreed. This issue was discussed among members of RAMP Tech over the last year. A new climate station south of Fort McMurray will be installed in 2011.

2.3 WATER QUALITY

2.3.1 Comments and Recommendations by Dr. George Dixon

The following comments and recommendations were made by George Dixon to improve the Water Quality component of RAMP.

34. *RAMP could consider the establishment of a standing external review committee to comment on the activities of the Technical Program Committee* (Ap C, p. 2, para 1).

RAMP agrees that there is value in having a review committee for all regional monitoring that would ensure broader program linkages. This is a governance related issue, which will be addressed by the government monitoring initiatives.

35. *QA/AC: If a protocol for accuracy in data entry is in place it should be reported; if one is not in place it should be implemented* (Ap C, p. 2, para 3).

RAMP follows standard QA/QC procedures for data entry and data management, as outlined in the RAMP Design and Rationale document (RAMP 2009b) and annual technical reports. The water quality group has tasked the Implementation Team with the development of a formal, stand-alone QA/QC protocol for data entry and management, which will be reported during the next RAMP Tech meeting in fall 2011. AENV representatives will provide their QA/QC protocols to RAMP for consideration.

36. *I suggest that RAMP management clearly define who they want to inform (and about what) and develop a communications strategy to achieve their goals* (Ap C, p. 2, para 3).

A communications strategy has been developed in 2010 by the RAMP Communications Subcommittee, supported by external contract communication resources. RAMP has increased and improved its communication activity through several undertakings, including:

- Proactively issuing communication regarding release and availability of new information from RAMP, including posting of monitoring results and reports on the RAMP website;

- Implementing processes, including providing direct means to contact the communication group, in order to better respond to time sensitive media requests;

- Identification of a spokesperson for RAMP; and

- Proactive preparation of responses to journal articles or newspaper articles related to oil sands activities and the environment.

37. *The combination of stressor based and effects based monitoring, as undertaken by RAMP, is the best strategic approach, and significant progress in integrating the two (common stations for water, sediment and biological endpoints etc.) has been made over the last five years. Having said that, there is still progress to be made as outlined on page 3-5 of the design document (RAMP 2009b). This should be given priority* (Ap C, p. 3, para 1).

RAMP continues to look at ways to increase integration of the various components. The measurement of the various RAMP components has been harmonized to the extent possible for existing locations. The measurement of the various components is harmonized to the extent possible for all new locations to RAMP, ensuring that all components of RAMP can be compared at the same spatial and temporal scales.

38. *I suggest that the RAMP program consider supplementing the current monitoring program with in situ assimilative devices (i.e. semi-permeable membrane dialysis (SPMD)) where appropriate to measures PAHs and metals* (Ap C, p. 3, para 4, p. 4, para 1).

RAMP is open to the use of passive-sampling devices for PAHs and metals in water, and is aware of recent deployment of different types of such samplers by AENV, EC, and academic researchers. Use of a method that could be used in a routine way to provide reliable, repeatable, quantitative data for surveillance monitoring is essential for integration of such protocols in a program like RAMP. In 2011, AENV is doing further method validation and refinement for the various passive-sampling devices they are currently using; RAMP would like to review these results before committing to wide-scale use of passive samplers. In the interim, RAMP agreed to add analysis of PAHs in grab samples of water from RAMP water quality stations in 2011 and 2012, at ultra-low detection limits (i.e., MDL ~5 ng/L). These results will provide additional information about appropriate methods and technologies to monitor PAHs.

39. *I note that for the sediment quality variables, pore water naphthenates are not included as one of the measured parameters. Since these are oil-sands chemicals of concern, what is the basis for exclusion? While I recognize that they are often considered too hydrophilic to partition to sediment, it might be appropriate to demonstrate that they are in fact not present at significant concentrations* (Ap C, p. 4, para 2).

This issue relates to groundwater-surface water connections and potential seepage issues around developments, as well as partitioning of sediment-borne chemicals to pore water. To date, groundwater issues have been considered to be outside of RAMP's monitoring mandate (the proposed Groundwater Management Framework being developed by AENV under the Lower Athabasca Regional Plan includes a substantial sampling component to examine these issues); other seepage-oriented programs may exist at specific development sites that could provide such data. Clearer understanding of groundwater influences on surface water will be pursued by the RAMP Implementation Team with groundwater experts for the 2011 Technical Report and onwards. Given uncertainties currently associated with naphthenic acids analyses (i.e., the development of appropriate laboratory analyses to detect NAs associated with oil sands processes), incorporation of naphthenate analyses in pore water would not likely be appropriate for routine surveillance monitoring at this time, but could be incorporated in future.

40. *Standard Operating Procedures (SOPs) of the analytical labs involved should be included in the RAMP SOPs Reports, particularly if different labs are being used for the analytical work through time* (Ap C, p. 4, para 3).

RAMP has received from its contract laboratories descriptions of analytical methods used for its analysis; wherever possible, Variable Method Values (VMV codes) are included in the RAMP database for each observation, which indicate the specific analytical procedure used (these are government-wide standards and codes for tracking analytical methods). RAMP will inquire further with its contract laboratories about receiving complete SOPs for each analysis, which will be included in the next version of the Design & Rationale document and posted to the RAMP website.

41. *If a change in analytical techniques takes place, there has to be some way of translating the results from the new method so that they can be compared to previous data for the purposes of trend analysis. One way of achieving this is (for the first year of introduction of the new method) to complete duplicate analysis using both the old and new methods, the presumption being that the relationship between the two could be used to convert the data to a common base. I suggest that this approach, or an appropriate equivalent, be added to standard procedures as the program moves forward* (Ap C, p. 4, para 4).

This has usually been the case for RAMP methodologies, and will be continued. For example in 2011, RAMP will analyze hydrocarbons in water using the existing TRH method as well as the CCME Four-Fraction method, with the goal of shifting to the CCME method fully in 2012, to be consistent with a similar method shift recently made by Alberta Environment. RAMP will formally add this transition protocol to its SOPs.

42. *I suggest that the current approach of undertaking selected studies on seasonal variability in water quality data to supplement the fall monitoring program be given increased priority. This may not, at least initially, require expanded sampling, but rather re-examination of existing data to answer the following question. Are estimates of the potential for cumulative impact on aquatic biota based on fall data the same as estimates based on data for other seasons?* (Ap C, p. 5, para 1).

RAMP agreed to examine increased seasonal sampling. In 2011, this will include compilation and analysis of existing regional water quality data that are collected seasonally in the region to assess seasonal variability. Datasets to be examined will include continuous and monthly water quality monitoring from the Muskeg River, monthly data from the Athabasca River, and any other high-frequency water quality datasets that can be identified from RAMP members through a letter of request that RAMP Tech sent out this spring. RAMP will also install continuous conductivity meters in two tributaries (the lower Steepbank and lower MacKay) from spring to fall 2011, to examine temporal variability in water quality. Based on the findings of that study, RAMP will add ten times/year water quality sampling in 2012—coordinated with hydrology measurements—in the lower Ells and MacKay rivers, to assess temporal variability. Further expansion of within-year sampling at individual locations will be deferred until more direction is provided from the government monitoring initiatives.

43. *For both river systems and lakes, efforts should be made to assess the number of baseline sites that would be appropriate given the current number of test sites and increase the number of reference sites to that level* (Ap C, p. 5, para 2, p. 5, para 3).

RAMP will continue to focus on maintaining and increasing the number of baseline stations in the program. However, the appropriate ratio of test:baseline stations must be based on the questions being asked and the analytical design adopted for RAMP. As such, beyond continuing to add baseline stations in 2011 and 2012 as previously planned, this is a larger redesign issue that will be re-examined once the government monitoring initiatives are complete and RAMP is provided more direction.

44. *Acid-Sensitive Lakes - I suggest that the data be fully analyzed to determine the strength and weaknesses of fall sampling and that these be clearly stated. Stated another way, fall sampling has to be more fully justified* (Ap C, p. 7, para 2).

The ASL component lakes are sampled in the fall. The rationale for a fall sampling program was that the water chemistry of the lakes would be stable at this time of year, and the lakes would be fully mixed (stratification is broken down). In order to address issues of the timing of the sampling of the ASL component lakes and the possibilities of a spring acid pulse, Alberta Environment (AENV) conducted a seasonal study of ten representative ASL component lakes for five years. The results of this study were summarized in the 2008 RAMP Technical Report (RAMP 2009a) and have been referred to in the

2010 Technical Report (RAMP 2011). The AENV study showed that much of the water in these shallow lakes (median depth 1.8 m) freezes during the winter and the lake chemistry changes dramatically. Large decreases in pH and increases in Gran alkalinity are observed during the winter, accompanied by low oxygen levels and high levels of sulphide (strong sulphide odour). In spring, the lakes recover from the low pH and high alkalinities as the ice melts and oxygen is re-introduced. A spring pulse in acidity, seen as a subtle decrease in pH or alkalinity, could not be detected in the spring when all these events were occurring. The major change in pH during the recovery from anoxia in these shallow lakes (an increase) was in the direction opposite of that expected from a spring acid pulse (a decrease).

Despite the failure to detect a spring acid pulse in the AENV study, RAMP agrees that a spring acid pulse may still occur, especially in the deeper lakes where the winter oxygen deficit is less pronounced. The RAMP Implementation Team was asked to develop a pilot-scale sampling plan to attempt to detect a spring pulse in acidity, which will be conducted in 2012 and included in the 2012 RAMP Technical Report. The proposed study will use hydrosonde probes placed in a subset of the ASL component lakes before the melt. pH and other relevant variables will be measured throughout the spring season.

2.3.2 Comments and Recommendations by Dr. Monique Dubé

The following comments and recommendations were made by Monique Dubé to improve the Water Quality component of RAMP.

45. *Demonstration that existing baseline stations are outside of aerial contamination, in light of the recent Kelly et al. (2009) work* (Ap D Deficiencies).

Ongoing AENV and other studies of aerial deposition will help to address this question specifically with respect to PAHs and metals; results from these other studies will be used to help define the development footprint and to define baseline stations. The regional occurrence of aerial deposition requires re-examination of the appropriate definition of "baseline"—that is, must a station be completely pristine (i.e., unexposed to any human influences) to be considered baseline, or is it a matter of gradient or degree of exposure? The RAMP Tech Supergroup (focusing on water quality, benthic invertebrates, sediment quality and acid-sensitive lakes monitoring) agreed that definitions of baseline and test should be based on expectation of degree of impact through defined impact pathways, and that ongoing deposition studies would assist in clarifying these pathways and their relevance. Based on results of those studies, and definitions developed from the government monitoring initiatives, definitions of test and baseline in RAMP will be assessed and reconsidered as necessary.

46. *The entire basis of an effects based design is a defensible baseline to compare test sites to. There has been little characterization or justification of regional baseline for reviewers to assess. A clear picture of variation locally to regionally has not been established. Baseline should be established on a local, parameter-specific basis and compared to regional baseline. Variability also requires quantification inter-annually and by season* (Ap D Deficiencies).

Further explanation and discussion of regional baseline calculations and characteristics was provided in Section 6.2 of the 2010 Technical Report (RAMP 2011) and Appendix A2 of this report. In the absence of regional or site-specific water quality objectives/thresholds provided by regulators or regional organizations such as CEMA, RAMP has developed a set of regional water quality benchmarks to address these two issues in its own assessments, from data collected by RAMP at baseline stations since 1997. These regional baseline ranges are intended to represent the range of natural variability in water quality in the region, for use in screening RAMP water quality data collected at both baseline and test stations. The intent of these benchmarks is to identify regionally meaningful changes in water quality.

The characterization of regional baseline conditions will continue and be built upon in the 2011 report. Given the confusion regarding screening water quality against regional baseline characteristics, the Implementation Team will consider eliminating the use of regional baseline data in the 2011 Report, and instead just focus on the other analyses conducted, including screening to published benchmarks and site-specific historical data using trend analyses.

47. *Explanation is required as to why some companies are within RAMP and others are not* (Ap D, p. 11, para 7).

In recent years, many of the approvals for new projects have stated that the company must be a part of a regional aquatics monitoring program. Since then, many new members have joined RAMP and there are now very few companies that are not involved with RAMP in the lower Athabasca region (Table 6).

Table 6 Members and non-members of RAMP in the lower Athabasca region[1], as of May 2011.

RAMP Member	Non-Member
Suncor Energy Inc.	Petrobank Whitesands
Syncrude Canada Ltd.	Statoil Canada Kai Kos Dehseh
Shell Canada Energy	Connacher
Canadian Natural Resources Ltd.	
Imperial Oil Resources	
Nexen Inc.	
Total E&P Canada Ltd.	
Husky Energy	
Hammerstone Corp.	
ConocoPhillips Canada	
Devon Energy Corp.	
MEG Energy Corp.	
Dover Operating Corp.	
SilverBirch Energy	
Cenovus Energy	
Japan Canada	

[1] Companies in these lists are in operational phases and have received approval. The list does not include projects or companies that are still conducting EIA and in the application phase.

48. *Combine data: Water quality data stored in the RAMP data base is only RAMP data. This division of data is archaic and limiting to the understanding of change in the basin and the ability to manage it* (Ap D, p. 12, para 1).

RAMP does not have ownership rights or control over data collected by others and; therefore, cannot manage and provide these data for others via its database. RAMP agrees that greater availability and coordination of regional data from all sources would be very beneficial to all parties. Recently, RAMP Tech has asked industry and government to identify what water quality monitoring they undertake, including the location and frequency of sampling and variables measured, in an effort to gain a better understanding of the monitoring in the region.

49. *Hydrometric monitoring should be alarmed or automatic notification and a maximum allowable response time specified* (Ap D, p. 12, para 4).

The comment refers to the Climate and Hydrology component. Agreed, this can be done through a telemetry network whereby the Implementation Team will receive email updates if a station goes down. The establishment of a telemetry network is ongoing and will be implemented at all hydrology stations in 2012.

50. *Variability must be shown on report figures to have a realistic limit on interpretability* (Ap D, p. 13).

Variability among years and among stations is shown in water quality graphs and tables in RAMP annual reports. Additional information was provided in the 2010 Technical Report (RAMP 2011) regarding variability within and among regional clusters over time and space.

51. *Need to know how often results are different from local reference, as well as sub regional reference, as well as inter-annual viability. Significant effects can exist within the range of natural variability, and are important for detecting cumulative effects* (Ap D, p. 13).

This is related to responses provided for recommendations #46 and #50. Additional information was provided in the 2010 Technical Report (RAMP 2011, p. 3-67 to 3-68) regarding variability within and among regional clusters over time and space.

The following sections from RAMP (2011) are reproduced below as they provide supporting information for the response to this recommendation.

Comparison to Regional Baseline Concentrations

To allow for a regional comparison, untransformed data for 15 of the 21 water quality measurement endpoints from all *baseline* stations sampled by RAMP from 1997 to 2010 (fall only) were pooled from each cluster of similar stations. Descriptive statistics describing *baseline* water quality characteristics for each group were calculated; for each water quality cluster, the 5th, 25th, 50th (median), 75th, and 95th percentiles were determined for comparison against station-specific data. The number of observations varied by cluster for each of the fifteen selected water quality measurement endpoints. The median rather than the mean was used as an indicator of typical conditions; given water quality data are characteristically positively skewed. Regional *baseline* ranges did not include, and were not applied to lakes sampled by the RAMP Water Quality Component in 2010, to address concerns expressed by the RAMP 2010 Peer Review (AITF 2011) in combining water quality data from streams and lakes in regional *baseline* ranges.

Data for the fifteen selected water quality measurement endpoints were presented graphically in the context of relevant regional variability by presenting data for each station for all years of sampling by RAMP to allow assessment of any temporal trends. Where possible, stations located upstream and downstream on specific watersheds were presented together, to allow assessment of any differences in values or trends between upstream/downstream locations.

Development of Regional *Baseline* Concentrations Descriptions of regional *baseline* water quality conditions were developed from existing data collected by RAMP since 1997 from *baseline* locations throughout the study area. These ranges of regional natural variability in water quality were used as one method of screening water quality observed at all stations in fall 2010, to assess whether water quality conditions at the time of sampling were similar to, or differed from, those typically observed in the region.

This analytical approach is similar to that of the Reference Condition Approach to biomonitoring (Bailey *et al.* 2004), also is used in the RAMP Benthic Invertebrate Communities component, and incorporates elements of control charting (Morrison 2008), which also is a feature of RAMP Benthic Invertebrate Communities and Acid-Sensitive Lakes components. This approach is more fully described in the RAMP Technical Design and Rationale document (RAMP 2009b). It also shares similarities with CCME's prescribed approach for developing site-specific water quality objectives (SSWQOs), which uses the 90th percentile of upstream water quality observations to define benchmarks for assessment of water quality in a given waterbody, typically downstream of some kind of development (CCME 2011).

Multivariate data analysis was used to develop descriptions of regional *baseline* water quality that were then applied to water quality measurements from *baseline* and *test* stations. In this approach, water quality data from all RAMP *baseline* water quality stations from 2002 to 2010 were pooled using cluster analysis. Cluster analysis was applied to the RAMP water quality variables. Similar approaches to consolidation and analysis of large water quality datasets are common in the water quality assessment literature (e.g., Boyacioglu and Boyacioglu 2010, Astel *et al.* 2007, Singh *et al.* 2004, Jones and Boyer 2002, Güler *et al.* 2004).

52. *Assess contamination of PAHs due to air emissions and describe or illustrate if baseline stations are inside or outside of the McMurray Geologic Formation. A review of the detection limits for PAHs is also required* (Ap D, p. 13).

It is possible to plot the location of the McMurray formation on a map relative to the location of sampling stations; however, there are other factors that influence this analyses, including details regarding where the formation surfaces (i.e., outcrops, surface exposure etc) vs. where it is buried at depth, the heterogeneous nature of formation regarding concentration of oil/hydrocarbons, influence of groundwater seeps., etc. These issues (and others) all influence whether a station is actually "exposed" or directly influenced by the formation rather than its geographical location.

As stated in the response to Recommendation #38, RAMP agreed to add analysis of PAHs in grab samples of water from RAMP WQ stations in 2011 and 2012, at ultra-low detection limits (i.e., MDL ~5 ng/L). These results will provide additional information about appropriate methods and technologies to monitor PAHs.

53. *Now that an effects based monitoring approach has been adopted by RAMP as recommended in 2004 it is critical that the comparison between test and baseline sites is valid to detect a change in time or space if a change exists. Reviewing the water quality component emphasizes the need for refining / justifying / and significantly improving how the background/baseline is established* (Ap D, p. 6, para 5).

Agreed. As stated in the response to Recommendation #45, ongoing AENV and other studies of aerial deposition will help to address this question specifically with respect to PAHs and metals; results from these other studies will be used to help define the development footprint and to define baseline stations. The regional occurrence of aerial deposition requires re-examination of the appropriate definition of baseline, that is, must a station be completely pristine (i.e., unexposed to any human influences) to be considered baseline, or is it a matter of gradient or degree of exposure? RAMP Tech agreed that definitions of baseline and test should be based on expectation of degree of impact through defined impact pathways, and that ongoing deposition studies would assist in clarifying these pathways and their relevance. Based on results of those studies, and definitions developed by the government monitoring initiatives, definitions of test and baseline in RAMP will be assessed and reconsidered as necessary.

54. *As was recommended in 2004, activities of other monitoring programs and studies by government, academia and industry must be integrated and reported along with RAMP's considering the significance of the development in the region* (Ap D, S.4.2 Linkages).

Integration of all other monitoring programs in the region is beyond the scope of RAMP; however, this issue is currently being evaluated by government monitoring initiatives. RAMP's role within such an integrated framework will be strongly influenced or dictated by these initiatives.

55. *Provision of the actual percentage of change (rather than >10%) for each watershed is recommended up front in executive summaries. Annual overlay maps required* (Ap D, S. 4.3).

The grouping of watersheds based on a range of percent land change (i.e., <5%, 5-10%, >10%) is summarized in the executive summary of each annual technical report. The actual percent land change for each watershed is documented in Chapter 2 of the technical report and extracted from the 2010 Tech Report into Table 7 in this document.

Table 7 Percent of total watershed areas with land change in 2010.

Watershed	Total Watershed Area (ha)	Watershed Area with Land Change (%)						
		Focal Projects		Other Oil Sands Projects in RAMP FSA		Total		Watershed Total (%)
		Not-Closed Circuited (%)	Closed-Circuited (%)	Not-Closed Circuited (%)	Closed-Circuited (%)	Not-Closed Circuited (%)	Closed-Circuited (%)	
Minor Athabasca River Tributaries	160,730	5.35	16.91	-	-	5.35	16.91	22.25
Muskeg	146,000	3.53	8.26	-	-	3.53	8.26	11.79
Steepbank	135,491	2.98	0.32	-	-	2.98	0.32	3.30
MacKay	557,000	0.24	0.08	-	-	0.24	0.08	0.32
Tar	33,261	4.44	17.65	-	-	4.44	17.65	22.09
Calumet	17,354	0.20	1.03	-	-	0.20	1.03	1.23
Firebag	568,174	0.69	0.05	-	-	0.69	0.05	0.73
Ells	245,000	0.32	0.07	-	-	0.32	0.07	0.38
Christina	1,303,805	0.25	0.02	0.10	0.03	0.35	0.05	0.40
Hangingstone	106,641	-	-	0.01	0.04	0.01	0.04	0.05
Mills Creek	890	5.31	23.31	-	-	5.31	23.31	28.62
Shipyard Lake	4,047	13.48	79.26	-	-	13.48	79.26	92.75
Fort Creek	3,193	61.57	0.93	-	-	61.57	0.93	62.50
Horse	215,741	-	-	0.13	0.05	0.13	0.05	0.18
McLean	4,712	1.77	23.42	-	-	1.77	23.42	25.19
Original Poplar	13,856	1.21	2.22	-	-	1.21	2.22	3.43
Upper Beaver	28,711	2.77	6.72	-	-	2.77	6.72	9.48
FSA Total	**3,544,606**	**0.91**	**1.51**	**0.05**	**0.01**	**0.95**	**1.53**	**2.48**

The change in land use over time can be viewed in the map interface of the publically available website (http://www.ramp-alberta.org/data/map/).

56. *It is critical for a common understanding of where test sites are located and the development related activities they are exposed to as well as their location relative to the Fort McMurray Formation. Reviewers need to verify and understand the exposure conditions to determine the adequacy of the monitoring program design. Every year a map or series of maps are absolutely required that overlays this information* (Ap D, p. 13, p. 5, para 3, p. 10, para 2).

As stated in the response to Recommendation #52, it is possible to plot the location of the McMurray formation on a map relative to the location of sampling stations; however, there are other factors that influence this analyses, including details regarding where the formation surfaces (i.e., outcrops, surface exposure etc) vs. where it is buried at depth, the heterogeneous nature of formation regarding concentration of oil/hydrocarbons, influence of groundwater seeps., etc. These issues (and others) all influence whether a station is actually "exposed" or directly influenced by the formation rather than its geographical location.

Alternatively, for fish, liver MFO induction could also be used to evaluate exposure of stations/reaches to inducing agents such as hydrocarbons. Environment Canada has done some liver MFO induction in tributaries to the Athabasca River. Once data from studies conducted in 2009 and 2010 are available from Environment Canada, these results could help establish which sites are beyond the influence of inducing agents related to the formation.

57. *Increase water quality sampling in the mainstem Athabasca River* (Ap D, p. 13, p. 9, para 3).

The draft federal water quality monitoring plan also recommends increased sampling in the Athabasca River mainstem. RAMP is open to increased sampling in the Athabasca River mainstem and looks forward to better understanding if RAMP will play a role in implementing the federal monitoring plan.

58. *Reduce variability in the water quality sampling program. Clustering lakes and streams together for one group of baseline stations is ecologically disastrous and is absolutely inflating the variation of the natural condition to which test sites are being compared. Variability also requires quantification inter-annually and by season* (Ap D, p. 13, pg 7, para 4, p. 7, para 5).

Various additional analyses of regional-baseline data were undertaken for the 2010 Technical Report (RAMP 2011) to examine the influence of lakes and other waterbodies on regional-baseline ranges. These analyses indicated that lake data were not inflating the range of natural variability in the single cluster in which they were included. However, lake data were removed from all regional-reference analyses in the 2010 report to address this concern.

RAMP water quality stations on the Athabasca River upstream and downstream of development (i.e., stations ATR-DC-E/W, upstream of Donald Creek and ATR-DD-E/W, downstream of Fort Creek) are monitored seasonally and monthly sampling occurs at two AENV stations on the Athabasca River (i.e., ATR-UFM, upstream of Fort McMurray and ATR-OF, downstream of the Embarras River) to look at inter-annual variability in the mainstem. A pilot study will be conducted in 2012 to look at the variability across months at a subset of tributary stations to determine the range of inter-annual variability. The tributaries will be selected once RAMP receives more information on other monitoring programs taking place in tributaries, to avoid overlap in monitoring activities.

59. *Accelerate NA analysis methodology. PAHs and NAs are two predominant contaminate classes of concern neither of which are being adequately or accurately quantified or measured. Reassess PAH monitoring in light of Kelly et al. (2009) findings* (Ap D, p. 13).

Agreed. RAMP supported method analysis in 2010 by collecting extra samples for use by various laboratories in method development, and will do so again in 2011.

PAHs will be incorporated into water quality analyses in 2011 and 2012, as described in the response to Recommendation #52.

60. *Calculate within year, season and system type (lotic/lentic) baseline by parameter before regional. Consider the longer temporal record; pre 1997* (Ap D, p. 13).

Additional water quality data comparisons were made in the 2010 Technical Report between AENV and RAMP monitoring stations on the Athabasca River (RAMP 2011, p 5-33 to 5-39 and Appendix A2 of this document, which will be expanded in the 2011 report (including comparisons with longer-term [pre-1997] datasets, where possible). Lentic (lake) data were excluded from pooled data comparisons because of the small number of locations with data.

61. *Spatial comparison of the Water Quality Index requires clarification. Cannot compare spatially with different parameters and benchmarks. Clarification of method and application are required* (Ap D, p. 13, p. 8, para 3).

This is related to the response to Recommendation #46. Additional clarification of the Water Quality Index (WQI) and how it is applied was provided in the 2010 Technical Report (RAMP 2011, p. 3-75).

The following section from RAMP (2011) provides supporting information for the response to this recommendation.

Water quality at each RAMP monitoring station in fall 2010 was summarized into a single index value, ranging from 0 to 100, using an approach based on the CCME Water Quality Index. This index is calculated using comparisons of observed water quality against user-specified benchmark values, such as water quality guidelines or background concentrations. It considers three factors: (i) the percentage of variables with values that exceed a given user-specified benchmark; (ii) the percentage of comparisons that exceed a given user-specified benchmark; and (iii) the degree to which observed values exceed user-specified benchmark values. A detailed description of the index and how it is calculated is found at http://www.ccme.ca/ourwork/water.html?category_id=102. Its specific application to RAMP is described below.

Index calculations for RAMP water quality data used regional baseline conditions, as the benchmark for comparison. Specifically, individual water quality observations were compared to the 95th percentile of baseline concentrations (for the appropriate water quality station cluster) for each water quality variable.

Variables included in the calculation of the water quality index included all RAMP water quality measurement endpoints with the exception of total nitrogen, which was excluded because of autocorrelation with nitrate+nitrite and ammonia, both of which were included in index calculations. Index values were calculated for all baseline and test stations. Calculation of water quality index values for all stations sampled by RAMP in fall since 1997 (n=423) yielded index values ranging from 76.3 to 100.0. It should be noted that historical index values calculated for specific observations may change annually, given 95th percentile values for individual variables included in the index may change with addition of new baseline data to the RAMP data record.

Water-quality-index scores were classified using the following scheme:

- 80 to 100: Negligible-Low difference from regional baseline conditions;
- 60 to 80: Moderate difference from regional baseline conditions; and
- Below 60: High difference from regional baseline conditions.

This classification scheme, based on similarity to regional baseline conditions, differs somewhat from that used by CCME to classify water quality based on water-quality guidelines. Specifically, only three categories were used (versus five used by CCME), to ensure consistency with classification schemes used for other RAMP components. A classification of a "Negligible-Low" difference from baseline, corresponds with CCME guideline-based index classes "Good" and "Excellent"; RAMP classification of a "Moderate" difference from baseline generally corresponds with CCME class "Fair"; and RAMP classification of a "High" difference from baseline corresponds with CCME classes "Marginal" and "Poor". Although the CCME index is typically calculated using comparisons against water quality guidelines, it is customized for each station where it is applied to suit local conditions and concerns, and the use of regional norms as benchmarks, as is done by RAMP, is an appropriate use of this index (Government of Canada 2008, S. Pappas, Environment Canada, *pers. comm.* 2009).

Because the WQI, as applied by RAMP incorporates regional-baseline data, the WQI will be discontinued in future RAMP technical reports if the regional-baseline approach is also discontinued and the analysis will instead focus on screening to published benchmarks and site-specific historical data.

62. *Impact criteria used in EIAs and in RAMP must be 1) consistent and 2) must be tied to some level of decision or action both in future EIAs as well as for RAMP. Relate those in EIAs to RAMP and vice versa (WQ)* (Ap D, p. 8, para 4).

A summary of potential effects and related variables was summarized in the RAMP Design and Rationale document (RAMP 2009b). This assessment formed, in part, the basis for selecting specific monitoring endpoints and effect sizes in RAMP. However, given the differences in criteria and thresholds across oil sands projects, it is difficult to fit the monitoring to all EIA predictions. This issue needs to be resolved by Alberta Environment given they are responsible for developing the terms of reference of EIAs.

63. *Justify rationale for blanket use of mouth sampling stations as the watershed "cumulative effects" indicator stations. In the absence of understanding where monitoring stations are relative to development activities, which stations lie outside of or within the McMurray Geological Formation, and how far monitoring stations are from the mouth relative to other non-RAMP activities, tributaries, changes in surficial geology, etc., blanket use of mouth sites to assess cumulative effects requires better justification* (Ap D, p. 13, p. 9, para 3).

Use of river-mouth stations is typically only one means of assessing cumulative effects of development on a watershed, as most watersheds include multiple RAMP sampling stations, particularly those with multiple active or proposed projects (e.g., Muskeg, Steepbank). RAMP undertakes a mapping exercise each year to describe development activities so that the degree of disturbance and baseline and test status can be defined for each station.

64. *Tier decisions are required. At least a two level tiered response framework is required. Exceeding the first trigger would increase the frequency or detail of monitoring for confirmation and second level trigger investigation for casual identification. Further, the relationship between water quality monitoring and other monitoring components for triggering action requires identification* (Ap D, p. 9, para 5).

Annual monitoring results are submitted to AENV each year with flags on results that are not within baseline variability. RAMP supports a tiered approach to determine action and next steps and have in some cases conducted follow-up studies in watercourses where unusual results were observed (e.g., the 2009 slimy sculpin study in the lower Steepbank River produced low catch numbers. The study was conducted again in 2010 to determine if 2009 was an anomaly).

RAMP anticipates that an effects criteria framework will be developed under the LARP (Lower Athabasca Regional Plan) (triggers and limits) that will help to provide context to RAMP data.

65. *Harmonize components of RAMP. Harmonization within the aquatics program has improved but remains inadequate with respect to fisheries as well as with respect to linkages between water quality and quantity. All monitoring programs and studies within the region (including those being conducted by government agencies, academia and industry) must be integrated and reported considering the significance of the development* (Ap D, p. 13, p. 11, para 2, p. 11, para 5).

RAMP continues to harmonize sampling between components, most recently by conducting fish sampling in reaches where water, sediment and benthos are sampled. RAMP has discussed and recommended that all monitoring groups have an opportunity to share information and monitoring results in order to better integrate the work done in the region. It is expected that further harmonization with other monitoring in the region (i.e., groundwater, air quality) will be facilitated by AENV based on the government monitoring initiatives.

2.4 BENTHIC INVERTEBRATE COMMUNITIES AND SEDIMENT QUALITY COMPONENT

2.4.1 Comments and Recommendations by Dr. Kelly Munkittrick

The following comments and recommendations were made by Kelly Munkittrick to improve the Benthic Invertebrate Communities and Sediment Quality component of RAMP.

66. *The overall initiative needs to be tied together in a more transparent and public fashion. Monitoring or Surveillance? Components need to be tied together and need a range of initiatives that are linked with similar and overlapping components. There needs to be sufficient linkages to overlap the programs (i.e. so that baseline assessments are tied to indicators useful for EEM and performance, as well as are important in EIA and CEA evaluations)* (Addendum to Ap E, p 1, para 1,2).

This will be considered in the review of the overall scope of RAMP and other monitoring in the region that is currently being undertaken by government monitoring initiatives.

67. *Could try and time fall sampling events better in terms of ecological timing. Studies on how much variability occurs over the month period would be valuable* (Ap E, p. 2, para 3).

It is difficult to mobilize crews based on water temperature, discharge, etc., which can vary substantially within seasons among years. We are looking at relative importance of those variables to indices of benthic community composition, and would be able to make adjustments to indices, if it turns out adjustments will make a difference to an assessment. The Implementation Team has been tasked with undertaking a literature review to ascertain likely effects of within-season temporal variability on benthic data.

68. *Limited opportunity for finding reference sites in large rivers – focusing on smaller sites will (at minimum) double the number of sites available for each decrease in the order of the river under study; these could also be standardized between watersheds and add another layer to the analysis (1st order vs. 1st order, 2nd order vs. 2nd order….etc.)* (Ap E, p. 3, para 2).

A focus on smaller tributaries is a good idea, and will be re-examined once the government monitoring initiatives have provided further direction to RAMP. A field survey to establish new baseline stations on small tributaries to the Athabasca River will be conducted in 2011.

69. *The reports need to more explicitly specify what the normal ranges of variability are and how they are changing (or if they do change) from year to year. It is also critical to document and report the normal range of variability for each baseline site relative to the regional baseline to better understand how individual baseline sites are changing from year to year and contributing to increase or decreases in overall regional variability of the baseline condition* (Ap E, p. 5, para 2).

Natural causes of variation of measurement endpoints were further explored in the 2010 RAMP Technical Report (RAMP 2011, Section 6.4). Variations within baseline reaches and over time, were documented. Supporting information from RAMP (2011) for the response to this recommendation is provided in Appendix A2.

70. *Need to identify real triggers and how a range of concerns with differences based on the type of variability quantified. [See examples a-d provided]* (Ap E, p. 6, para 2).

The proposed triggers outlined in the review are similar to what RAMP currently uses for the Benthic Invertebrate Communities component (Table 5). More information regarding triggers in relation to quantified changes in benthic communities is provided in the 2010 RAMP Technical report (RAMP 2011, Chapter 3, p. 3-78to 3-79).

The following section from RAMP (2011) provides supporting information for the response to this recommendation.

Temporal Trends and Spatial Comparisons

Possible changes in benthic invertebrate communities were evaluated by comparing measurement endpoints in reaches designated as test to upstream baseline reaches and/or to pre-development conditions with analysis of variance (ANOVA). When necessary, the measurement endpoints were $\log_{10}$-transformed to meet assumptions of normality and homogeneity of variances. One-way ANOVAs were conducted for each benthic invertebrate community measurement endpoint with each reach-year (or lake-year, as appropriate) combination as the factorial variable. Planned linear orthogonal contrasts (Hoke *et al*. 1990) were then used to identify differences between baseline and test reaches (or lakes), between baseline and test periods, and differences in time trends between lower test reaches and upper baseline reaches (or lakes, as appropriate). In all cases, the comparisons were tested against the residual error of the overall one-way ANOVA.

Analysis of variance was used to test for variations over time for reaches or lakes that have been exposed to oil sands development since RAMP started in 1997. The ANOVA used variations within reaches (or lakes) to judge the significance of linear time trends. Linear contrasts were used to carry out the analysis of variance and to test the specific hypothesis:

- H_1: No linear time trend in mean values of measurement endpoints during the period of sampling.

RAMP has produced data for some reaches such as lower Jackpine Creek (JAC-D2) during both the *baseline* period for that reach and now when it is classified as a *test* reach. For those reaches, linear contrasts were developed that test the following null hypotheses:

- H_2: No difference from before to after exposure to oil sands development in mean values of measurement endpoints.

Where a *test* reach can also be compared with a *baseline* reach, evidence of an effect is derived from a change from before to after exposure to oil sands development, in the difference between *test* and *baseline* reaches. Linear contrasts were thus used to test the following specific hypotheses where the data allowed:

- H_3: No change from before to after exposure in difference between *baseline* and *test* reach mean values of measurement endpoint.

- H_4: No difference in linear time trends during the period of exposure to oil sands development.

The statistical power associated with these various hypothesis testing procedures is high with an error-degrees-of-freedom that is frequently > 100. The ability to detect differences is quite substantive, with the detectable effect sizes much less than the within-reach-standard deviation (i.e., small differences, Cohen 1977, Kilgour *et al.* 1998). Statistically significant differences; therefore, may be minor, subtle, or otherwise trivial. The nature of statistically significant differences was therefore examined to determine if the difference was consistent with a negative change in the benthic invertebrate community. A decrease in taxa richness, Simpson's Diversity, evenness and percent EPT would each be considered a negative change or difference. An increase or decrease in abundance could be considered a positive or negative change. Excessively high abundances (i.e., on the order of 100's of thousands of organisms per m^2) would be considered a negative change if the fauna was dominated by one or a few taxa (see Kilgour *et al*. 2005), and might be consistent with a nutrient enrichment effect (Lowell *et al*. 2003). In addition, non-effect-related variation was tested for significance. This was determined by testing the "remainder" variation, which is based on the remaining treatment sums of squares, left over after considering the specific effects-based contrasts. A significant "remainder" test indicates that there is a considerable amount of noise in the data and can put into question other contrasts that may be statistically significant, but that do not account for as much of the total variation (DFO and EC 1995).

All information collected by RAMP is, in turn, provided to the environmental regulatory authorities in support of ongoing assessment and the provision of various monitoring requirements stipulated in Alberta EPEA approvals for oil sands projects. In addition, the results of the monitoring program are evaluated on an annual basis by RAMP Tech for the purpose of refining future monitoring activities.

71. *It may also be possible to use WQ clustering as a basis for developing sub-regions or analytical units; as it is there are no opportunities for grouping the sites based on specific characteristics* (Ap E, p. 6, para 6).

RAMP has explored, in prior years and this year, factors that influence variation in indices of composition and found that habitat type is by far the most critical factor. Other factors, e.g., river size (width), and slope, etc., are not very important influences on variation in measurement endpoints in the RAMP study area. The data are going to be further assessed within and outside RAMP to identify modifying factors that will be subsequently used to develop a better "model".

72. *Harmonize components of RAMP (was either directly stated or implied by all reviewers)* (Ap E, p. 9).

RAMP agrees that there are still some watercourses where components are not harmonized. For all new stations added to the program, monitoring of all components is implemented; however, further work will be done to harmonize all existing stations.

73. *Incorporate a tier analysis. Analyze within river, and within year before regional. Need to increase the ability to detect reasonable changes (stated by others including M. Dube)* (Ap E, p. 9).

The RAMP Technical report now incorporates tiered analyses. In the 2010 report, benthic data from test reaches included (1) a spatial assessment within a watercourse (upstream vs. downstream, i.e., baseline vs. test, where feasible) using rigorous statistical hypothesis tests of differences, (2) a temporal (within or between years) assessment within a reach, and (3) comparison to regional baseline ranges of variation (RAMP 2011, p. 3-77 to 3-81).

74. *Increase the number of sites where development is anticipated in the future. Increase the timeline for site-specific reference data* (Ap E, p. 9).

RAMP attempts to collect three years of baseline data at all stations/reaches prior to any development. It is sometimes difficult to predict the three year time span given projects are started based on the market and the approval process for applications, which are often variable. In 2011, RAMP will attempt to establish new baseline stations/reaches in watercourses where development is not currently planned.

75. *Place replicate samples within riffle habitat units* (Ap E, p. 9).

In 1998, triplicate samples were collected in each of three riffles in three tributaries (Golder 1999). The triplicate samples were used to calculate within-riffle variance. As defined by Environment Canada's guidance for EEM programs for the pulp & paper and metal mining sectors (Environment Canada 2010), the number of replicates per station (or riffle in this case) should be enough to obtain estimates of measurement endpoints to within ±20% of their true value. Using this guideline, it was determined that a single replicate within a riffle would be adequate to estimate abundance, richness, diversity and percent EPT taxa. The unit of replication and spatial distribution of "stations" within reaches is consistent with the recommendations from Environment Canada's EEM guidance documents for river sites. A pilot study will be conducted in 2011 to document the within-station variation of depositional habitats and to verify the within-station variation in erosional habitats.

76. *In terms of cumulative effects there still needs to be some effort in the Athabasca River. Increase sampling in the mainstem Athabasca River. Consider mainstem riparian zone sampling with D-ring or kick nets of some type* (Ap E, p. 9, p. 3, para 3).

The Athabasca River mainstem was sampled in 1997 only, in areas upstream and downstream of the principal oil sands projects. Mainstem sampling was discontinued because the shifting sands were dominated by a few tolerant taxa, and there was a realization that the benthic invertebrate community in the mainstem would not react as quickly to development-related stressors as would communities in the tributaries (RAMP 2009b). In addition, the variability in the results was quite high given the shifting sand environment. Given the interest in the effects on the Athabasca River from oil sands development, a pilot study will be conducted in 2011 to determine an appropriate means of collecting benthos from the mainstem. Operators will be requested to provide benthic monitoring data that may exist for the mainstem, for possible inclusion in the RAMP technical report. It is expected that further information on this issue will be provided by the government monitoring initiatives.

77. *Reduce variability in the sampling program. The only defensible way to proceed is with an EEM design approach of multiple references and multiple exposed sites. The approach should have several components for assessing and reporting variability including 1) local baseline SD vs. regional SD on an annual basis; 2) the same on a seasonal basis; 3) the same on an inter-annual basis. At the very least, a study of variability should be an immediate priority based on existing benthic data, a comparison with literature, and supplemented if required with filed monitoring next fall* (Ap E, p. 9, p. 3, para 5, p. 4, para 4, p. 6, para 2).

We have explored, in prior years and this year, factors that influence variation in indices of composition and found that habitat type is by far the most critical factor. Other factors, e.g., river size (width), and slope, etc., are not very important influences on variation in measurement endpoints in the RAMP study area. The data are going to be further assessed within and outside RAMP to identify modifying factors that will be subsequently used to develop a better "model".

78. *Include a measure of variability on report figures* (Ap E, p. 9, p. 4, para 2).

The large amount of data makes it difficult to present the variability within a reach/year that is easily interpreted to the reader; however, RAMP will examine alternative ways of illustrating the among and within reach/year variability and will ensure that the within-reach/variability is in incorporated into figures in future reports.

79. *Incorporate tier decision for interpretation of differences. Need to know how often it is different from local references, as well as sub-regional references as well as inter-annual variability. Need to fall outside of normal 3 years in a row – should tier the triggers better than this – what are the consequences to monitoring exceedances. Would like to see at least a two level tiered response – exceeding first trigger would increase the frequency or detail of monitoring for confirmation (equivalent to extent and magnitude of EEM)and second level trigger investigation monitoring (equivalent to IOC in EEM)* (Ap E, p. 9, p. 6, para 4).

The RAMP Technical Reports now incorporate tiered analyses. In the 2010 report, benthic data from test reaches included (1) a spatial assessment within a watercourse (upstream vs. downstream, i.e., baseline vs. test, where feasible) using rigorous statistical hypothesis tests of differences, (2) a temporal (within or between years) assessment within a reach, and (3) comparison to regional baseline ranges of variation. It is anticipated that larger issues related to the management/decision framework will be addressed by the government monitoring initiatives. Currently, RAMP has identified impact criteria to evaluate observed changes in measurement endpoints. The results are also provided to the regulatory agencies for further evaluation.

2.4.2 Comments and Recommendations by Dr. Joseph Flotemersch

The following comments and recommendations were made by Joseph Flotemersch to improve the Benthic Invertebrate Communities and Sediment Quality component of RAMP.

80. *Recommend including citations where necessary* (Ap F, p. 2, para 4).

Agreed. The Implementation Team has tried to incorporate references and as much supporting information as possible in the 2010 Technical Report (RAMP 2011) and will continue to look for studies to provide context to the RAMP data.

81. *At a minimum, I would suggest some pilot samples be collected using alternative methods* (Ap F, p. 2, para 5).

This was initiated in 2010 with the use of a traveling kick-net to collect samples in the same location as Hess cylinder sampling to compare results. The comparison of the two gear types from this pilot study can be found in Section 6.3 of the 2010 Technical Report (RAMP 2011) and Appendix A2 of this report. There will be a pilot study in 2011 to determine an appropriate way of collecting benthic data from the mainstem Athabasca River.

82. *If changes in habitat are a concern, then I would suggest that a multi-habitat method be considered. Habitat should be sampled in proportion to their presence in the reach* (Ap F, p. 2, para 6).

The “dominant” habitat type is sampled in each reach and this dominant habitat type does not change over time.

83. *It is critical that the methods being used for sampling are better documented to support the collection of data through time as field crews change. More details on lab processing should also be included to facilitate replication of the methods by a different lab* (Ap F, p. 3, para 2).

RAMP developed a set of Standard Operating Procedures for each monitoring discipline at the start of the program. A copy of these SOPs is provided in an appendix of the RAMP Design and Rationale document (RAMP 2009b). These SOPs have been used by both consulting teams involved in the implementation of the program since 1997. In addition, field crews undertake training to familiarize each member with various sampling methods and protocols. Although a small number of members of a crew may change from year to year, an emphasis has been placed on ensuring that crew leaders are experienced in conducting monitoring field work and have undertaken RAMP sampling in the past. In addition, a majority of other crew members have typically conducted RAMP sampling on a consistent basis from year to year. As a result, the degree of change among crew members is often quite limited.

Regarding laboratory processing, RAMP has used the same taxonomic laboratory and procedures since the beginning of RAMP. The methods used are outlined in the SOPs and annual technical reports.

84. *Consider development of a predictive model for baseline areas* (Ap F, p. 4, para 1).

The "Reference Condition Approach" (RCA) in Canada and elsewhere is a sampling design that has an underlying notion that it is very difficult to find baseline sites that "exactly" match a test site in terms of flow volumes, flow velocities, water temperature, substrate texture, etc. When this is a concern, differences in composition between any two sites, say baseline and test, are perhaps due to natural phenomena. When natural variation is considered to be significant, a possible design involves sampling a high number of baseline locations and the relationships between measurement endpoints and flow volume, flow velocity, substrate texture, etc. among baseline sites are modeled. These models explain the background variation and can also be used, technically, to predict the "expected" value of measurement endpoints for which there are "reference" models.

RAMP was designed on the premise that the upstream baseline sites are an "adequate" match, and thus allow us to predict the condition of the benthic invertebrate communities that ought to occur in the downstream test sites. In some cases the upstream baseline sites are not good matches, which is acknowledged and discussed. RAMP also has baseline data for most of test reaches prior to becoming test. These baseline data are a "perfect" reference, unless natural short term (10-20 year) variations in climate are expected that drive variations in measurement endpoints.

In addition, RAMP has limited its selection of test sites such that they are generally quite similar in channel width (20 to 40 m across) and substrate, either depositional (sand) or erosional (cobble/boulder). In many ways, RAMP has a design that already incorporates a "model" such as the RCA approach by identifying substrate texture as a primary driver of the variation in measurement endpoints. Other potential influences have also been examined including: flow volume, flow velocity, and mean annual air temperature. In 2010, mean annual flow was shown to be related to taxa richness in erosional reaches. Like any RCA, RAMP is exploring the various sources of variation as it becomes possible to test for those effects. The influence of climatic conditions; however, is difficult to test with a limited number of years of data.

2.5 FISH POPULATIONS

2.5.1 Comments and Recommendations by Dr. John Post

The following comments and recommendations were made by John Post to improve the Fish Populations component of RAMP.

85. *All Fish Populations Activities - The program could be much more effective in addressing the RAMP objectives by sampling throughout the basin and including small-bodied less mobile species that better represent local toxicological condition* (Ap G, p. 10, para 3).

Agreed. The goal of the fish assemblage monitoring is to obtain greater spatial coverage across the region and to harmonize with the sampling design of the benthic and water quality components. The sentinel species programs on tributaries to the Athabasca River and on the Athabasca River focus on small-bodied, less mobile fish species and are conducted every three years. RAMP has been collaborating with a group from Environment Canada that has been looking at physiological indicators in small-bodied species captured during the RAMP fish assemblage monitoring program.

86. *Fish Assemblage Monitoring - There are several take home messages from the community metrics/integrated index pilot study: (a) many more sites are needed (b) bigger sites are needed so more individuals are captured (c) the program must identify maximum impact and minimum impact sites so the index is appropriately scaled along this gradient, (d) integrate the chemical, physical and hydrological and benthos components similarly and create a metric that uses the most discriminating of each of these components. This last comment could be addressed using canonical correlation to identify the most important axes for each of these components Field sampling for the development of this index should be stratified by habitat type and have at least 30 sites per strata. The best index would include sites outside the oil sands area to incorporate the maximum range of community metrics* (Ap G, p. 10, para 5).

In 2010, the fish assemblage pilot study focused on developing sampling protocols to obtain the required number of individuals, reach length and fishing effort where statistical comparisons of measurement endpoints could be completed over time. We agree that more sites are needed, and the fish assemblage monitoring will follow the benthic design at all reaches in September 2011. The two-year pilot study was focused primarily on protocol development and feasibility of the study, which was only conducted at a subset of reaches.

The pilot study was conducted to establish the amount of effort and number of fish required to estimate measurement endpoints that are statistically robust and allow for spatial comparisons across time. In 2010, each reach was divided into sub-reaches and from the results, it was estimated that at least five sub-reaches within a reach was required to obtain estimates of measurement endpoints (

Table 8).

Table 8 Number of sub-reaches required to obtain estimates of measurement endpoints of fish assemblages to be within 20% of the true sub-reach average.

Community Index	Minimum/ Maximum Value	Standard Deviation (SD)	Sample Size (n)
Total Abundance (# fish per m)	0.20	0.10	7
	0.60	0.40	12
Richness	3.50	1.20	3
	4.10	1.50	4
Simpson's Diversity	0.55	0.12	2
	0.65	0.14	2
Evenness	0.61	0.19	3
	0.80	0.15	2
Assemblage Tolerance Index (ATI)	5.60	0.75	1
	7.00	0.40	1

The fish assemblage monitoring will be integrated with the water quality, sediment quality and benthos components to provide supporting information. RAMP recognizes that harmonization with the hydrology component needs to be done. The Implementation Team will be working together in 2011 to identify hydrologic measurement endpoints that would be useful for biological and chemical processes.

87. *Fish Inventory and Fish Fence - Conduct age-structured demographic analyses wherever possible to estimate rates of growth (size-at-age), survival, recruitment* (Ap G, p. 12, para 2, p. 4, para 3, p. 6, para 4).

The collection of ageing structures will resume during the fish inventory surveys on the Athabasca and Clearwater rivers in 2011. In addition, archived fish ageing structures are continuously being submitted to a lab for ageing from all years to develop a more complete dataset.

In 2011, five ageing structures will be taken from each size class (200 mm to 700 mm, with 100 mm increments) in each reach in each season for the following species: goldeye, walleye, northern pike, longnose sucker, white sucker, and lake whitefish.

The collection of ageing structures for small-bodies species will take place during the sentinel species programs.

88. *Fish Fence - Continue the Muskeg River fish fence spawning survey in all years with sufficiently low spring discharge. Also extend the spawning fish fence program to other trap-able tributaries. Further investigation should be completed on spawning habitat, egg survival, fry survival, rearing habitat and toxicological assessments on early life history stages* (Ap G, p. 12, para 3, p. 6, para 6).

The Department of Fisheries and Oceans required three years of fish fence monitoring on the Muskeg River. This was completed in 2003, 2006, and 2009. An expanded fish fence monitoring program is currently outside the scope of RAMP, although we agree this information would be useful in obtaining more biological information on fish species that use the tributaries for spawning. If additional monitoring of spawning fish populations and spawning habitat becomes an activity under RAMP, alternative gear types (i.e., underwater camera, hoopnets, etc.) will be investigated given the difficulty of installing fish fences in all hydrologic conditions. Alternative gear types would allow for multiple tributaries to be monitored each year.

It is recommended that the use of fish fences or investigations/monitoring into spawning fish populations in tributaries be re-examined once the government monitoring initiatives are complete and RAMP is provided more direction. Directed studies on early life history stages of large-bodied species would obtain important biological information, which would provide context for data currently collected by RAMP.

89. *Technical Report Format - Organize the Technical Report as a cumulative living document in which data and analyses grow with each subsequent year. In addition, organize the analysis by topic rather than by river/site* (Ap G, p. 12, para 2, p. 7, para 3).

The format of the annual technical report focuses on specific watersheds within the oil sands region to allow stakeholders, industry and regulatory agencies easy access to monitoring information specific to watersheds of interest. This approach is also aligned with the movement towards harmonizing the monitoring components within a watershed. RAMP agrees, that a component-based approach to the technical report would allow for more thorough discussion and analysis of data within a component.

There are advantages to both approaches and as RAMP evolves, this issue will be addressed by RAMP Tech to determine the best approach to present the results.

90. *Sampling Design - The second philosophical shift that I will argue for is one that is front and centre in the Whittier and Hughes review (Hughes and Whittier 2008). This works needs to be done at a large number of clearly stratified and random sites, not at a small number of fixed sites* (Ap G, p. 13, para 2).

The development of a probabilistic design for monitoring is a major change in scope for all components of RAMP. There are concerns about whether a probabilistic design will still meet the site-specific sampling requirements given the random selection of sampling sites each year using this approach. Certainly there is currently a strong need to incorporate site-specific monitoring in support of monitoring requirements stipulated in EPEA approvals for individual oil sands projects. RAMP will re-examine this recommendation once government monitoring initiatives are complete and RAMP is provided more direction.

91. *A more informative approach involves development of mechanistic models of physical, hydrological and biological processes that control success of various species followed by application to putative impact sites to examine deviations in success. This requires several philosophical changes. First develop these models within the Athabasca basin at un-impacted sites. Second recognize that the various components of the ecosystem are linked in some cases strongly and in some weakly and coordinate sampling of all components including hydrology, chemistry, and biota, both spatially and temporally. Additional data on benthic prey abundance, assessments of reproductive effort and success of rearing juveniles in pristine sites would provide the models to assess impacts of development on success. Of course this is best done in the context of the fish community analysis (which I discuss below) and the sentinel species program should be imbedded within it* (Ap G, p. 13, para 2; p. 3, para 3; p. 8, para 2; p. 9, para 2; p. 12, para 1).

RAMP is continuing to work to harmonize all of the monitoring components on a spatial scale to maximize the collection of data from common stations/reaches within each waterbody. Water quality, sediment quality and benthos are typically sampled at the same sites and the fish assemblage monitoring will follow the same design, beginning in 2011. In 2009, fish assemblage work was conducted at the same reaches as the slimy sculpin program (i.e., sentinel species program). In 2010, Environment Canada collected fish from the RAMP fish assemblage reaches to look at physiological indicators. If possible, RAMP will continue to foster collaboration with Environment Canada should they continue their research and monitoring activities in the oil sands region. It is hoped that information collected by Environment Canada would assist RAMP in identifying additional measurement endpoints for evaluating the health status of fish populations.

RAMP is attempting to continuously add new baseline stations when possible to assess fish populations that are not influenced by development. These sites will also be used for the other monitoring components.

To date, RAMP has not considered developing specific mechanistic models for the purpose of long-term monitoring. However, it is recognized that RAMP is collecting data that could be used for this purpose. There are other initiatives undertaken by some industry members through their approval conditions focusing on developing a better understanding between fish population size/biomass in relation to key habitat variables/thresholds; however, these are site-specific in nature and not focused on impact scenarios. Given the use of mechanistic models does represent a shift in monitoring philosophy, RAMP will re-examine this recommendation once the government monitoring initiatives are complete and RAMP is provided more direction.

92. *Sampling Design - I recommend strongly that a whole watershed design with random (or at least regular) sampling along all waterways from low order streams to the mainstem Athabasca River be implemented for the hydrology, chemistry, benthos and fish components in an integrated design. A spatial data base such as this could indentify "hot spots" of concern in various measures, provide time series of whole basin measures and facilitate assessments of spatial and temporal cumulative effects* (Ap G, p. 14, para 2; p. 7, para 3).

The transition from a site-specific, control-impact design (with some regional stations) to a completely randomized design is a large change in scope of the program (with different monitoring objectives) not only for the Fish Populations component but for all monitoring components. RAMP will re-examine this recommendation once the government monitoring initiatives are complete and more direction is provided.

93. *Fish Assemblage Monitoring - The fish assemblage pilot suggests that useful aggregative metrics can be developed for long term assessments of basin wide impacts. Further work (i.e. at many sites that are stratified by habitat and random) should be explored to develop a family of sensitive indices. Embedded in these should be more detailed process oriented sentinel species approaches* (Ap G, p. 14, para 3).

A variety of measurement endpoints developed for the pulp and paper EEM program and for the RAMP benthic invertebrate communities component were used for the analysis of the 2010 fish assemblage data, and more measurement endpoints will be evaluated as the program continues to identify the most appropriate endpoints to assess changes over time and space.

In 2010, Environment Canada conducted sentinel species work within the RAMP fish assemblage monitoring (at four reaches) to obtain a complete assessment of the ecological (RAMP) and physiological (EC) conditions of the fish community. The results of this study have not been reported at this time; however, RAMP anticipates an opportunity to review this information once it becomes available.

94. *Literature Review - A general comment is that there is no reference in the Technical Report to the literature that is accumulating on impacts of oil sands chemicals on biota. If this literature has not been summarized for the RAMP team then I suggest that it be reviewed so that the RAMP team can be kept at the leading edge of the field* (Ap G, p. 15, para 1).

RAMP agrees that reference to existing literature should be included in the report and every attempt is made to incorporate information and data from other studies in the area, when possible. In addition, RAMP will continue to facilitate opportunities to share information between monitoring groups and will gather gray literature from RAMP members to use as context for current data.

95. *All Fish Populations Activities - RAMP has not developed a coordinated program assessing the impacts of environmental contaminants on critical life stages of organisms commonly used in physiological and toxicological assays. If the goal is to determine cumulative impacts, then we need to know where to look for them, and lab and field based experimental systems are a good start* (Ap G, p. 15, para 2).

Studies conducted by Environment Canada and other groups (i.e., CONRAD, universities, etc.) have conducted toxicological assays on fish to look at impacts of oil sands processes on fish populations. A summary of all research related to toxicity of PAHs and naphthenic acids in fish are provided in NRC (2010). RAMP has assisted with the collection of fish for these studies (i.e., provided fish to CONRAD for their work on naphthenic acids and to EC for their fish physiology work). RAMP will continue to collaborate and assist with this research to better understand potential impact pathways/responses. The results of the ongoing research will be valuable to RAMP to determine the best endpoints to continuously monitor in fish at a larger scale.

96. *Sentinel Species - Re-instate lethal sampling for sentinel species program. Consider the use of habitat models for target species when analyzing the data to assess whether there are deviations from the models in the test sites* (Ap G, p. 15, para 3).

RAMP determined that data from non-lethal sentinel species programs have not provided conclusive results and the Fish Subgroup decided to return to lethal sentinel species sampling in 2010 and continuing on a three-year rotation to minimize the potential impact of sampling on the fish populations.

Draft Habitat Suitability Index models have been developed for many of the fish species in the lower Athabasca region and these can be used as guidance when assessing results from the sentinel species programs. Habitat data are collected during the sentinel and fish assemblage programs.

97. *All Fish Populations Activities - The RAMP program has provided key data on which to develop a rigorous monitoring program but now needs to focus on stratified random sampling to appropriately characterize spatial and temporal variability in the Athabasca watershed. In fact, the focus needs to shift from the idea of variability in data to variability in processes* (Ap G, p. 16, para 2).

As stated in the response to Recommendation #92, the transition from a site-specific, control-impact design (with some regional stations) to a completely randomized design is a large change in scope of the program (with different monitoring objectives) not only for the Fish Populations component but for all monitoring components. RAMP will re-examine this recommendation once the government monitoring initiatives are complete and more direction is provided.

98. *All Fish Populations Activities - The approach that I will argue strongly for, here and elsewhere, is an assessment of rates and processes in relation to development rather than statistical assessment of patterns which often lead to little in the way of insight into the biology of the fish or potential development impacts on biology* (Ap G, p. 4, para 2).

RAMP has focused more on non-lethal monitoring activities to minimize the potential impact on fish populations from sampling and; therefore, has relied on statistical assessments of fish population characteristics rather than a detailed understanding of specific whole-organism metrics such as gonad development/fecundity, liver size, growth rates, etc. However, in 2010, RAMP transitioned back to lethal sentinel programs on tributaries and the Athabasca River mainstem, which will help the program to focus on measurement endpoints describing rates and processes. Data collected by EC in 2010 will also be reviewed to assess whether additional processes should be monitored over time.

99. *All Fish Populations Activities - Fish growth needs to be interpreted in a whole ecosystem context including info on density, prey abundance and flow regime which alters metabolic rates* (Ap G, p. 5, para 1).

RAMP agrees that a whole-ecosystem approach should be taken for tributaries, which will be facilitated by undertaking fish assemblage monitoring in 2011 (a pilot study was conducted in 2009 and 2010). The objective of the fish assemblage monitoring program is to assess changes in the fish assemblage of a watercourse in relation to changes in water quality, sediment quality, prey abundance (benthos), hydrology and physical habitat characteristics.

RAMP is continuing to try to integrate all of the components and will be assessing potential hydrologic measurement endpoints that would be useful from a biological perspective in the coming year.

100. *Fish Inventory - Incorporate historical data -- The Athabasca River correspondence analysis has data from the 1980's. Can this also be included in the time series plots to broaden the time horizon and as a baseline pre-development?* (Ap G, p. 5, para 2).

This was completed in the 2010 Technical Report for the Athabasca River where inventory data from 1987 to present was included in the analyses (RAMP 2011). Data prior to 1997 (pre-RAMP) is not considered baseline on the Athabasca River given there was development in the area at this time. Historical fish data from the ASRD FWMIS were used as baseline for comparison with data collected in 2009 and 2010 for the fish assemblage monitoring program. RAMP will continue to "mine" data from AOSERP reports and EIAs for the purpose of strengthening the baseline database for tributaries of the region.

101. *Fish Tissue - Individual vs. population-based analyses: Data presentation and analysis shown in figure 5.9-21 would be much more useful if it involved measurements in individuals rather than population means. It should be ANCOVA with size as a covariate to assess if [Hg] differs among sites and years given the underlying relationship with body size* (Ap G, p. 9, para 3).

This was completed in the 2010 Technical Report (RAMP 2011, page 5-402) analyzing fish mercury data across years and all lakes that have been sampled by RAMP. The results of the ANCOVA indicated that there are significant differences in mercury concentrations in each of the three species between lakes and years. Given the differences in characteristics of the lakes, including size, water chemistry and physical habitat, the differences in the accumulation of mercury in fish could be related to lake-specific characteristics.

The following section from RAMP (2011) provides supporting information for the response to this recommendation.

Most of the sampled lakes are in the upper (southern) portion of the RAMP RSA (i.e., Gregoire Lake, Christina Lake, and Winefred Lake) while some are on the eastern border of the RAMP RSA (Big Island and Gardiner lakes) and Lake Claire is to the north in close proximity to the Athabasca River Delta. Generally, mercury concentrations in lake whitefish and walleye from Net Lake are higher than all other sampled lakes.

Spatial comparisons using an ANCOVA for each species indicated that there are significant differences in mercury concentrations in fish between lakes ($p<0.01$ for all species). However, there are several factors that could influence the concentration of mercury in fish, including the size of the waterbody, the amount of vegetation or wetlands near the waterbody, the quality of the water (particularly the concentration of mercury), DOC and pH, as well as the amount of mercury found in the sediment (Heyes *et al.* 2000). When factoring in size of lake as a predictor of the mercury load in the system, there was no significant correlation between lake size and concentration of mercury in fish ($p=0.57$). Other information for these lakes including water quality and physical characteristics were not available and; therefore, could not be included in the analyses.

2.5.2 Comments and Recommendations by Dr. William Franzin

The following comments and recommendations were made by William Franzin to improve the Fish Populations component of RAMP.

102. *Sentinel species - The use of physiological indicators would provide earlier warning of potential effects of oil sands discharges/activities on fish populations* (Ap H, p. 3, para 1).

The use of physiological indicators is being evaluated through collaborative work with Environment Canada. RAMP and Environment Canada shared field resources in fall 2010 during the RAMP fish assemblage work and sentinel monitoring program. At that time, Environment Canada collected samples to evaluate a variety of physiological responses (e.g., MFO activity, steroid levels, etc.). The challenge will be to understand whether an observed physiological change translates to a potential effect on fish health. Elevated MFO activity is a good example whereby research has not identified a strong linkage between MFO induction activity and a decline in fish health, although it has been very useful as an indicator of exposure to inducing agents (e.g., PAHs, organochlorines). RAMP will continue to collaborate with

Environment Canada in 2011, if possible, and look for labs that can conduct physiological analyses on a commercial scale.

103. *All Fish Populations Activities - Genetic data to provide assessment of the presence/absence of local versus migratory stocks seems essential to all the purposes being served by fish population sampling* (Ap H, p. 3, para 2).

RAMP agrees that more knowledge on local versus migratory fish populations would be useful to obtain a sense of whether they are useful species for assessing localized changes. Other organizations, including ASRD, CEMA, and academia have looked at genetic characterization of certain species in the lower Athabasca region (Burke 2008); however, this type of work is currently outside of the scope of RAMP's annual monitoring activities. RAMP will use this research as supporting information for data collected annually.

104. *Fish Fence – recommend using mobile gears such as hoop nets to monitor spawning runs in tributaries instead so that spawning runs can be monitored in any hydrologic conditions and capture all variability* (Ap H, p. 4, para 1).

RAMP agrees that it is important to assess spawning fish populations in tributaries to determine which tributaries are being used and evaluate the variability in the strength/richness of spawning runs over time. Unfortunately, fish fences have not proven to be a reliable method to monitor spawning fish populations (cannot be used in all hydrologic conditions, labour intensive, costly). Accordingly, hoop nets or other types of gear could be considered to ensure successful deployment during high water years. Alternate approaches may prove to be less labour intensive (and more cost-effective) such that a greater number of tributaries could be monitored. Once direction from the government monitoring initiatives is provided, this topic will be re-examined by RAMP as it is currently outside of the scope of the program.

105. *All Fish Populations Activities - I believe this program should be rolled into a much more extensive probabilistic sampling design that would sample the whole river from below the rapids just above Fort McMurray to the major distributary channels of the Athabasca delta. Such a program could include many of the existing sample sites as well as many more in different parts of the river* (Ap H, p. 4, para 3).

As stated in the response to Recommendation #92, the transition from a site-specific, control-impact design (with some regional stations) to a completely randomized design is a large change in scope of the program (with different monitoring objectives) not only for the Fish Populations component but for all monitoring components. RAMP will re-examine this recommendation once the government monitoring initiatives are complete and more direction is provided.

A new baseline reach, upstream of Fort McMurray will be established in 2011. Historically, sampling has focused around mouths of tributaries that are adjacent to development. Planning and evaluation of fish sampling in the Delta will be conducted in 2011 with plans to implement sampling in 2012. It is a logistical challenge to sample in the Delta with boat electrofishing, alternate gear types will also be assessed to determine the best methods.

106. *Fish Inventory - Another aspect that would greatly improve the program is to collect samples for DNA analysis to detect possible presence of sub-populations of the KIR species* (Ap H, p. 4, para 3).

Related to the response to Recommendation #103, RAMP agrees that more knowledge on local versus migratory fish populations would be useful to obtain a sense of whether they are useful species for assessing localized changes.

Other groups (i.e., CEMA) have done DNA work on fish populations of the Athabasca River (Burke 2008). RAMP will review this information to assess whether it will be helpful for future monitoring.

This type of work is currently outside of the scope of RAMP; however, RAMP will use this research as supporting information for data collected annually.

107. *Fish Inventory/Sentinel Species - Sampling should perhaps concentrate more on the smaller species in the river such as Flathead Chub, Lake Chub, and Spottail Shiner etc as has been done for Trout-perch* (Ap H, p. 4, para 3).

In recent years, the fish inventories have been standardized to look at the whole fish assemblage; however, boat electrofishing does target larger-bodied fish species. The addition of the summer inventory has helped to increase the capture success of flathead chub, juvenile large-bodied species and small-bodied species that are resident to the mainstem. In addition, the fish assemblage monitoring on the tributaries primarily assesses small-bodied fish species.

Sentinel species programs on the mainstem and tributaries focuses on the small-bodied species. The selection of the target species was based on reconnaissance surveys to determine which species could be captured in adequate numbers to conduct a scientifically robust program. Given the differences in fish communities between rivers, it was not possible to target the same sentinel species on all rivers (i.e., slimy sculpin on the Steepbank and Muskeg rivers, longnose dace on the Ells River and trout-perch on the Athabasca River).

RAMP will review historical data from the Athabasca River to determine if other small-bodied species are captured in sufficient numbers to conduct analyses and determine the most appropriate season for this work.

108. *Fish Inventory - Sampling gear suggestions: In order to improve catchability of some smaller species in the mainstem, other gears in addition to boat electrofishing should be used such as bottom trawls and beach seines. These can be deployed from the same boat* (Ap H, p. 4, para 3).

Currently the inventories on the Athabasca and Clearwater rivers are conducted to assess fish species that are of importance to stakeholders; however, if the objective of the inventories changes to assess localized populations and the whole fish assemblage, then alternative gear methods will be evaluated.

RAMP recognizes that the inventory does not monitor the complete fish assemblage but focuses largely on larger-bodied species of interest to local stakeholders, particularly key indicator resource species. An emphasis has been placed on ensuring inventory methods are consistent from reach to reach and year to year to evaluate possible changes in KIRs (Key Indicator Resources) over time. The inventory has built on an existing database provided by Syncrude Canada developed from their inventory activities undertaken between 1987 and 1996.

RAMP conducted a pilot study in 2002 using different gear types on the Athabasca River; an evaluation of these results will be conducted if the objective of the Athabasca inventory changes.

An evaluation of historical small-bodied species data will be conducted in 2011 in the same manner as large-bodied species to look at trends over time and whether the current gear type is adequate to conduct statistical analyses on small-bodied species.

109. *All Fish Populations Activities - PIT Tag use: In order to obtain knowledge on age and growth of individual fish the program should consider mass marking of all caught and released fish with less damaging and less losable tags such as PIT tags* (Ap H, p. 4, para 4).

RAMP agrees that the use of PIT tags would be useful for tracking fish between tributaries and the mainstem as well as fish that are pit-tagged in other studies (company-specific programs, CEMA projects, etc.). As a first step, RAMP will purchase a scanner in 2011 to see if fish that have been tagged in other programs are captured during RAMP monitoring activities. The addition of pit tagging into the program will also be considered once RAMP receives further direction from the government monitoring initiatives.

110. *Athabasca Inventory - To see if there are differences among years you might have taken average catches by species by season for all years and tested 2009 against those averages to see if there were differences e.g. box and whisker plots with an average drawn across the graph* (Ap H, p. 5, para 1).

This was completed in 2010 Technical Report (RAMP 2011, page 5-81 to 5-92) and included in Appendix A2 as supporting information to this response. In the 2010 report, the average catch was presented across all years, including years prior to the start of RAMP, which provided some historical context. The data were also presented to indicate the periods of sampling across years (i.e., pre-RAMP: 1987 to 1996; RAMP: 1997 to 2004; and RAMP standardized sampling: 2005 to 2010). By calculating average catch for each period, the changes across years could be discussed taking into account any changes in fishing effort and sampling methods and identifying variability over time.

111. *Athabasca Inventory - Graphing and Analytical suggestions: Figures 5.1-27 and 5.1-31 to 5.1-36 should all be bar graphs or scatter plots because the data on the x axes are category variables not continuous variables (this was noted in some other areas as well). In all of these bar graphs you could have computed means and done tests of the data against means to see if any years were significantly different. Another approach with the inventory sampling is to do some species accumulation curves to determine if the number of samples is sufficient to develop an asymptote of species numbers. – a standard procedure to determine if sampling is sufficient. Another way to do that is to sub-sample the whole database using bootstrap techniques to determine for each species how many samples are required to reach an asymptote in the numbers caught* (Ap H, p, 5, para 1).

The graphical presentation of length-frequency distributions was changed to bar graphs in the 2010 Technical Report (RAMP 2011, see p. 5-83 to 5-89). An analysis of the Athabasca and Clearwater inventory data will be conducted in 2011 to determine the number of fish in a sample that is required to reach an asymptote in a species accumulation curve. Similar to the analyses for the fish assemblage monitoring on the tributaries in 2010 where we looked at the number of sub-reaches required to obtain estimates of species richness, this can also be completed for the inventory reaches. See chapter 6 in the 2010 Technical Report (RAMP 2011) and Appendix A2 of this report for the results of the fish assemblage monitoring program.

112. *Athabasca Inventory - Obviously the fish inventory work has to be approached differently than the test/baseline approach that is used for the sentinel species component, need to identify a baseline reach upstream of development and the town of Fort McMurray (reference-condition approach) to assess changes between baseline and test reaches* (Ap H, p. 5, para 2).

In spring 2011, RAMP will be conducting a reconnaissance survey for a baseline reach upstream of Fort McMurray to add into the Athabasca Inventory program. Evaluation of other potential sites on tributaries to support the fish assemblage and sentinel species monitoring will also occur in 2011. This issue has been discussed in great detail in the past, particularly with regard to the high level of mobility exhibited by many fish species of the lower Athabasca River and the ability to monitor fish that strongly reflect localized baseline and test conditions.

113. *Fish Tissue - I believe far too many metals are being analyzed far too frequently in the RAMP program. It is unlikely that metals in tissues of adult fish will change suddenly in one year so a three or five year rotation for tissue samples for metals would be more appropriate, including mercury. If some metal appears elevated in a composite then a more thorough sample might be analyzed* (Ap H, p. 5, para 3).

RAMP already conducts the Athabasca River/Clearwater River fish tissue programs on a three-year rotation and the regional lakes on a five-year rotation (as per the schedule of ASRD). RAMP will look at the historical data to determine which metals are always below detection limit to see if they could be eliminated from the suite of variables (*N.B.* It is also recognized that it may be important to continue to document the absence of specific metal burdens in fish rather than assume there has been no change over time, particularly given the wide-scale interest in RAMP). RAMP will also discuss which variables should be added to the suite to capture any oil sands-specific compounds that may affect fish health.

RAMP will continue to analyze fish tissue for metals and tainting compounds until further direction is received from the government monitoring initiatives.

114. *Fish Tissue - Unless the regional lakes are subject to potential effects of air pollution by being in the airshed of the Oil Sands probably RAMP should leave the sampling of regional lakes to the province and/or Health Canada. There are enough ways to use the consultants' time and the industry's funds without doing what should be a federally or provincially funded program* (Ap H, p. 6, para 2).

Sampling of lakes in Alberta for mercury does not fall under the jurisdiction of Health Canada but rather to Alberta Health and Wellness (AHW). Currently, mercury data from RAMP are provided to the Government of Alberta (GOA) to develop waterbody-specific fish consumption guidelines. These data are also provided to Health Canada recognizing their role in establishing consumption national guidelines for general and subsistence consumers. In other regions of Alberta, the GOA relies on other groups to collect mercury data to establish consumption guidelines. RAMP will continue to assist in this program until directed otherwise by the government monitoring initiatives given the activity is important to the stakeholders in the region.

115. *Sentinel species - It is possible that developing some suitable physiological indicators such as MFO/EROD assays might detect potential changes in Trout-perch biology ahead of the more physical measures including fecundity and egg size. It would be good if it were possible to develop an index of YOY year class strength perhaps at 1+ age (these will have passed the test of first overwintering). Alternatively analyses declining year class strength over time would provide an early indication of population level effects* (Ap H, p. 6, para 3).

In 2009 and 2010, Environment Canada looked at physiological indicators in trout-perch from the Athabasca River (in collaboration with RAMP in 2010). RAMP has transitioned from non-lethal to lethal sentinel species surveys given it is difficult to assess YOY (young-of-year) year classes in fractional-spawner species such as trout-perch.

As stated in the response to Recommendation #102, the use of physiological indicators is being evaluated through collaborative work with Environment Canada. RAMP and Environment Canada shared field resources in fall 2010 during the RAMP fish assemblage work and sentinel monitoring program. At that time, Environment Canada collected samples to evaluate a variety of physiological responses (e.g., MFO activity, steroid levels, etc.). The challenge will be to understand whether an observed physiological change translates to a potential effect on fish health. Elevated MFO activity is a good example whereby research has not identified a strong linkage between MFO induction activity and a decline in fish health, although it has been very useful as an indicator of exposure to inducing agents (e.g., PAHs, organochlorines). RAMP will continue to collaborate with Environment Canada in 2011, if possible, and look for labs that can conduct physiological analyses on a commercial scale.

116. *Sentinel species - I suggest that the Slimy Sculpin sampling program is not sampling YOY fish at all unless there are age data to back up the claim of YOY fish reaching 50mm in October* (Ap H, p. 6, para 4).

RAMP agrees that it is difficult to accurately assess the YOY size classes for non-lethal sentinel species; therefore, RAMP has transitioned to lethal surveys to look at traditional EEM measurement endpoints, including age, GSI, LSI, and weight-at-age of adult fish.

117. *Sentinel Species - Sculpins taken in the lethal sampling or incidentally to other programs should be aged by otoliths to develop an age length key for the species* (Ap H, p. 6, para 4).

RAMP agrees that the next slimy sculpin survey should be a lethal survey given the inconclusive results from the non-lethal program in 2009. Concerns have been raised regarding the frequency of RAMP sampling, and other monitoring/research programs, and the potential impact of these sampling events on sculpin populations of tributary habitats. This is particularly true for sculpin species given their territorial behaviour and limited home range (characteristics that make them a good sentinel species). A collaborative effort of lethal fish sampling in the tributaries is required among all groups to avoid impacting more sensitive species (i.e., sculpins).

118. *Sentinel Species - Another point about the sentinel program is that it seems that two baseline sites is unlikely to be sufficient for such a program, a minimum of three and better about 5 should be sought. This is especially true in this program where the differences between baseline sites are greater than the differences between the baseline and test sites* (Ap H, p. 7, para 1).

Agreed. Two new baseline sites will be established in 2011 and incorporated into the program for all components. These baseline sites, should they provide appropriate habitat for slimy sculpin, will be included in the next scheduled sentinel species program in 2012.

119. *Database - There is a great urgency to data mine the existing RAMP database to learn what has been found in much more detail, species by species. Ageing needs to be completed for all of the samples in a timely manner* (Ap H, p. 8, para 1).

As of December 2010, the RAMP database has been publicly available to allow researchers the opportunity to analyze the data. Archived ageing structures from the fish inventory programs are currently being aged. Typically, however, any ageing structures taken in a given year are analyzed and reported in the annual technical report (i.e., fish tissue, sentinel species, fish inventories, and fish fence programs).

120. *Fish Inventory - the Summer inventory on the Clearwater should not continue as in 2008-09 but rather be rolled into a broader Athabasca River assemblage monitoring program as was done on the smaller rivers in 2009* (Ap H, p. 8, para 2).

This recommendation is suggesting that the Athabasca and Clearwater fish inventories should be conducted using a probabilistic sampling design by randomly sampling reaches from the upstream end of the Clearwater to the Athabasca Delta. RAMP agrees that it would be worthwhile to look into a randomized design for these rivers given their size and the influence of many tributaries; however, this is a major change from the current scope of RAMP and will be re-examined once the government monitoring initiatives are complete and RAMP is provided more direction.

121. *Fish Inventory – Annual training of netters and consistent timing of hydrology in river to better standardize the fish inventory survey* (Ap. H, p. 13, para 2).

RAMP agrees that, in theory, the fish inventories should be conducted based on hydrology and not on calendar dates, although it is difficult to coordinate sampling programs based on hydrology given the change in climate (precipitation) that can influence the hydrologic conditions in a very short time period. Alternatively, results are often discussed in relation to the hydrologic conditions at the time of sampling. RAMP will be investigating potential hydrologic measurement endpoints that relate to biological processes, which can provide context for the fish data and aid our understanding of the observed variability in catch over time.

RAMP agrees that all netters on the boat should be trained prior to conducting the fish survey. RAMP will incorporate a training session for all new netters prior to conducting any fish inventory surveys.

2.6 ACID-SENSITIVE LAKES

2.6.1 Comments and Recommendations by Dr. Shaun Watmough

The following comments and recommendations were made by Shaun Watmough to improve the Acid-Sensitive Lakes component of RAMP.

122. *RAMP should explicitly acknowledge the fact that summation of results from EIAs may be somewhat arbitrary for certain parameters or attempts to standardize when possible* (Ap I, p. 2, para 2).

Agreed. The results from EIAs across projects are not standardized, which has made comparisons to EIA predictions somewhat difficult. RAMP will continue to refine the criteria to assess change but for some components (i.e., hydrology), a standardization of results has been used to assess change.

123. *In order to attribute source to any changes in lake chemistry an atmospheric model is needed* (*Ap* I, p. 6).

The development of an atmospheric model has not been completed by RAMP and is currently beyond the scope of the Acid-Sensitive Lakes component. An atmospheric model may help to interpret changes in lake chemistry and can be developed once further direction has been received from the government monitoring initiatives and their plans for collection of data from lakes sensitive to acidification.

124. *It would be much clearer if the RAMP Technical Design and Annual Report documents stated very clearly from the onset what measure of change is being assessed for each component and the rationale for this measure. It should also clearly explicitly state the reasons for other measures and how they will be used in the assessment* (Ap I, p. 6, para 1).

The criteria for determining change and the associated measurement endpoints are outlined on p. 1-21 of the RAMP 2010 report (RAMP 2011). A more thorough description of the criteria for each component is provided in Chapter 3 of the technical report (RAMP 2011) and in the RAMP Design and Rationale document (RAMP 2009b). See Table 5 in this report for the criteria to determine change, taken from RAMP (2011).

For components where no effects criteria have been developed by organizations tasked with developing management frameworks for monitoring programs (i.e, AENV, CEMA, LARP), RAMP has developed criteria based on other programs (e.g., Pulp and Paper EEM).

125. *The expected effects of oils sands operations outlined in the RAMP Technical Design and Rationale Report (Figs 2.1 and 2.2) could be improved as it is not entirely clear why the figures for surface mines and in situ projects are so different. These figures need to be revised* (Ap I, p. 7, para 1).

Figures summarizing the potential impact pathways (i.e., linkage diagrams specific to aquatic resources) provided in the RAMP Technical Design and Rationale document (RAMP 2009b) and annual technical reports were developed using information provided in the various EIAs for proposed surface mine and *in situ* operations. Differences in the linkage diagrams reflect differences in bitumen recovery activities between surface mine and *in situ* operations and the potential influence of these activities on surface waters. RAMP will review this information to ensure it is presented clearly in the Design and Rationale document.

126. *The rationale for the choice of lakes and how changes will be interpreted should be explicitly stated* (Ap I, p. 10, para 1).

The rationale for the choice of lakes was presented in the 2010 Technical Report (RAMP 2011, see page 3-53) and the Technical Design and Rationale document (RAMP 2009b). The following section provides information to support the response to this recommendation (RAMP 2011).

Initially the ASL component lakes were chosen to represent the most acid-sensitive (poorly buffered) lakes. These lakes were to serve as early-warning indicators of acidification. Over the years, the number of lakes sampled was expanded so that the ASL component lakes now cover a large range of lake types from soft to hard water and poorly buffered to highly buffered.

The sampling design for the ASL component reflects the natural geographic distribution of lakes within the study region, which limits the ability to apply a more statistically defensible stratified sampling design. The 50 lakes represent a majority of the lakes within the RAMP region that are worth sampling including a large number of little ponds that are less than 0.5 km^2 in area. Beaver ponds were not considered to be permanent lakes. There are very few lakes close to the major oil sands developments (Syncrude and Suncor) that are not clearly influenced by the developments themselves. The closest lakes are those lakes in the Muskeg River uplands and the area NW of Fort McMurray, which are well represented in the ASL component lakes. Low alkalinity lakes are represented in the upland areas (Birch Mountains, Stony Mountains). Lakes to the northwest and Northeast of the oils sands region in the Caribou Mountains and Canadian Shield are remote from emission sources of NO_xSO_x and were selected as baseline lakes.

Change is interpreted statistically on specific measurement endpoints through the use of:

- Analyses of variance;
- Mann-Kendall trend analysis; and
- Control charts.

Critical loads are calculated using a modified Henrikson critical load model. The estimates of critical load are interpreted largely as estimates of lake sensitivity to acidification.

127. *Seasonal data should be evaluated to see if a) they are appropriate for detecting episodic effects and b) if they are, is there any evidence that episodic effects occur* (Ap I, p. 10, para 3).

In order to detect episodic effects on the ASL component lakes, in particular a spring pulse in acidity, seasonal samples from ten representative ASL component lakes were collected for five years by AENV (as recommended in CEMA 2004). The results of these analyses were summarized in the 2008 RAMP technical report (RAMP 2009a) and have been referred to in the 2010 Technical Report (RAMP 2011). The AENV study showed that much of the water in these shallow lakes (median depth 1.8 m) freezes during the winter and the lake chemistry changes dramatically. Large decreases in pH and increases in Gran alkalinity are observed during the winter accompanied by low oxygen levels and high levels of sulphide (strong sulphide odour). In spring, the lakes recover from the low pH and high alkalinities as the ice melts and oxygen is re-introduced. It was impossible to detect a subtle decrease in pH or alkalinity in the spring, when all these events were occurring. The major change in pH during the recovery from anoxia in these shallow lakes (an increase) was in the direction opposite of that expected from a spring acid pulse (a decrease). A separate study on the Steepbank, Firebag and Muskeg rivers conducted in 2003 also failed to detect a spring acid pulse on these rivers that could be attributed to sulphates and nitrates deposited on the snow during the winter (WRS 2003).

Despite the failure to detect a spring acid pulse in the AENV study, the group agrees that a spring acid pulse may still occur especially in the deeper lakes where the winter oxygen deficits are less pronounced. The RAMP Implementation Team was asked to develop a pilot-scale sampling plan to attempt to detect a spring pulse in acidity. The proposed study will use hydrosonde probes placed in a subset of the ASL component lakes before the melt.

128. *An effort should be made to link the chemistry to the biology. If it is not in the RAMP program, impacts on Biology cannot be assessed* (Ap I, p. 11, para 1).

Environment Canada has the archived plankton samples from the lakes and is looking into having them analyzed for future research. RAMP is helping Environment Canada with funding options for this research.

129. *Confirm that it is change in chemistry of acid sensitive lakes that is of interest and acknowledge that this does not likely reflect the response of the lakes in the region. Is eutrophication an issue? – examine N as a stressor in its own right, not just as a component of the PAI* (Ap I, p. 15).

The overall goal of the ASL component is to determine whether acidification is occurring. Eutrophication by nitrates is not the main consideration. The ability of nitrates to be assimilated and used as a nutrient by plants within each lake catchment was accounted for by applying the approach adopted by CEMA and AENV whereby any nitrogen deposition in excess of 10 kg/ha/y and 25 % of the first 10 kg/ha/y deposited N were considered acidifying (CEMA 2008, AENV 2007). This assumption was incorporated in the calculations of net potential acid input (PAI) which, following standard practice, was compared to the critical loads of acidity calculated for each lake.

130. *Evaluate whether the current seasonal sampling captures snowmelt in the 10 ASL. If yes – compare the chemistry of this sample with other seasons. If no – do a spring survey of selected ASL* (Ap I, p. 15).

This is related to the response to Recommendation #127 on the possibility of a spring acid pulse in the ASL component lakes. We don't have the information available from the seasonal sampling program to determine when the snowmelt actually happened in each year. In other words, we can't detect a spring pulse from the data although, as stated above, we suspect that the changes in lake chemistry associated with the introduction of oxygen to these lakes in the spring will mask such an effect. The group agrees that a spring acid pulse may still occur especially in the deeper lakes where the winter oxygen deficit is less pronounced. A one-time study to address this concern is planned for spring 2012, using hydrosonde probes to be placed in a subset of the ASL component lakes before the melt.

131. *Between year comparisons of measurement endpoints over the entire population of lakes is really only for descriptive purposes and should not be used for assessing potential impacts of acid deposition* (Ap I, p. 16).

The simple one-way ANOVA is only one of the several techniques that we use in a "weight of evidence approach" to detect change attributable to acidification. The group agrees that the one-way ANOVA is probably the least sensitive of the techniques employed. We also use the general linear model of ANOVA as well as Mann Kendall trend analysis to examine trends in measurement endpoints in each individual lake. The general linear model (GLM) regresses a measurement endpoint against year in each individual lake and determines the significance of the regressions over all the lakes collectively. The analysis can be applied to subsets of the 50 ASL component lakes so that effects within a physiographic region exposed to acidifying emissions can be compared to effects in the baseline lakes remote from acidifying deposition. These comparisons permit us to determine whether an effect is attributable to factors other than acidification. The Mann-Kendall trend analysis also examines trends in measurement endpoints within each individual lake.

132. *Clarify why the Critical Load Calculation is done and review the methodology* (Ap I, p. 16).

The methodology for calculating the critical load of acidity was reviewed explicitly in the 2010 Technical Report (RAMP 2011) and is reported below.

Critical loads are calculated because they have become the standard methodology for determining the risks of a lake for acidification when compared to the potential acid input from industrial emissions. The use of the critical load is an integral part of CEMA's Acid Deposition Management Framework (CEMA 2004, 2010) and the concept is applied in almost all impact assessments. In the RAMP ASL component, the CL is understood largely as a measure of acid-sensitivity.

The critical load was calculated using the standard Henriksen steady state model modified to account for both the buffering of weak organic anions and the lowering of ANC attributable to strong organic acids. The modified model assumed that DOC, with its associated buffering from weak organic acids (ANC_{org}) and reduction of ANC from strong organic acids (A^{-}_{SA}), was exported from the catchment basin to each lake in the same way that we assume the export of base cations (carbonate alkalinity) to each lake. The same model was used recently in the Phase 2, Stage 2 implementation of CEMA's Acid Deposition Management Framework (CEMA 2010). The modified Henriksen model is:

$$CL= ([BC]^*_0 + ANC_{org} - A^-_{SA} - ANC_{lim}) .Q$$

Where,

- BCo is the original base cation concentration before acidification.
- ANC_{lim} is the limiting acid neutralizing capacity of the lake required to maintain a healthy and functional aquatic ecosystem;
- $ANC_{org} = 0.00680^*$ DOC exp(0.8833*pH);
- A_{-SA} = 6.05 *DOC +21.04; and
- Q is the runoff to each lake from the catchment and lake area.

The modifications of the Henriksen model for organic acids and the empirical relationships for ANC_{org} and A_{-SA} are described in WRS (2006) and RAMP (2009b).

133. *Justify the rationale for assuming pre-industrial base cation concentrations are the same as currently observed* (Ap I, p. 17).

This was completed in the 2010 Technical Report (RAMP 2011). See page 3-96 as follows:

During the process of acidification of a catchment, base cations are released from the soils to the lake waters. In applying the Henriksen model in the ASL component, it was assumed that base cations have not increased in these lakes as a result of acidic deposition; that is, the current base cation concentrations are equivalent to the original values. This simplifying assumption was adopted for the following two reasons:

1. The discrepancy between the original and the current base cation concentrations in a lake is normally calculated by an equation presented in Braake *et al.* (1990) based on increases in sulphur concentrations in a lake resulting from aerial deposition. Calculations of BC_0 using the Brakke *et al.* (1990) equation indicated that there is an insignificant difference between the current and calculated original base cation concentrations in all 50 lakes. The results of these calculations are presented in Appendix H of the 2010 Technical Report (RAMP 2011).

2. A study by Whitfield *et al.* (2010a) in which the Magic Model was applied to the oil sands region concluded that, to date, sulphate deposition levels in the oil sands region have resulted in only a limited removal of base cations from the soil.

134. *Justify the rationale for the* ANC_{limit} *and change if necessary* (Ap I, p. 20).

This was completed in the 2010 Technical Report (RAMP 2011, see page 3-97) and described below.

The critical load concept as expressed in the Henriksen model assumes a dose-response relationship between a water quality variable and an aquatic indicator organism. In this case, the water quality variable is the acid neutralizing capacity (alkalinity) required to maintain a healthy fish population. In applying the Henriksen model in Europe, a critical threshold ANC_{lim} of 20 μeq/L was set to protect brown trout, the most common European salmonid and to ensure that no toxic acidic episodes occur to this species during the year.

In North America, the effects of acidification on biota have been historically related to pH rather than alkalinity or acid neutralizing capacity. Research on pH tolerance of a wide range of aquatic organisms has shown that a pH > 6 is required to maintain aquatic ecosystem functioning and protect both fish and other organisms (RMCC 1990, Environment Canada 1997, Jeffries and Lam 1993). Within a given region, lake pH has been empirically and theoretically related to alkalinity as an inverse hyberbolic sine function (Small and Sutton 1986) and this relationship has been used to equate the two variables for the purpose of critical load modelling (e.g., Jeffries and Lam 1993). The relationship between pH and alkalinity for the oil sands region was derived from a water quality survey conducted on lakes in the ALPAC forest management area (WRS 2001, see Appendix H in the 2010 Technical Report [RAMP 2011]). Over these lakes, a pH of 6.0 is associated with an alkalinity of ~75 μeq/L. This value was therefore chosen for ANC_{lim} in the Acid Deposition Management Framework for the Oil Sands Region (CEMA 2004) and has been applied in numerous studies (e.g., Gibson *et al.* 2010).

135. *Recalculate the impact of organic acidity following Lyderson et al. (2004)* (Ap I, p. 21).

The modifications of the Henriksen model for organic acids are described in WRS (2006) for CEMA and RAMP (2009b). They represent empirical equations derived for the RAMP lakes relating the increase in buffering attributable to weak organic acids and the decrease in buffering attributable to strong organic acids to pH and DOC in each lake. These derived relationships will be submitted for publication in a peer-reviewed journal in 2011. The impact of organic acidity following Lyderson *et al.* (2004) will be further investigated during the 2011 monitoring year.

136. *Choose one run-off value for calculating critical loads – I recommend the lake specific values from isotopes* (Ap I, p. 23).

This was completed in the 2010 Technical Report (RAMP 2011, see page 3-96 and section below). We are now only applying the calculation using the isotopic mass balance approach.

The runoff (Q) to each lake, was calculated from analysis of heavy isotopes of oxygen (^{18}O) and (^{2}H) in each lake conducted and provided by John Gibson (University of Victoria). With this technique, the natural evaporative enrichment of ^{18}O and ^{2}H in each lake is used to partition water losses between evaporation and liquid outflow and hence derive an estimate of runoff (Gibson *et al*. 2002, Gibson and Edwards 2002, and Gibson *et al*. 2010). This technique utilizes a different set of assumptions from traditional hydrometric methods, which extrapolate water yields from one or more gauged catchments to the ungauged lake catchments. Potential inaccuracies in the traditional hydrometric method, especially in low-relief catchments, have previously been recognized in lakes in the Athabasca oil sands region (WRS 2004).

137. *Evaluate whether the SSWC approach is appropriate for these lakes* (Ap I, p. 23).

The SSWC (Henriksen model) approach is probably not the most appropriate critical load model. The SSWS is a runoff-based model that assumes that the input of alkalinity to a lake is carried in surface runoff to the lake. In fact, many of the lakes in the oil sands have a high connectivity to the surficial groundwater aquifer. We have assumed this model because (1) it is relatively simple to calculate (2) requires very few data; and (3) it has become the standard in CEMA's and AENV's acid deposition management framework. It is a simple survey tool. Adoption of a more complicated dynamic model incorporating terms of groundwater seepage etc. would require a lot more information than we have on most of the catchments. Given the fact that most of these lakes are not that highly acid sensitive, the use of more complicated model may not be warranted. Whitefield *et al.* (2010a, 2010b) applied the MAGIC model in a NO_xSO_x project and concluded that acidification is unlikely in these lakes. We also state that the critical load is understood as a measure of acid-sensitivity rather than a true "critical load".

138. *PAI should not be used for estimating critical load exceedances* (Ap I, p. 23).

This has been reviewed and discussed in the 2010 Technical Report (RAMP 2011, see page 3-97) and summarized below.

We disagree with the statement that PAI should not be used in estimating critical load exceedances.

The three major reasons presented by the reviewer for not using the PAI are that the base cations are counted twice (once in the calculation of PAI and once in the Henriksen model itself), the N in the estimate of PAI is partially eutrophying rather than acidifying and that S is non-conservative and immobilized in the catchments. Using PAI would be a worst-case scenario by assuming that all the N and S are acidifying.

In the first case, we do not believe that we are counting the base cations twice. In the calculations of PAI we are using the base cation concentration in the estimates of dry fallout to neutralize a portion of the acidity in the NO_xSO_x deposition. The value of PAI expresses the net number of hydrogen ions deposited per hectare per year after this neutralization. The base cation concentrations in the Henriksen model correspond to the base cations that are released from the soil matrix in the lake catchments after displacement by these hydrogen ions (alkalinity export to the lake). In other words, we are actually talking about two separate sources of base cations. They are conceptually non-equivalent.

For the second point, we agree that some of the N inputs to the lake catchments will be eutrophying rather than acidifying. Our calculations of PAI, however, do not assume the all the N is acidifying. The ability of nitrates to be assimilated and used as a nutrient by plants within the lake catchment was accounted for by applying the approach adopted by CEMA and AENV whereby only nitrogen deposition in excess of 10 kg/ha/y and 25 % of the first 10 kg/ha/y deposited N were considered acidifying (CEMA 2008, AENV 2007).

For the third point, we agree that some of the S in the PAI is probably not acidifying and is immobilized in the catchment (Whitfield *et al.* 2010a). Proof of this assertion is the fact that we do not see large increases in sulphate in these lakes despite decades of elevated S deposition. This would indeed make comparison of the PAI to the critical load highly conservative and a “worst-case scenario”.

The Henriksen model is simplistic and the case could easily be made that it is not the best model to use for all lakes within the oil sands region. Comparison of the critical loads to the PAI is also problematic especially if the S and N deposited in the lake catchments are, to an unknown extent, non-acidifying and do not end up in the lake. However, the model requires very little data and can easily be applied as a survey tool. The potential errors listed above result in a “worst-case scenario” where the potential for acidification is overestimated. The approach therefore is highly conservative and protective of lake integrity. We do not feel that erring in this direction to be altogether a negative feature. Other approaches (models) would require far more data and resources, which may not be warranted. Comparison of the critical loads calculated from the Henriksen model to the PAI has become an integral part of CEMA’s acid deposition management framework and has been applied as recently as 2010 by CEMA in their Phase 2 Implementation of the CEMA Acid Deposition Management Framework (CEMA 2010).

139. *Remove Shewart charts from the analysis/reports* (Ap I, p. 25).

The use of Shewhart charts to track measurement endpoints is a well known ecological technique for detecting trends (See Gilbert 1987). In reality, we are simply plotting the variable against time and looking for trends graphically. Most of the RAMP components, including water quality and benthos, do this as well and compare their value to the 95th percentile to determine whether change has occurred. We are comparing our measurement endpoints to two and three standard deviations from the mean, a similar process. We assume that the reviewer’s objections concern the method of interpreting whether change has occurred. In particular, a three standard deviation change in a measurement endpoint would be catastrophic, a far too insensitive criterion. To address these concerns we have tightened our interpretation of these charts to include a variety of rules outlined in the 2010 Technical Report (RAMP 2011) to detect trends. In the 2011 report we will also tighten the control limits by excluding the current year’s data.

140. *Check database and reports for errors in units; it would be beneficial to check the text for consistencies or errors* (Ap I, p. 25-26).

The data were thoroughly checked prior to the release of the public database and QA/QC of the data is done annually prior to including new data in the database.

3.0 REFERENCES

AAFC (Agriculture and Agri-foods Canada). 2007. Soil Landscapes of Canada (v3.1.1). Geospatial data accessed at: http://www4.agr.gc.ca/AAFC-AAC/display-afficher.do?id=1227014964079&lang=eng

AENV (Alberta Environment). 2000. Spring Creek watershed post harvesting – phase II impacts of timber harvesting. Water Sciences Branch, Hydrology/Forecasting Section Report -7GF, 2000-314. Prepared for: Water Sciences Branch. Prepared by: John Taggart. Edmonton. December 2000. 64 pp.

AENV. 2007. Review of the Acid Deposition Management Framework and its Implementation. Alberta Environment Report. Environmental Policy Branch. pp 19.

AITF (Alberta Innovates Technology Futures). 2011. 2010 Regional Aquatics Monitoring Program (RAMP) Scientific Review. Prepared by: RAMP Review Panel. Prepared for: RAMP Steering Committee. 160 pp.

Astel, A., S. Tsakovski, P. Barbieri, and V. Simeonov. 2007. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. Water Research 41(19): 4566-4578

Ayles, B., M. Dube and R. Rosenburg. 2004. Oil Sands Regional Aquatics Monitoring Program (RAMP) Scientific Peer Review of the Five Years Report (1997-2001). Prepared for and submitted to the RAMP Steering Committee.

Bailey, R.C., R.H. Norris and T.B. Reynoldson. 2004. Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach. Kluwer Academic Publishers.

Borisko, J.P., B.W. Kilgour, L.W. Stanfield, and F.C. Jones. 2007. An evaluation of rapid bioassessment protocols for stream benthic invertebrates in southern Ontario, Canada. Water Quality Research Journal of Canada, 42: 184-193.

Boyacioglu, H. and H. Boyacioglu. 2010. Detection of seasonal variations in surface water quality using discriminant analysis. Environmental Monitoring and Assessment 162 (1): 15:20.

Brakke, D.F., A. Henriksen and S.A. Norton. 1990. A variable F-factor to explain changes in base cation concentrations as a function of strong acid deposition. Verh. Internat. Verein. Limnol. 24: 146-149.

Burke, L. 2008. Genetic Characterization of walleye populations in the Athabasca River basin. Prepared for: Cumulative Environmental Management Association (CEMA). Prepared by L. Burke, University of Alberta. October 2008. 51 p.

CCME (Canadian Council of Ministers of the Environment). 2003. Guidance on the Site-Specific Application of Water Quality Guidelines in Canada: Procedures for Deriving Numerical Water Quality Objectives. Canadian Council of Ministers of the Environment.

CCME. 2011. Canadian Environmental Quality Guidelines and Summary Table. Canadian Council of Ministers of the Environment. Internet version.

CEMA (Cumulative Environmental Management Association). 2004. Recommendations for the Acid Deposition Management Framework for the Oil Sands Region of North-Eastern Alberta. Report to NOXSOX Management Working Group of the Cumulative Environmental Management Association (CEMA.) pp 39.

CEMA. 2008. Proposed Interim Nitrogen (Eutrophication) Management Recommendations and Work Plan for the Regional Municipality of Wood Buffalo. (updated May 12, 2008). Prepared for the NOXSOX Management Working Group of the Cumulative Environmental Management Association (CEMA). 57 pp.

CEMA. 2010. The Assessment of Acid Deposition in the Alberta Oil Sands Region – Phase 2 of Stage 2 Implementation of the CEMA Acid Deposition Management Framework. Report by Golder Associates Ltd. to the Cumulative Environmental Management Association (CEMA). CEMA Report No. 08-1331-0023.

Cohen, J. 1977. Statistical Power Analysis for the Behavioral Sciences. Academic Press, New York. 474 pp.

Corkum, L. 1985. Water Quality of the Athabasca Oil Sands Area: A Regional Study. Alberta Oil Sands Environmental Research Program (AOSERP) Report L-85. Alberta Environment, Edmonton.

David, S.M., K.M. Somers, R.A. Reid, R.J. Hall, and R.E. Girard. 1998. Sampling protocols for the rapid bioassessment of streams and lakes using benthic macroinvertebrates. 2nd edition. Dorset Environmental Sciences Centre, Aquatic Science Section, Ontario Ministry of the Environment, Dorset, Ontario.

DFO and EC (Department of Fisheries and Oceans and Environment Canada). 1995. Further guidance for the invertebrate community survey for aquatic environmental effects monitoring related to federal Fisheries Act requirements. EEM 2, February 1995.

Dußling U., R. Berg, H. Klinger, and C. Wolter. 2004. Assessing the ecological status of river systems using fish assemblages. In Handbuch angewandte Limnologie, VIII-7.4, 20. Erg.Lfg. 12/04. C. Steinberg, W. Calmano, H. Klapper and R-D. Wilken (eds.). Ecomed: Landsberg am Lech, Germany; 1-84.

Elliott, J.M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Freshwater Biological Association, Scientific Publication No. 25.

Environment Canada. 1997. 1997 Canadian Acid Rain Assessment. Volume Three. The Effects on Canada's Lakes, Rivers and Wetlands.

Environment Canada. 2005. Revised pulp and paper EEM guidance document. Environment Canada. July 2005.

Environment Canada. 2010. 2010 Pulp and Paper Environmental Effects Monitoring (EEM) Technical Guidance Document. Environment Canada. November.

Gibson J.J., E.E. Prepas, and P. McEachern. 2002. Quantitative comparison of lake throughflow, residency and catchment runoff using stable isotopes: modelling and results from a regional survey of Boreal lakes. J. Hydrol. 262: 128-144.

Gibson, J.J. and T.W.D Edwards. 2002. Regional water balance trends and evaporation-transpiration partitioning from a stable isotope survey of lakes in northern Canada. Global Biogeochemical Cycles. 16: 10-1 to 10-14.

Gibson, J.J., S.J. Birks, S. Kumar, and P.M. McEachern. 2010. Inter-annual variation in water yield to lakes in northeastern Alberta: Implications for estimating critical loads of acidity. J. Limnol., 69 (Suppl. 1): 126-134.

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Company, New York, NY. 320 pp.

Golder (Golder Associates). 1999. Oil Sands Regional Aquatics Monitoring Program (RAMP) 1998. Submitted to RAMP Steering Committee. May 1999.

Government of Canada. 2008 Technical guidance document for Water Quality Index practitioners reporting under the Canadian Environmental Sustainability Indicators (CESI) initiative 2008. 48 pp.

Glozier, N.E., D.B. Donald, R.W. Crosley, and D. Halliwell. 2009. Wood Buffalo National Park Water Quality: Status and Trends from 1989-2006 in Three Major Rivers; Athabasca, Peace and Slave. Environment Canada, Saskatoon, SK. May 2009.

Grewer, D.M., R.F. Young, R.M. Whittal, and P.M. Fedorak. 2010. Naphthenic acids and other acid-extractables in water samples from Alberta: what is being measured? Sci Total Environ. 408(23): 5997-6010.

GSC (Geological Survey of Canada). 2006. Geospatial data describing surficial geology of northern Alberta. Accessed at: http://geopub.nrcan.gc.ca

Güler, C., G.D. Thyne, J.E. McCray, and A.K. Turner. 2004. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeology Journal, 10: 455-474.

Headley, J.V., K.M. Peru, S.A. Armstrong, X. Han, J.W. Martin, M.M. Mapolelo, D.F. Smith, R.P. Rogers, and A.G. Marshall. 2009. Aquatic plant-derived changes in oil sands naphthenic acid signatures determined by low-, high- and ultrahigh-resolution mass spectrometry. Rapid Commun. Mass Specrom. 23: 515-522.

Hebben, T. 2009. Analysis of Water Quality Conditions and Trends for the Long-Term River Network: Athabasca River, 1960-2007. Alberta Environment, Edmonton, AB. March 2009.

Heyes, A., T.R. Moore, J.W.M. Rudd, and J.J. Dugoua. 2000. Methyl mercury in pristine and impounded boreal peatlands, Experimental Lakes Area, Ontario. Canadian Journal of Fisheries and Aquatic Sciences. 57: 2211-2222.

Hoke, R.A., J.P. Geisy and J.R. Adams. 1990. Use of linear orthogonal contrasts in environmental data. Environmental Toxicology and Chemistry, 9:815-819.

Hughes, B. and T. Whittier. 2008. An assessment of the Regional Aquatics Monitoring Program (RAMP) Fish Survey. Prepared for Hatfield Consultants Partnership. Department of Fisheries and Wildlife. Oregon State University, Corvallis, Oregon.

Hughes, R.M. and D.V. Peck. 2008. Acquiring data for large aquatic resource surveys: the art of compromise among science, logistics, and reality. Journal of the North American Benthological Society 27:837-859.

Humphries, D. 2011. Senior Analyst, Organic Environmental Monitoring, Alberta Innovates – Technology Futures (AITF). Edmonton, AB, personal communication. Teleconference Meeting, April 2011, Vancouver, BC.

Imhof, J.G., J. Fitzgibbon, and W.K. Annable. 1996. A hierarchical evaluation system for characterizing watershed ecosystems for fish habitat. Canadian Journal of Fisheries and Aquatic Sciences, 53(Supplement 1):312-326.

Jeffries, D. S. and D.C.L. Lam. 1993. Assessment of the effect of acidic deposition on Canadian lakes: Determination of critical loads for sulphate deposition. Water Science and Technology 28 (3-5): 183-187.

Jones, R.D. and J.N. Boyer. 2002. FY2001 Annual report of the water quality monitoring project for the water quality protection program in the Florida Keys National Marine Sanctuary. Southeast Environmental Research Center Technical Report #T181. Florida International University, Miami, FL. 48 pp.

Kanno Y., J.C. Vokoun, D.C. Dauwalter, R.M. Hughes, A.T. Herlihy, T.R. Maret, and T.M. Patton. 2009. Influence of rare species on electrofishing distance–species richness relationships at stream sites. Transactions of the American Fisheries Society 138:1240-1251.

Karr, J.R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6(6):21-27.

Kelly, E.N., J.W. Short, D.W. Schindler, P.V. Hodson, M. Ma, and A.K. Kwan. 2009. Oil-sands development contributes polycyclic aromatic compounds to the Athabasca River and its tributaries. Proceedings of the National Academy of Sciences of the United States of America 106(52): 22346-22351.

Kilgour, B.W., K.M. Somers and D.E. Matthews. 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. Écoscience, 5:542-550.

Kilgour, B.W., K.R. Munkittrick, C.B. Portt, K. Hedley, J. Culp, S. Dixit, and G. Pastershank. 2005. Biological criteria for municipal wastewater effluent monitoring programs. Water Quality Research Journal of Canada, 40:374-387.

Lowell, R.B., S.C. Ribey, I.K. Ellis, E.L. Porter, J.M. Culp, L.C. Grapentine, M.E. McMaster, K.R. Munkittrick, and R.P. Scroggins. 2003. National assessment of the pulp and paper environmental effects monitoring data. National Water Research Institute, NWRI Contribution No. 03-521.

Lyderson, E., T. Larssen and E. Fjeld. 2004. The influence of total organic carbon (TOC) on the relationship between acid neutralizing capacity (ANC) and fish status in Norwegian lakes. Science of the Total Environment 326: 63-69.

Morrison, L.W. 2008. Use of Control Charts to Interpret Environmental Monitoring Data. Natural Areas Journal 28(1): 66-73.

Nero, V., A. Farwell, L.E. Lee, T. Van Meer, M.D. McKinnon, and D.G. Dixon. 2006. The effects of salinity on naphthenic acid toxicity to yellow perch: gill and liver histopathology. Ecotoxicol. Environ. Saf. 65(2): 252-64.

NRC (Natural Resources Canada). 2010. CanmetENERGY. Oil sands water toxicity: A critical review. By Brenda Miskimmin, Philip Fedorak, Robert Lauman, and Kristin Vinke. Devon, Alberta, Report No. 2010-089 (INT).

Pappas, S. 2009. Environmental Monitoring Scientist, Environment Canada, Vancouver, personal communication. Meeting December 23, 2009, Vancouver, BC.

Parsons, B.G., S.A. Watmough, P.J. Dillon, and K.M. Somers. 2010. Relationships between lake water chemistry and benthic macroinvertebrates in the Athabasca oil sands region, Alberta. Journal of Limnology, 69:118-125.

Peck, D.V., A.T. Herlihy, B.H. Hill, R.M. Hughes, P.R. Kaufmann, D.J. Klemm, J.M. Lazorchak, F.H. McCormick, S.A. Peterson, P.L. Ringold, T. Magee, and M.R. Cappaert. 2006. Environmental Monitoring and Assessment Program - Surface Waters Western Pilot Study: field operations manual for wadeable streams. EPA/620/R-06/003. U.S. Environmental Protection Agency, Washington, DC.

RAMP (Regional Aquatics Monitoring Program). 2009a. RAMP 2008 Technical Report. Prepared for the RAMP Steering Committee by Hatfield Consultants, Kilgour and Associates Ltd., Klohn Crippen Berger Ltd., and Western Resource Solutions. April 2009.

RAMP. 2009b. RAMP Technical design and rationale document. Prepared for the RAMP Steering Committee. Hatfield Consultants, Kilgour & Associates Ltd., Klohn Crippen Berger Ltd., and Western Resource Solutions.

RAMP. 2010. RAMP 2009 Technical Report. Prepared for the RAMP Steering Committee by Hatfield Consultants, Kilgour and Associates Ltd., Klohn Crippen Berger Ltd., and Western Resource Solutions. April 2010.

RAMP. 2011. RAMP 2010 Technical Report. Prepared for the RAMP Steering Committee. Hatfield Consultants, Kilgour and Associates Ltd., and Western Resource Solutions.

Reynoldson, T.B., C. Logan, T. Pascoe, and S.P. Thompson. 2004. CABIN (Canadian Aquatic Biomonitoring Network) invertebrate biomonitoring field and laboratory manual. Environment Canada, National Water Research Institute.

RMCC (Research and Monitoring Coordinating Committee). 1990. The 1990 Canadian Long-Range Transport of Air Pollutants and Acid Deposition Report. Part 4: Aquatic Effects. Federal – Provincial Research and Monitoring Committee. Ottawa, Ontario. 151 pp.

Singh, K.P., A. Malik, D. Mohan, and S. Sinha. 2004. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)--a case study. Water Research 38(18): 3980-3992.

Small, M.J. and M.C. Sutton. 1986. A regional pH-alkalinity relationship. Water Research 20: 335-343.

Stanfield, L.W. and B.W. Kilgour. 2006. Effects of Percent Impervious Cover on Fish and Benthos Assemblages and In-Stream Habitats in Lake Ontario Tributaries. In, R. M. Hughes, L. Wang, and P. W. Seelbach (eds). Influences of landscape on stream habitats and biological assemblages. American Fisheries Society, Symposium 48: 577-599, Bethesda, Maryland.

Whitefield, C.J., J.E. Ahern, J.B. Cosby, and S.A. Watmough. 2010a. Modelling boreal lake catchment response to anthropogenic acid deposition. Journal of Limnology 69(Suppl. 1): 135-146, 2010 - DOI: 10.3274/JL10-69-S1-14.

Whitefield, C.J., J.E. Ahern, J.B. Cosby, and S.A. Watmough. 2010b. Modelling catchment response to acid deposition: a regional dual application of the MAGIC model to soils and lakes in the Athabasca Oil Sands Region, Alberta. Journal of Limnology, 69(Suppl. 1): 147-160, 2010 - DOI: 10.3274/JL10-69-S1-15.

WSC (Water Survey of Canada). 2001. Hydrometric Technician Career Development Program. Accessed at: http://www.smc-msc.ec.gc.ca/wsc/CDP/

WRS (Western Resource Solutions). 2000. Critical loads of Acidity to 162 lakes Samples by Alberta-pacific Forest Industries during 1998. Report by Western Resource Solutions to Syncrude Canada Ltd.

WRS. 2001. Critical loads Of acidity to 162 lakes sampled by Alberta-Pacific forest industries during 1998: Prepared for Syncrude Canada Limited, Fort McMurray, Alberta by Western Resource Solutions, Calgary, Alberta.

WRS. 2003. Analysis of the Water Quality of the Steepbank, Firebag and Muskeg Rivers during the Spring Melt (1989-2001). Final Report prepared by Western Resource Solution for the Wood Buffalo Environmental Association, Fort McMurray, Alberta. 73 pp.

WRS. 2004. Calculations of critical loads of acidity to lakes in the oil sands region. Report by Western Resource Solutions to the NO_x-SO_x Management Working Group.

WRS. 2006. Critical loads of acidity to lakes in the Athabasca oil sands region – modification of the Henriksen Model for lake organic content. Final report to: NO_xSO_x Management Working Group, Fort McMurray, Alberta. January, 2006.

Zubot, W. 2010. Syncrude Canada Ltd. Edmonton, AB, personal communication. Meeting, April 2010, Vancouver, BC.

APPENDICES

Appendix A1

Letter to RAMP from the Assistant Deputy Minister, Monitoring and Science Division, Alberta Environment

Government of Alberta
Environment

Monitoring and Science Division
10 Floor, South Petroleum Plaza
9915 – 108 Street
Edmonton, Alberta T5K 2G8
Telephone: 780-427-0029
Fax: 780-427-0923
www.alberta.ca

AR 43329

March 7, 2011

Mr. Tomas Romero
Steering Committee Chair,
Regional Aquatics Monitoring Program
ConocoPhillips Canada
P. O. Box 130
Postal Station M
Calgary, Alberta T2P 2H7

Dear Mr. Romero:

I am writing to provide advice on the recommendations emerging from the 2010 Regional Aquatics Monitoring Program scientific review. In their report, the peer reviewers' provide a number of technical and non-technical recommendations for improvement of the Regional Aquatics Monitoring Program, some of which represent significant program changes if implemented.

The peer reviewers' technical recommendations for Regional Aquatics Monitoring Program focused on expanding the temporal and spatial scope of the program, better integrating all environmental monitoring programs within the region, embedding the current monitoring within a decision-making framework, addressing the issue of inflating baseline variability, and better integrating the analysis of physical and biological information. The reviewers also provided recommendations on program management that deal with science oversight and governance matters.

Some of these recommendations touch on areas that are currently being examined through ongoing federal or provincial monitoring initiatives. For example, several of peer reviewers' recommendations fall within the mandate of the recently appointed provincial environmental monitoring panel to resolve (i.e., integration of existing monitoring programs, governance and science oversight). Others such as the spatial and temporal expansion of the Regional Aquatics Monitoring Program, touch on, and have the potential to conflict with, Environment Canada's surface water quality monitoring plan, or recommendations on the physical monitoring network emerging from the provincial environmental monitoring panel.

I understand that the Regional Aquatics Monitoring Program has already taken steps to implement some of the reviewers' recommendations around statistical analyses and data presentation. The timely implementation of these recommendations is commendable. However, due to the overlap and potential conflict with the federal and provincial monitoring initiatives underway, it is advisable that major changes to the Regional Aquatics Monitoring Program are deferred until after these initiatives are substantially complete.

...2

Much more will be known about the future environmental monitoring, evaluating and reporting system for the Athabasca oil sands by the summer of 2011, once government has had an opportunity to consider recommendations from the provincial environmental monitoring panel. I believe that at that point, the Regional Aquatics Monitoring Program may have enough information to better understand their role within the environmental management system and will be in a better position to examine the peer reviewers' recommendations that require significant program changes.

In closing, I would like to acknowledge the significant contribution the Regional Aquatics Monitoring Program has made over the years to the information base on aquatic ecosystems within the lower Athabasca River basin, and thank program members for their diligence and ongoing commitment to improving the program.

Sincerely,

Bob Barraclough
Assistant Deputy Minister
Monitoring and Science Division

cc: Ernie Hui, Alberta Environment
Rick Brown, Alberta Environment

Appendix A2

Supporting Information from RAMP (2011)

A2 SUPPORTING INFORMATION FROM RAMP (2011)

Sections A2.1 to A2.5 contain supporting information from Chapter 6 of the RAMP 2010 Technical Report for several recommendations (RAMP 2011). Section A2.6 and A2.7 contains supporting information from Chapter 5 of the RAMP 2010 Technical Report (Section 5.1, RAMP 2011) for recommendations #60 and #110, respectively.

A2.1 NAPHTHENIC ACIDS IN WATER

A2.1.1 Background

Formally, naphthenic acids are a broad group of alkyl-substituted carboxylic acids, with the general formula $C_nH_{2n+Z}O_2$, where n is the number of carbon atoms (typically between 10 and 20), and Z is a negative number corresponding to twice the number of rings in the molecule (i.e., 0, -2, -4, etc.). This group includes numerous compounds with various cyclic and acyclic (aliphatic) structures.

Grewer *et al.* (2010) provides a history of the analysis and interpretation of naphthenic acids in oil sands process waters (OSPW) and ambient surface water samples. Information from this study and other sources has been briefly summarized below.

Naphthenic acids became associated with the environmental chemistry of the oil sands region when MacKinnon and Boerger (1986, cited in Grewer *et al.* 2010) indicated that observed toxicity of oil sands tailings pond waters was likely associated with “polar organic carboxylic acids (naphthenic acids)”. This assertion was partly based on their observation that the acid-extracted organic compounds associated with toxicity was very similar in composition to commercial preparations of naphthenic acids, using a Fourier transform infra-red (FTIR) spectrum analysis (Grewer *et al.* 2010).

FTIR-measured concentrations of “naphthenic acids” in oil sands process waters (OSPW) are in the tens to low-hundreds of mg/L (Han *et al.* 2009, Grewer *et al.* 2010), which are concentrations that have been shown to cause toxicity to aquatic organisms (Nero *et al.* 2006). Given concerns about potential accidental release of naphthenic acids to local receiving waters through seepage from tailings facilities, this method also was applied to ambient surface waters samples in various site-specific and regional environmental monitoring programs, including those conducted by RAMP and AENV. From 1997 to 2008, RAMP samples were analyzed by ALS Environmental using this method, with a method detection limit of 1 mg/L.

Different high-resolution techniques were developed and applied to the measurement of “naphthenic acids” in the oil sands region in the mid-2000s, largely in response to concerns regarding potential effects of OSPW toxicity on effective tailings pond reclamation strategies. It became clear that the FTIR method (as well as the newer, high-resolution methods) measured many more acid-extractable organic compounds than those classically defined as “naphthenic acids” by the formula listed above. This included longer-chain acids, more highly oxidized species (i.e., O_3 to O_7, not just O_2), and those with more complex oxy-groups, such as SO_2 to SO_6, and NO_4 (Headley *et al.* 2009, Grewer *et al.* 2010). Assessments of samples of OSPW, commercial naphthenic acids preparations, and ambient river water samples using both low-resolution FTIR and an ultrahigh-resolution method (electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry, or ESI-FT-ICR MS) by Grewer *et al.* (2010) found that most acid-extractable acids present in these mixtures, including in a commercial naphthenic acids mixture, did not fit the accepted definition of naphthenic acids or their oxidized derivatives. They also found that the FTIR method gave generally comparable results to the high-resolution method at high (OSPW-type) concentrations but overestimated naphthenic acids concentrations in ambient river water samples. Fewer than 10% of acid-extractable organics measured by Grewer *et al.* in river water samples from various locations in Alberta were classic naphthenic acids, with ≥70% of these compounds being aliphatic (non-cyclic) fatty acids, particularly palmitic and stearic acids, which are common components of biological cell membranes and routinely found in river waters. Given the complexity of acid-extractable organics found in OSPW and surface-water samples, Grewer *et al.* (2010) suggested the replacement of the term "naphthenic acids" for these analyses with something better representative of the range of compounds measured, such as "oil sands tailings water acid-extractable organics (OSTWAEO)". Given many of these constituent compounds also are present in surface waters outside the oil sands region, the more general term of "acid-extractable organics" is used in this section.

These recent studies have demonstrated the need to improve analytical techniques used to identify acid-extractable organics in OSPW, define those with greatest potential for environmental change, and apply this knowledge to future environmental monitoring programs. Not only do new, high-resolution methods (combined with meaningful toxicological data) potentially allow for more accurate and precise identification of concentrations of concern for this suite of compounds as a whole, precise speciation of many individual acid-extractable organics in a single sample may allow for identification of unique "fingerprints" of different OSPWs. Such "fingerprints" could then be compared with those in ambient surface water samples to potentially identify specific sources of any OSPW-associated organics observed in an ambient sample.

At least four different laboratories are currently developing or using high-resolution analytical techniques for quantification and speciation of naphthenic acids mixtures in water, including:

- AITF (formerly ARC, Vegreville, AB), which uses a GC/MS-ion-trapping method, and was the laboratory used by AENV and RAMP in 2009 and 2010 for analysis ambient water quality samples;
- ALS Environmental Ltd. (Edmonton, AB), who have developed a high- resolution gas chromatography/mass spectrometry (GC/MS, operating at 10,000 resolution), selected-ion method, targeting the following selected ions: m/z 286.2278 (9-FCA), 267.1780 (napthenic acids) and 267.0836 (^{13}C-tetradecanoic);
- Dr. Jon Martin's laboratory at the University of Alberta (Edmonton, AB), which uses an ultra-high-resolution quadrupole, time-of-flight mass spectrometry (Q-TOF MS) and Fourier transform ion cyclotron resonance mass spectrometry (FT-ICRMS); and
- AXYS Analytical Services Ltd. (Sidney, BC), which uses a high-resolution liquid chromatography/MS/MS method (currently being used to analyze samples collected with passive samplers as part of AENV's ongoing Contaminant Load Study in the Athabasca River).

In 2009, AENV began using AITF for analysis of "naphthenic acids" in surface waters collected for routine monitoring at AENV's Long-Term Regional Network (LTRN) locations. In 2009, RAMP also shifted its naphthenic acids analysis from ALS (using low-resolution FTIR) to AITF, to match the analytical method being used by AENV. AITF's method in 2009 was based on a GC/MS-ion-trapping method, and provided a method detection limit of 20 µg/L. Results in fall 2009 using this higher-resolution technique indicated concentrations of naphthenic acids (acid-extractable organics) of 0.035 to

0.848 mg/L, consistent with previous RAMP data (based on FTIR analysis), which typically returned values of <1 mg/L (RAMP 2009a).

A2.1.2 Analyses of 2010 RAMP Water Samples for Naphthenic Acids

A2.1.2.1 Methods

Recognizing current uncertainties and ongoing method development in the identification and quantification of acid-extractable organic acids, in 2010 RAMP collected triplicate samples in spring, summer and fall for analysis of these compounds. One set was provided to AITF as previously proposed in the RAMP 2010 sampling design; a second set of samples was provided to Dr. Deib Birkholz at ALS Environmental (Edmonton) for analysis using their HRGC/MS-selected-ion method; and a third set of samples was provided to Dr. Jonathan Martin at University of Alberta. Recognizing the value of these ambient water samples for method development and validation, AITF provided speciation data at no additional cost to RAMP, ALS provided analysis of a subset of samples provided at a significant discount, and Dr. Martin's laboratory also used these samples in their research.

As of the time of reporting, complete analyses of RAMP 2010 samples had been undertaken and data shared with RAMP by AITF and ALS (data from four stations [MIC-1, CAR-1, ATR-MR-E, ATR-FR-CC] could not provided from ALS due to matrix interferences that confounded quantification). Samples provided to University of Alberta had not yet been fully analyzed and reported.

In spring 2010, AITF modified their analytical method to reduce the mass-unit range of compounds measured in an attempt to eliminate some of the compounds not classically defined as naphthenic acids from their results. The AITF 2010 data provide results that may be compared with those from 2009, despite inconsistencies between these methods (D. Humphries, AITF, *pers. comm.*, April 2011).

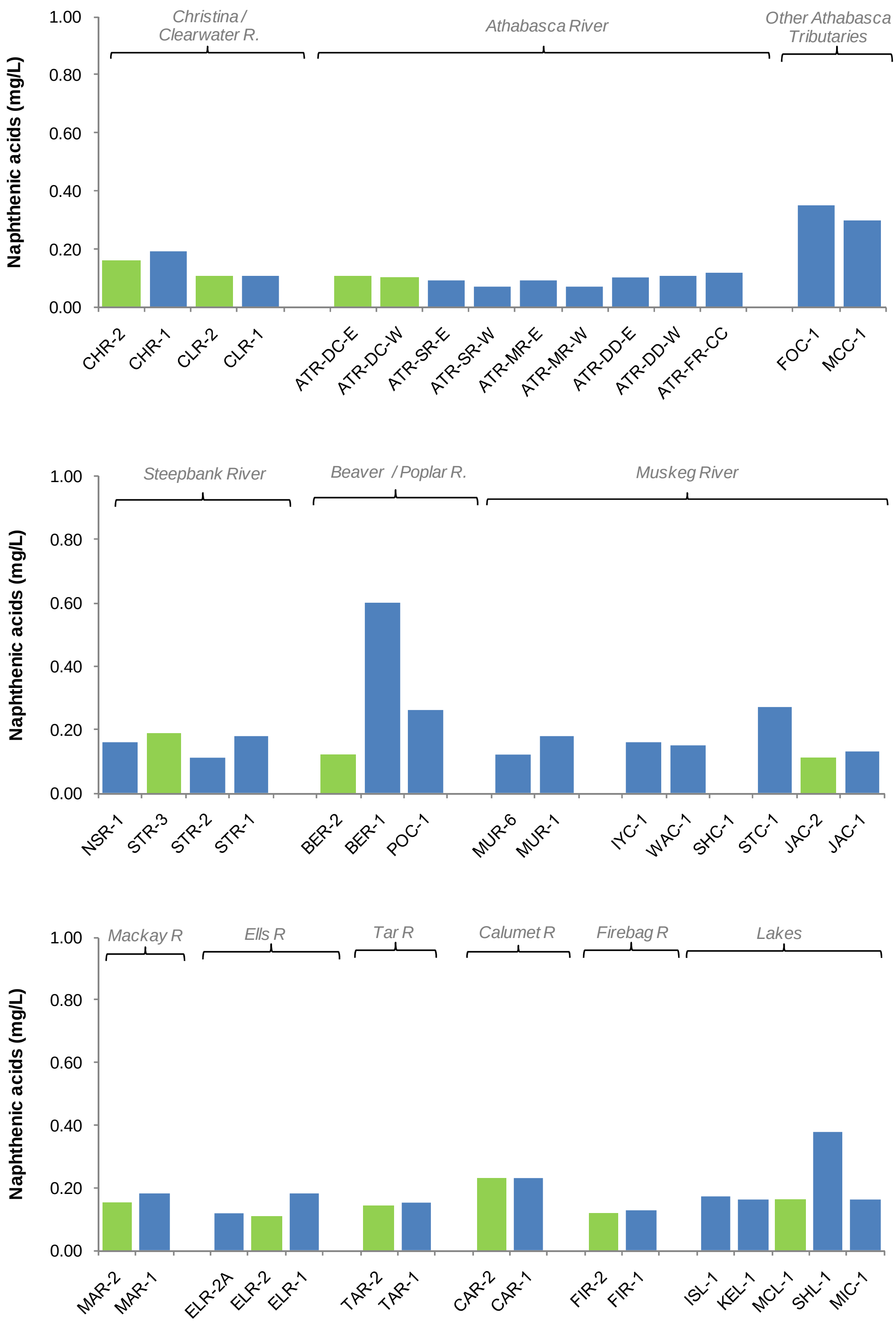
A2.1.3 Results and Discussion

Comparison with 2009

Figure A2.1-1 presents results of naphthenic acids analyses performed by AITF from RAMP water quality stations in fall 2010. Observed concentrations and spatial patterns among stations were generally similar between 2009 and 2010 (Figure A2.1-2).

Concentrations at most stations were below 0.2 mg/L. The highest concentration (i.e., 0.6 mg/L) was observed in lower Beaver River (*test* station BER-1), downstream of the Mildred Lake Settling Basin, followed by Shipyard Lake (*test* station SHL-1), Fort Creek (*test* station FOC-1) and McLean Creek (*test* station MCL-1), which are all small watersheds downstream of oil sands developments. The next highest concentrations were in the Calumet River, with similar concentrations in *baseline* station CAR-2 and *test* station CAR-1 (Figure A2.1-1). Concentrations in the Athabasca River mainstem showed gradual increases moving downstream in fall 2010, particularly downstream of the Muskeg River (Figure A2.1-1). In fall 2009, concentrations were gradually decreasing moving downstream along the entire river (Figure A2.1-2).

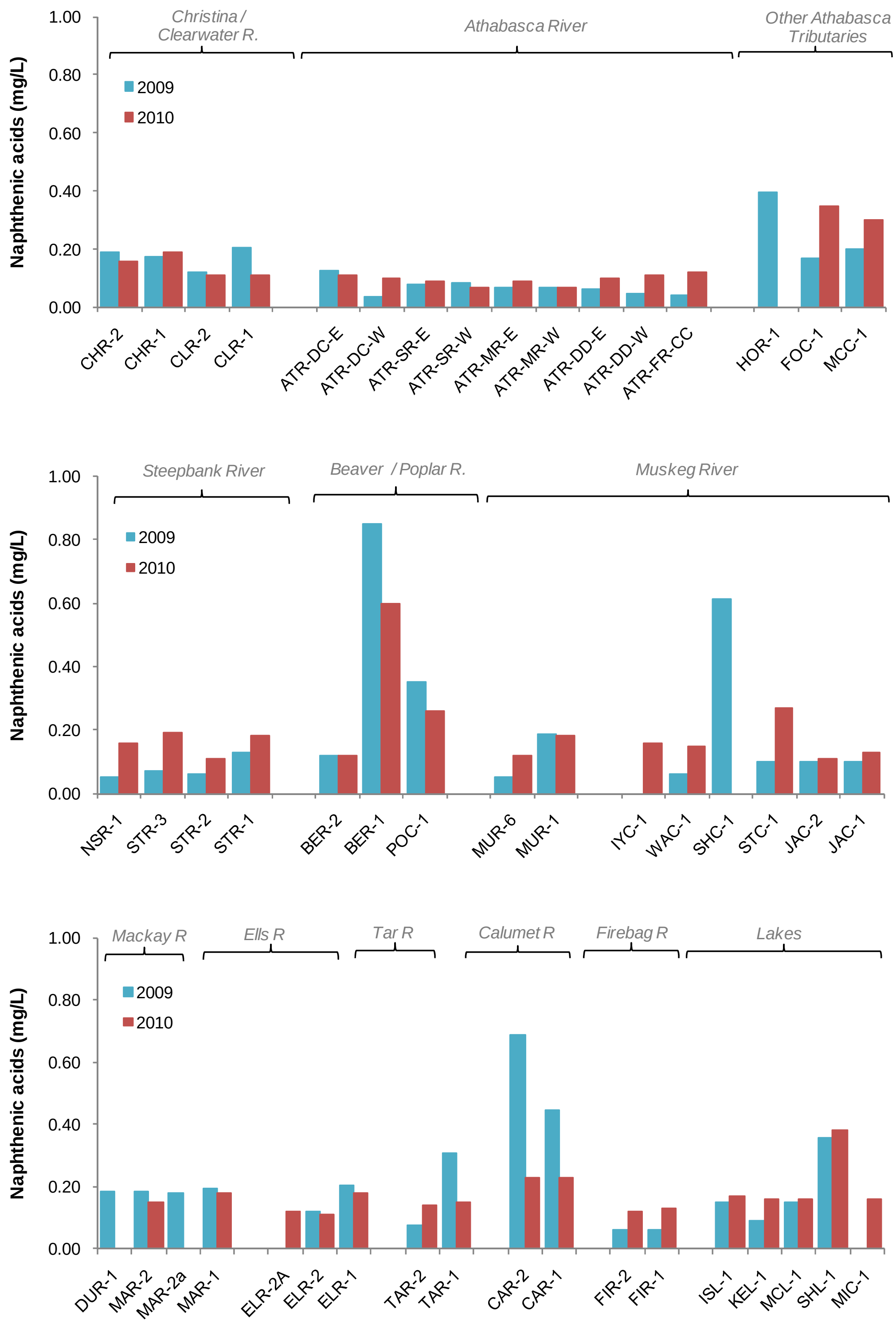
Figure A2.1-1 Concentrations of acid-extractable organic acids (naphthenic acids) in the RAMP FSA, fall 2010.



Note: results were adjusted to allow for comparisons with 2009 results.

Note: green denotes *baseline* stations and blue denotes *test* stations.

Figure A2.1-2 Concentrations of acid-extractable organic acids (naphthenic acids) in the RAMP FSA, fall 2009 and 2010 results.



Comparison of Methods

A comparison of concentrations of naphthenic acids reported by AITF and ALS for fall 2010 is presented in Figure A2.1-3. The method developed and applied by ALS is intended to be specific to specific ions classically defined as naphthenic acids (i.e., $C_nH_{2n+Z}O_2$). Analyses of all fall 2010 RAMP samples using this method returned all non-detectable values (at detection limits ranging from 2 to 5 µg/L), except at lower Beaver River (*test* station BER-1, 0.093 mg/L), lower Firebag River (*test* station FIR-1, 0.034 mg/L), McClelland Lake (*test* station MCL-1, 0.020 mg/L), and McLean Creek (*test* station MCC-1, 0.011 mg/L). Concentrations measured at these locations using the ALS method were approximately one-sixth to one-thirtieth of the corresponding concentration found using the AITF method.

Lower Beaver River exhibited the highest concentration using either method; this creek is known to receive seepage from the Mildred Lake Settling Basin, although most is captured at the creek's head and pumped back into the holding basin (W. Zubot, Syncrude Ltd., *pers. comm.*, April 2010). McLean Creek also showed relatively high concentrations in the AITF method and has a highly modified upper watershed. Although the lower Firebag River and McClelland Lake are both defined as *test*, these stations have very little development in their upper watersheds (see Section 2), and exhibited AITF-determined levels of acid-extractable organics that were similar to the lowest ("background") values of all stations measured in the RAMP 2010 dataset by AITF.

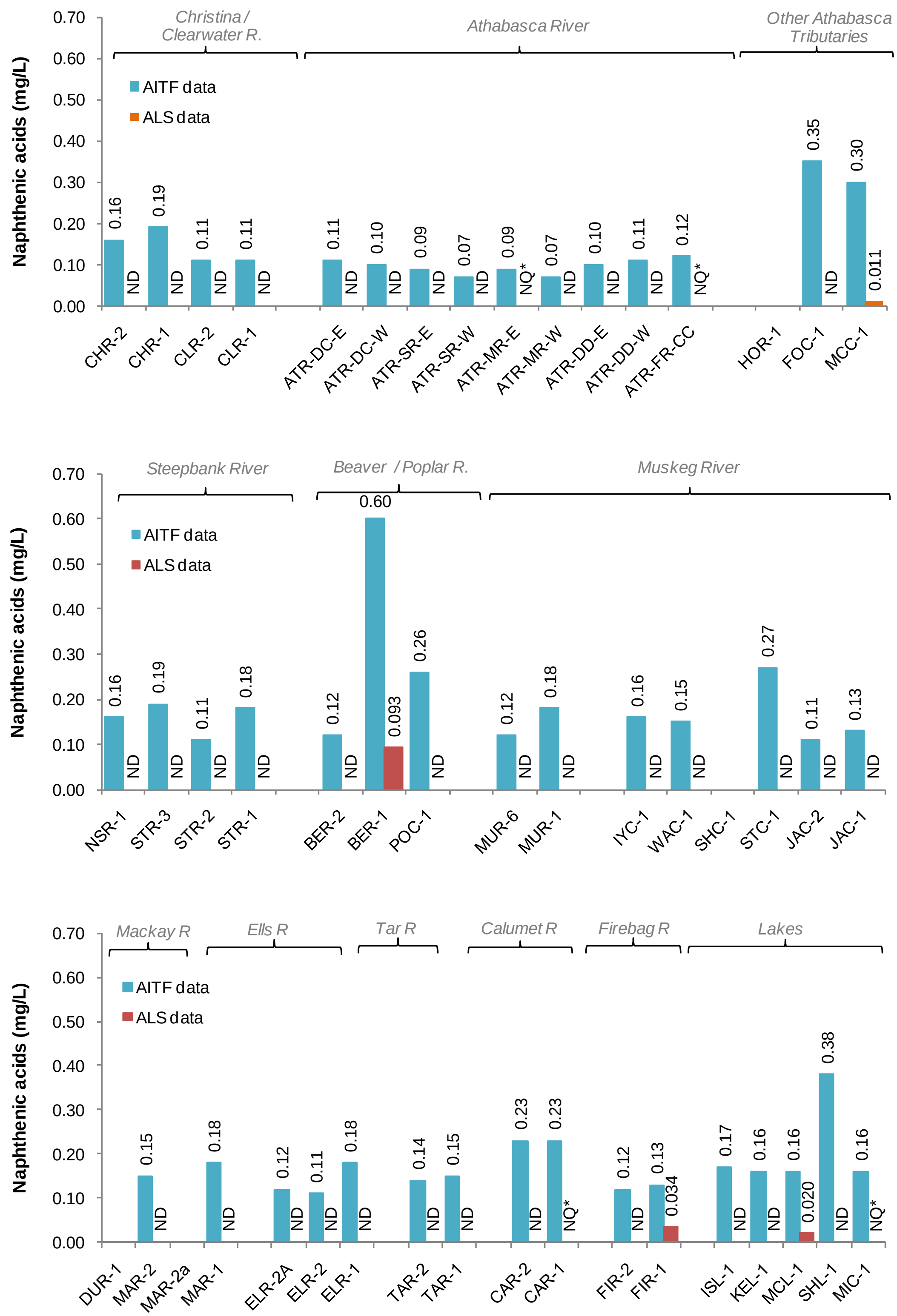
Comparison of data derived through these different methods suggests that: (a) the AITF method measures many more organic compounds than simply naphthenic acids; and (b) concentrations of organic acids conforming to the classic naphthenic acids formula ($C_nH_{2n+Z}O_2$) in ambient waters of the lower Athabasca watershed are low, with the majority of compounds detected by other methods likely being other acid-extractable organic compounds.

A2.1.4 Need for Clarity and Agreement Moving Forward

The environmental chemistry of acid-extractable organics ("naphthenic acids") in the oil sands region is continuously being clarified. However, analytical methods remain in flux, with numerous approaches currently being used or developed returning very different results and none having associated, endpoint-specific toxicological data for comparison. It is apparent that each of these methods is measuring a different set of compounds.

While each of these different methods may have advantages for specific applications, for effective environmental monitoring of the ambient aquatic environment in the oil sands region, it is important that a standard method for measurement of naphthenic acids and/or other acid-extractable organic compounds be identified for routine use by regulators, RAMP, other site-specific monitoring programs, and academic researchers. This method should be based on measurement of specific compounds or groups of compounds that have potential for toxicity at environmentally relevant concentrations. This may require a Toxicity Identification Evaluation (TIE) using OSPW samples or an analogous desktop study based on chemical characterization of OSPW and surrounding ambient surface waters in the region. In the absence of a clear toxicological understanding of what compounds are important, it will be difficult to develop and refine an appropriate test for acid-extractable organics in regional surface waters that can be linked to meaningful benchmarks of potential environmental change.

Figure A2.1-3 Concentrations of acid-extractable organic acids (naphthenic acids) measured by AITF for RAMP using three different quantification methods, fall 2010.



*NQ = not quantifiable due to sample matrix interferences.

A2.2 WATER QUALITY REGIONAL *BASELINE* ASSESSMENT

A2.2.1 Background

Although RAMP water quality data are screened against generic water quality guidelines published by the CCME, AENV, or other provincial jurisdictions (where CCME and AENV guidelines do not exist), use of such generic guidelines may not be appropriate in all circumstances, given natural, site-specific variability in water quality. For example, in the lower Athabasca River and its tributaries, concentrations of various metals may exceed generic guidelines, because insoluble metals present in suspended particulates may indicate high concentrations of metals. However, these particulate metals often are not bioavailable and thus typically contribute little toxicity relative to dissolved metals; this phenomenon has been documented in the region by several authors, including Corkum (1985), Hebben (2009), and Glozier *et al.* (2009). In its guidance for derivation of site-specific water quality objectives, CCME (2003) indicates that it is appropriate to develop site-specific objectives where "the generic water quality guideline for a substance is lower than the upper limit of background at a site under investigation", as is the case for many water quality variables in waterbodies monitored in the RAMP FSA.

Additionally, although RAMP collects water quality data from both *baseline* (upstream) and *test* (downstream) locations in several watersheds, this is not possible in some watersheds, where no *baseline* station may be available for use as an uninfluenced (reference) location for comparison with downstream conditions (e.g., very small watersheds, or those where substantial alteration occurred previous to RAMP's existence).

In the absence of region- or site-specific water quality objectives or thresholds provided by regulators or regional organizations such as CEMA, RAMP has developed a set of regional water quality benchmarks to address these two issues in its own assessments, from data collected by RAMP at *baseline* stations since 1997. These regional *baseline* ranges are intended to represent the range of natural variability in water quality in the region, for use in screening RAMP water quality data collected at both *baseline* and *test* stations. The intent of these benchmarks is to identify regionally meaningful changes in water quality.

Methods used to develop these regional *baseline* ranges are described in Section 3. Put simply, groups of stations that exhibit similar water quality over time are identified through cluster analysis, and water quality data from *baseline* stations (specifically, 5^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} percentiles) within these clusters are used for screening purposes. Observed values outside the central 90% of values (i.e., below 5^{th} percentile or above 95^{th} percentile) are flagged as being outside the documented range of natural variability (although it should be noted that 10% of *baseline* values will, by definition, fall outside this range). This approach is similar to "background concentration procedure" examples outlined in CCME (2003), which defined site-specific objectives using 90^{th} or 95^{th} percentiles of background values, or two standard deviations from the mean, which statistically is similar to using the central 95% of observations. It is also similar to reference-condition-approach (RCA) or bioassessment methods used for benthic invertebrate monitoring, which also are used by the Benthos and Sediment component of RAMP and discussed therein.

The recent RAMP Peer Review (AITF 2011) raised questions about the use of these regional *baseline* ranges as benchmarks in both the Water Quality and Benthic Invertebrate Communities components, particularly with respect to the pooling of spatial and temporal variability in the creation of these ranges. The following analysis of regional water quality characteristics and *baseline* ranges was undertaken to provide context for future discussion of these questions.

The suitability of regional *baseline* ranges as a representation of the range of natural variability in RAMP water quality assessments should consider the following:

1. Similarity of water quality at *baseline* stations within each cluster among stations and among years (i.e., consistency of cluster membership);
2. Variability of *baseline* water quality among stations within clusters; and
3. Variability of *baseline* water quality among years within clusters.

A2.2.2 Consistency of Cluster Membership

Previous RAMP assessments of this regional *baseline* approach have focused primarily on the first of the questions listed above (i.e., cluster membership). This topic has been discussed in some detail in previous RAMP technical reports, the RAMP Design and Rationale Document (RAMP 2009b), and elsewhere in this report (i.e., Section 3 and Appendix D).

Since this regional *baseline* method was adopted by RAMP in 2004, various modifications to the statistical approaches to clustering of water quality data have been made, mainly related to data selection and treatment prior to clustering. In 2010, multiple approaches to data pre-treatment and clustering were taken, to assess the potential effect of the analytical techniques used on the final clustering outcome, as discussed in Section 3. In all cases, three groups of stations with consistently similar water quality characteristics over time were identified, namely:

- **Cluster 1 –** Athabasca River mainstem;
- **Cluster 2 –** Tributaries predominantly located along the west bank of the Athabasca River, including the MacKay, Ells, Tar, and Calumet rivers; and
- **Cluster 3 –** Tributaries predominantly located along the east bank of the Athabasca River, including the Muskeg, Steepbank, and Firebag rivers.

These groups of watersheds exhibit various physiographic and hydrographic similarities. Obviously, the Athabasca River is substantially larger than all of its tributaries in the oil sands region, with only 14% of its drainage area occurring downstream of Fort McMurray (WSC 2011); its upper reaches flow through several different landforms and anthropognic developments, including industrial and municipal discharges. Relative to western tributaries to the lower Athabasca River, eastern tributaries generally are characterized by lower gradients, greater proportions of their headwaters comprised of poorly-drained muskeg and peatlands (GSC 2006, AAFC 2007). Annual runoff in eastern tributaries is more dominated by freshet than the higher-gradient western tributaries, with less extreme high flows, particularly in summer and fall, than western tributaries (RAMP hydrology data, RAMP database www.ramp-alberta.org).

Southern tributaries (i.e., Clearwater, Christina, Horse rivers) have not always grouped consistently within these three clusters, and over time have alternately grouped with either western tributaries (as in 2010) or, less frequently, with the Athabasca River mainstem. However, these southern tributaries do not consistently group separately either, based on their water quality. It should be noted that no water quality data from these rivers are used to generate regional *baseline* ranges for comparison, because all of these watersheds contain development upstream of RAMP sampling locations.

A2.2.3 Variability Within and Among Clusters

A2.2.3.1 Water Quality Characteristics Among Clusters

Stations within these groups/clusters exhibit consistent similarities in water quality, which have been observed repeatedly over time. Generally, concentrations of most metals are higher in the *baseline* Athabasca River stations (Cluster 1) than at *baseline* stations in tributaries sampled by RAMP, particularly for metals present primarily in particulate form, such as aluminum, antimony, copper, lead, silver, and titanium (total suspended solids are generally higher in the Athabasca River as well). However, this trend is reversed for several metals that are present predominantly in dissolved form, particularly boron, lithium, and manganese, which are present in higher concentrations in tributaries. Total dissolved solids and most major ions also are higher in tributaries than in the Athabasca River mainstem, with the notable exception of sodium, potassium, chloride and sulphate, which are all lower in eastern tributaries (Cluster 3) than either western tributaries (Cluster 2) or the Athabasca River mainstem.

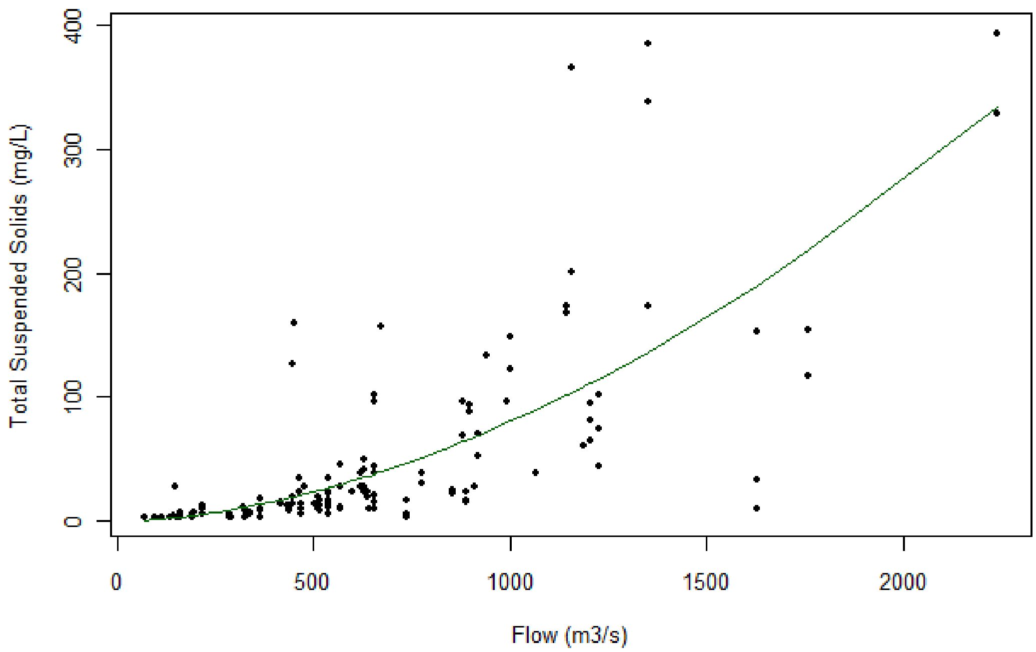
Indicators of organic substances—including total and dissolved organic carbon, total Kjeldahl nitrogen (TKN), total phenolics, and true colour—are typically much higher in tributaries than in the Athabasca River. These organic variables are generally higher in western tributaries than in eastern tributaries. Total sulphides follow a similar pattern.

Differences between tributary groups also are apparent, with water quality stations in eastern tributaries (Cluster 3) exhibiting an ion balance dominated by calcium/bicarbonate, whereas sodium, chloride and sulphate occur at greater concentrations in western tributaries (Cluster 2). Concentrations of most metals are higher in western tributaries than in eastern tributaries.

Spearman's rank correlations for within-cluster water quality data collected by RAMP from 2002 to 2010 (tabulated in Appendix D) reveals additional consistency in differences in water quality among clusters and across years. In the Athabasca River mainstem, concentrations of many variables are significantly correlated ($p<0.01$) with suspended solids (TSS), including most total metals (Al, Sb, As, Ba, Bi, Cr, Co, Cu, Fe, Pb, Li, Ni, Th, Ti, U, V), nutrients (TN, TKN, TP) and organic compounds (TOC, DOC, total phenolics). Conversely, conductivity and most major ions are negatively correlated with TSS, and weakly correlated (negatively or positively) with metals. These correlations suggest a primary influence of river flow on measured water quality in the Athabasca River mainstem, given the positive relationship between Athabasca River flow and TSS (Figure A2.2-1), and the converse, negative relationship between river flow and conductivity.

In tributaries to the lower Athabasca River, most total metals also are highly correlated with suspended materials ($p<0.01$; Cluster 2: Al, Be, Bi, Cd, Cr, Co, Cu, Fe, Pb, Hg, Ag, Th, Tl, Sn, Ti, V, Zn; Cluster 3: Al, Ba, Be, Cd, Cr, Co, Fe, Pb, Mn, Hg, Tl, Th, Sn, Ti, U, V). In all tributaries, several metals typically occurring in dissolved form were positively correlated with major ion concentrations (especially Ba, B, Li, Mn, Sr). Positive correlations of most metals with major ions were stronger in western tributaries, while correlations of metals with suspended solids were stronger in eastern tributaries. Sodium and (especially) chloride in western tributaries did not correlate as strongly with other major ions as in eastern tributaries, which could perhaps indicate point-source influences on water quality at these *baseline* stations of saline seeps, which are known to occur in the region.

Figure A2.2-1 Relationship between river discharge and total suspended solids in the lower Athabasca River, 1997 to 2009.



Contrary to their behaviour in the Athabasca River mainstem, organic compounds and most nutrients in these tributaries were at most weakly correlated with suspended materials (although weakly positively in eastern tributaries, and weakly negatively in western tributaries). These compounds (TOC, DOC, TN, TKN, colour, total phenolics) were strongly correlated with each other in all tributaries; sulphide also was highly correlated with these variables.

Together, these tendencies suggest three dominant components of water quality in streams of the RAMP FSA: (i) particulate-associated materials, which are predominantly comprised of particulate metals; (ii) major ions and some dissolved metals, likely associated with groundwater sources in tributaries; and (iii) organic compounds, which covary among themselves, and are associated with suspended materials in the Athabasca mainstem, but appear to vary more independently of other water quality variables in tributaries.

A separate correlation analysis of all water quality data from 2009 and 2010 only, focused on examining correlations between naphthenic acids measured using high-resolution methods and other water quality variables, identified significant, strong correlations ($p<0.01$) between naphthenic acids and indicators of flow in the Athabasca River mainstem (i.e., positive correlations with most metals and organic compounds, and negative correlations with major ions and conductivity), indicating higher concentrations of acid-extractable organics at higher flows. However, correlations were reversed in tributaries, where naphthenic acids were strongly, positively correlated with major ions. In western tributaries, naphthenic acids were strongly associated with all major ions, TDS, and conductivity, whereas in eastern tributaries, naphthenic acids were strongly associated only with chloride and sulphate. Considered in parallel with the possible range of acid-extractable organic compounds measured by this analysis (see Section 2.1), these patterns suggest that these compounds measured in the Athabasca River mainstem are likely predominantly comprised of fatty acids and other related organic acids originating upstream of Fort McMurray, while organic acids in western tributaries are largely associated with influences of groundwater (which may have high organic-acid and TDS concentrations [AENV 2009]), and that organic acids measured at stations in eastern tributaries may have mixed origins, from both groundwater and from biological decomposition in these watersheds.

A2.2.3.2 Among-Year Variability Within Clusters

Year-specific regional *baseline* ranges for selected RAMP water quality measurement endpoints are shown in Figure A2.2-2, specifically an indicator of suspended materials (TSS), ion content (total alkalinity), dissolved metal (total boron) and organic content (DOC). Each figure presents 5^{th}, inter-quartile, and 95^{th} percentile ranges for each variable in each year (blue) from 1997 to 2010, as well as the overall (1997 to 2010) *baseline* range (red) used for water quality screening. Each plot also includes a representative average daily river discharge from September 1 to 15 for each year; these discharge data (collected by RAMP and Water Survey of Canada) are for the Athabasca River downstream of Fort McMurray, and the mouths of the Muskeg River (RAMP station S7, compared with Cluster 2) and Mackay River (RAMP station S26, compared with Cluster 3).

Examination of the TSS plots indicates greatest inter-annual variability of suspended materials in the Athabasca River mainstem, with highest TSS generally occurring in years of highest flow (i.e., 2004 and 2010). Clear relationships with flow are not apparent in data for tributaries. For the Athabasca River, the defined 95^{th} percentile for 1997 to 2010 is near the median concentration in high-flow years, but above the 95^{th} percentile for all other years. For the western tributaries, the 1997 to 2010 95^{th} percentile is within the 95^{th} percentile of three of eleven years of data, and the 75^{th} percentile is below the 95^{th} percentile for most years.

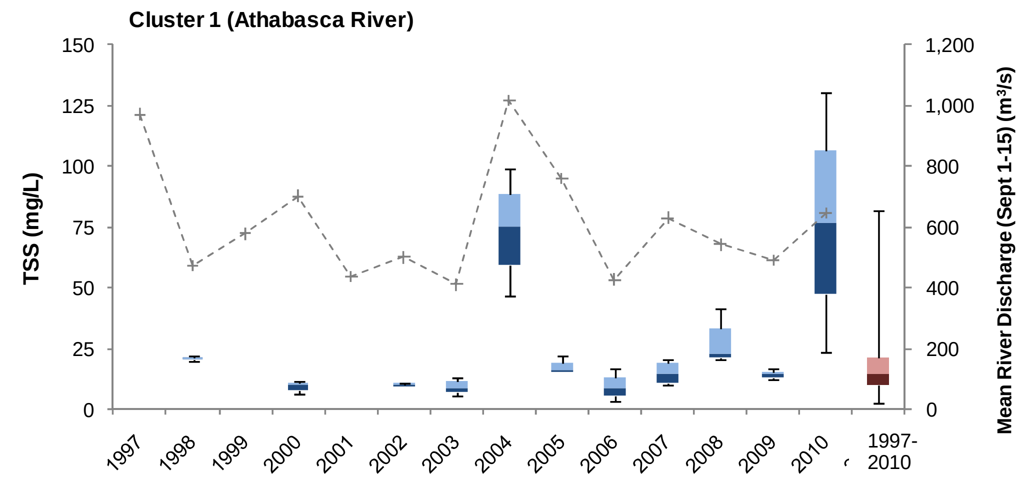
For total alkalinity, an influence of flow on ion composition was seen most clearly in eastern tributaries, while alkalinity in the Athabasca River and western tributaries showed more consistent values over all years. In both the Athabasca River and eastern tributaries, the 1997-2010 95^{th} percentile exceeded individual 95^{th} percentiles annually; for western tributaries, the cumulative 95^{th} percentile was below annual 95^{th} percentiles observed in 1998 and 1999.

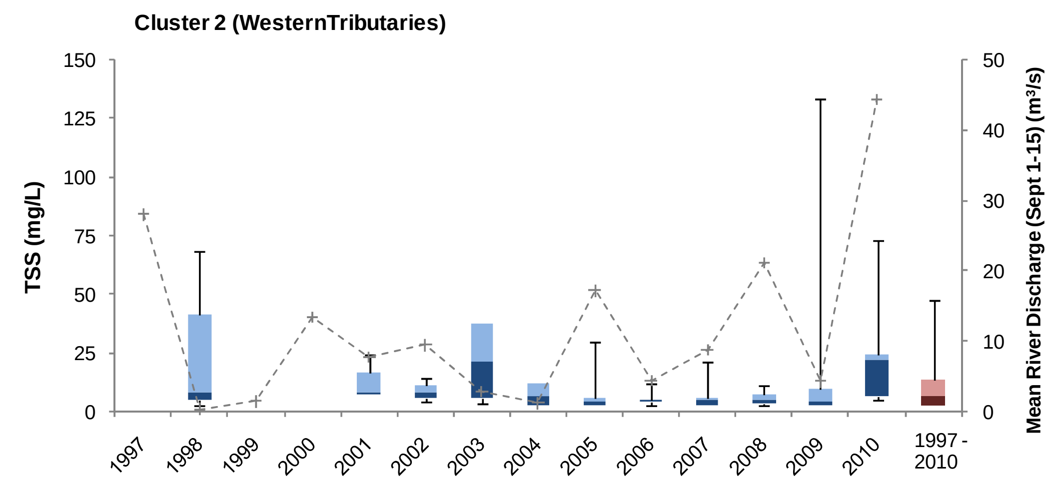
For total boron, concentrations were generally low and consistent in the Athabasca River relative to tributaries, and the associated 1997 to 2010 95^{th} percentile was correspondingly tight. In tributaries, a slight inverse relationship with river flow was suggested, particularly in eastern tributaries. In both sets of tributaries, the 1907 to 2010 95^{th} percentile is below that of three years, and the cumulative 75^{th} percentile is below that of several years.

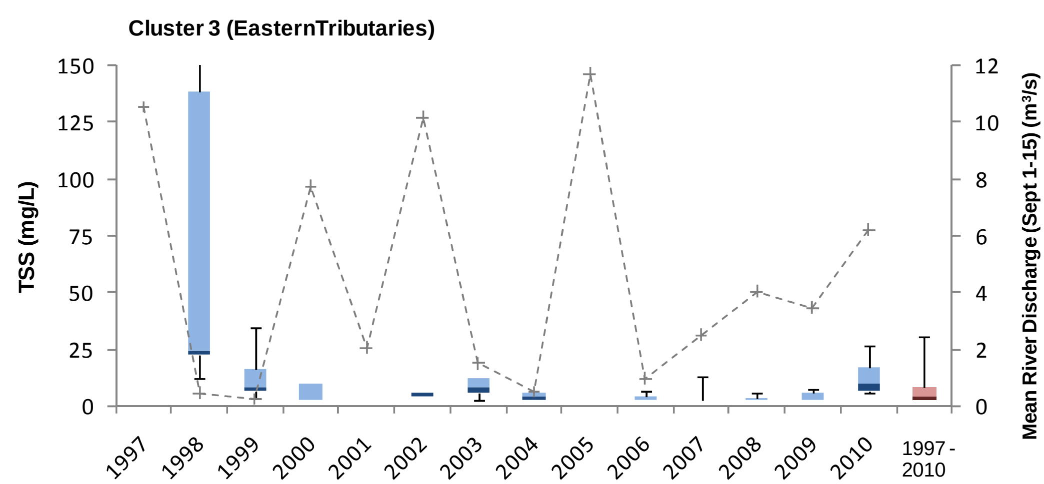
The DOC plot shows the clear influence of flow on organic content in the Athabasca River, and a weaker but still apparent influence of flow on DOC in tributaries. In all groups, the cumulative 95^{th} percentile falls near or below that of multiple years of observations.

Figure A2.2-3 shows average within-year coefficient of variation (CV, equal to standard deviation/mean, from all annual observations) against among-year CV (standard deviation of annual means/grand mean), for selected major ions, suspended solids, nutrients and organic compounds, and metals. Where among-year variability exceeds within-year variability, data fall above the 1:1 line on each chart. Data falling below this 1:1 line indicate greater within-year variability than among-year variability.

Figure A2.2-2 Annual and cumulative regional baseline ranges among clusters, 1997 to 2010, compared with fall river discharge (dashed line)[1].

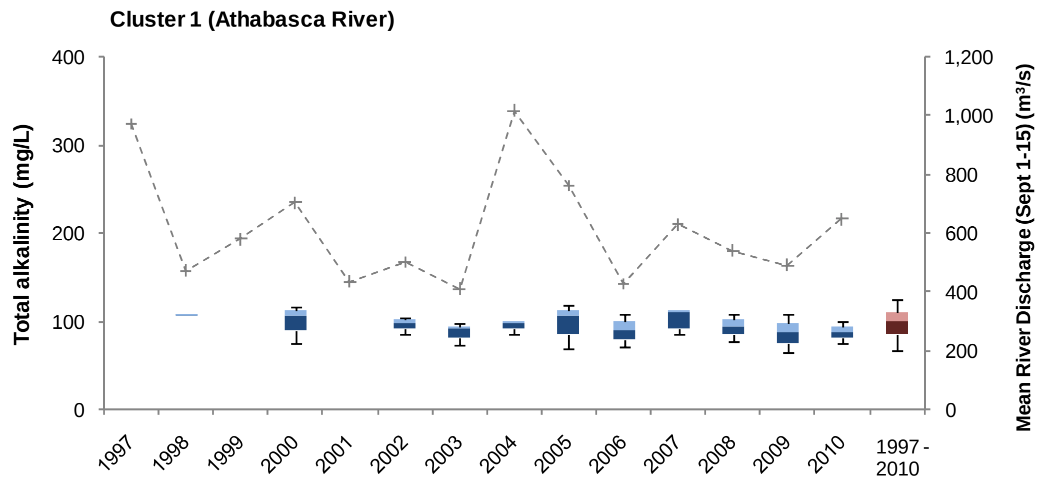


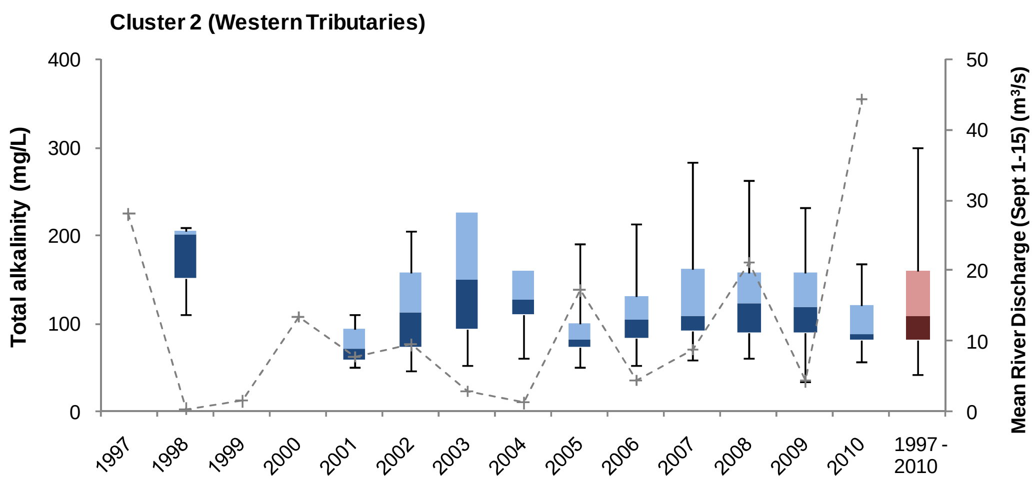


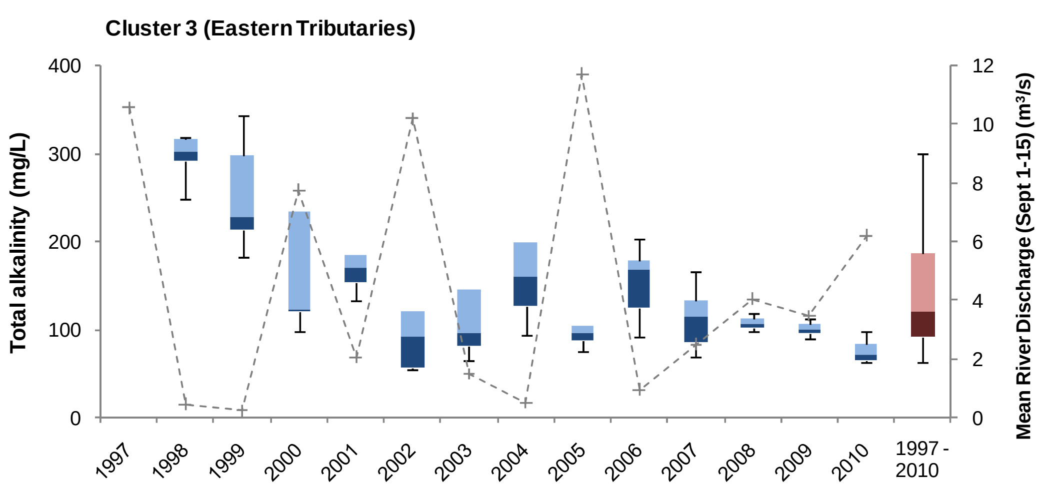


[1] River flows: Cluster 1-Athabasca R. (WSC 07DA001); Cluster 2-MacKay R. (RAMP S26); Cluster 3-Muskeg R. (RAMP S6).

Figure A2.2-2 (Cont'd.)

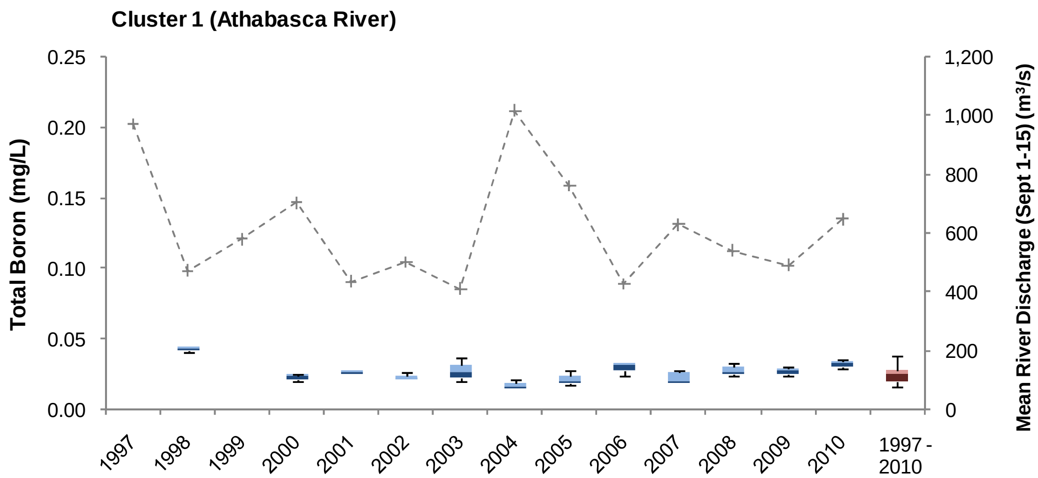


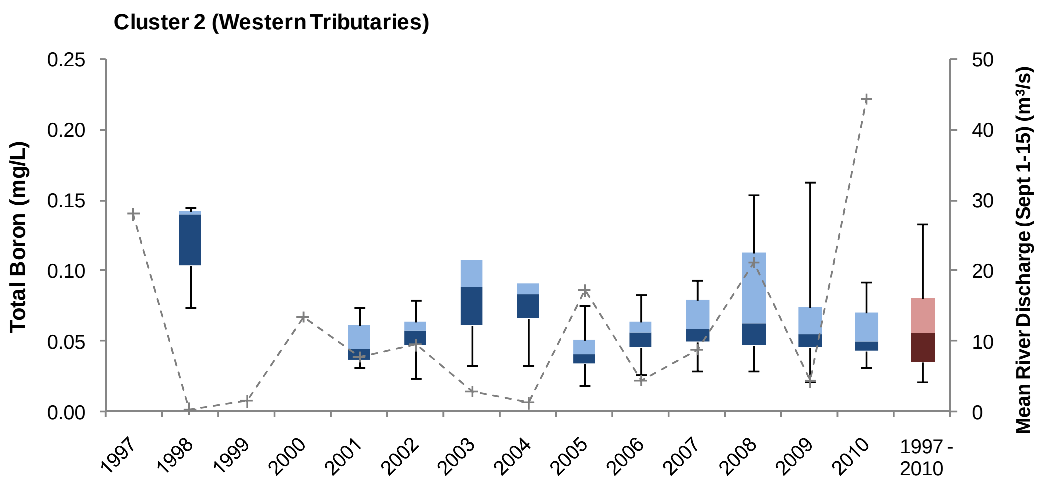


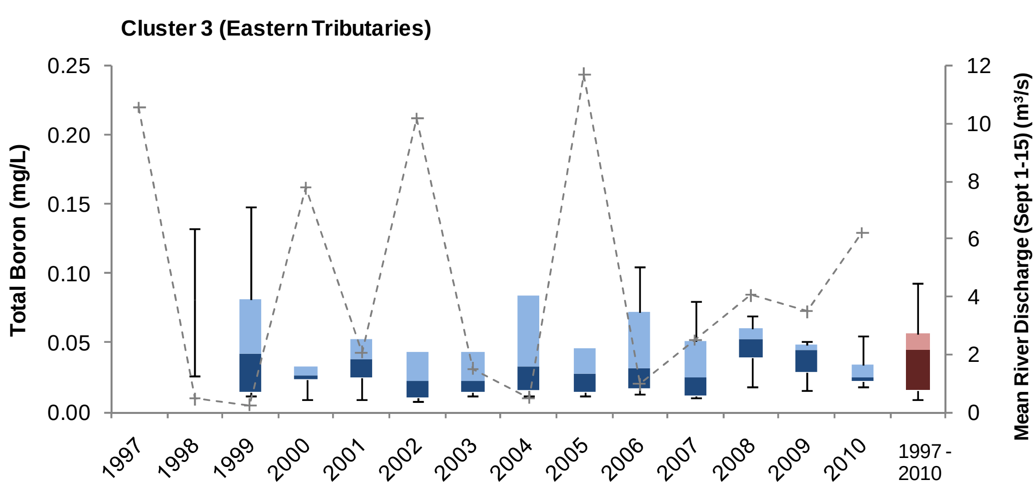


[1] River flows: Cluster 1-Athabasca R. (WSC 07DA001); Cluster 2-MacKay R. (RAMP S26); Cluster 3-Muskeg R. (RAMP S6).

Figure A2.2-2 (Cont'd.)

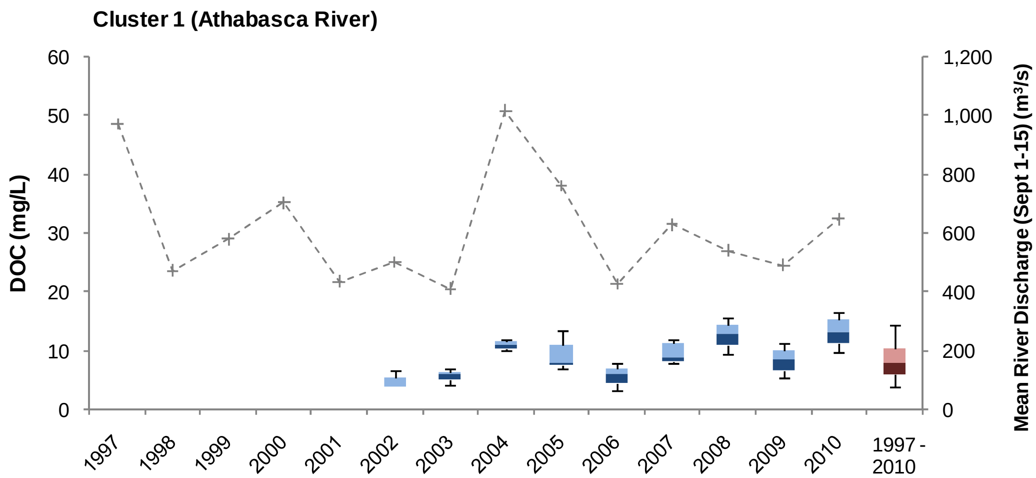


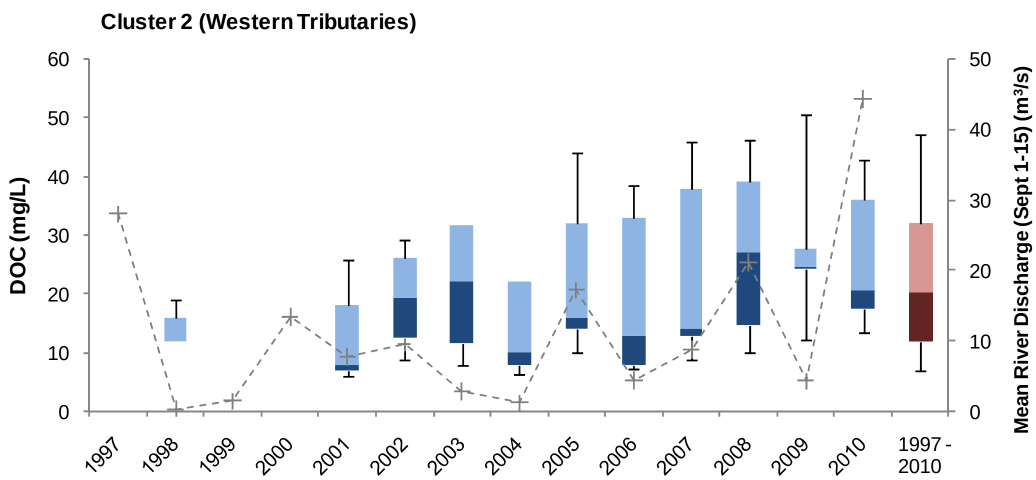


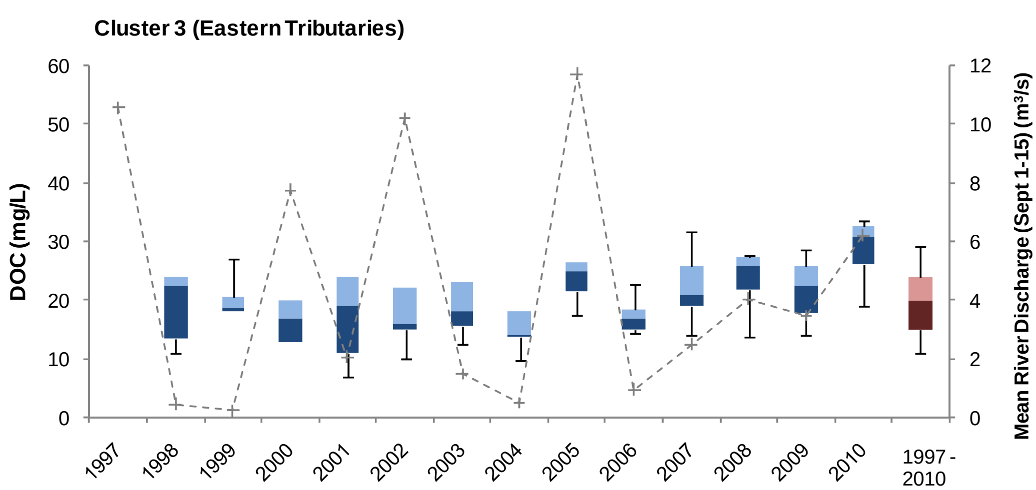


[1] River flows: Cluster 1-Athabasca R. (WSC 07DA001); Cluster 2-MacKay R. (RAMP S26); Cluster 3-Muskeg R. (RAMP S6).

Figure A2.2-2 (Cont'd.)



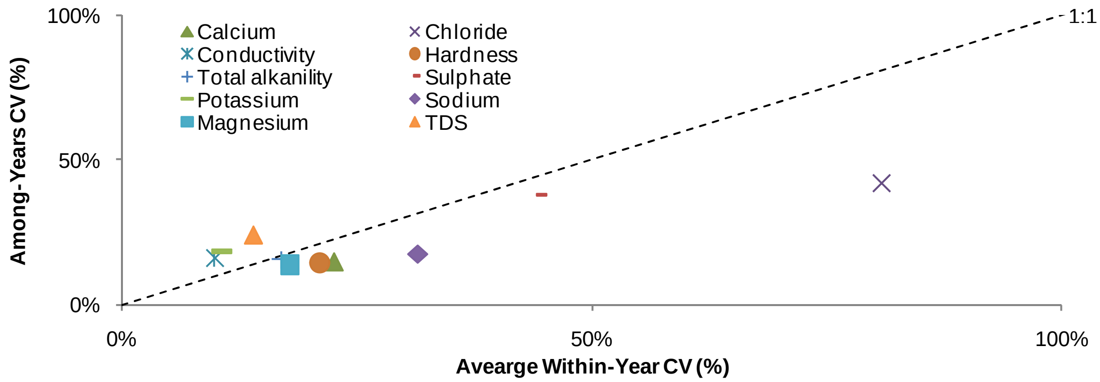




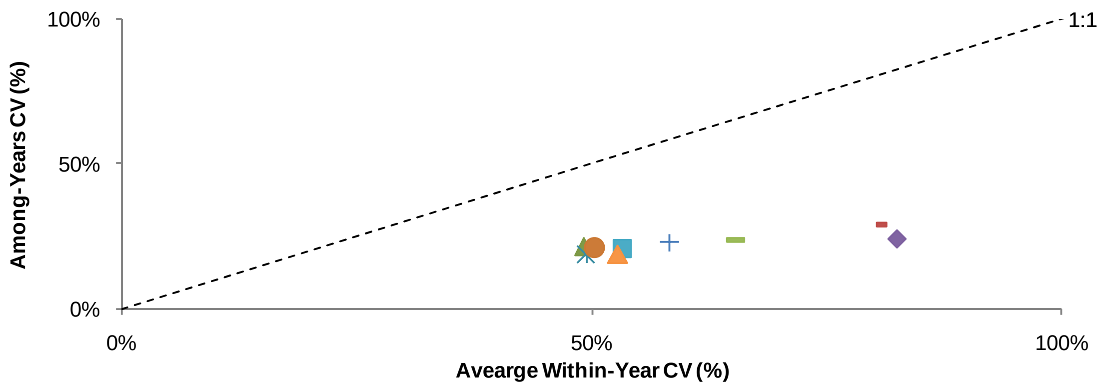
[1] River flows: Cluster 1-Athabasca R. (WSC 07DA001); Cluster 2-MacKay R. (RAMP S26); Cluster 3-Muskeg R. (RAMP S6).

Figure A2.2-3 Comparison of within-year and among-year variability for selected water quality variables measured by RAMP, 1997 to 2010.

Major ions: Cluster 1 (Athabasca River)



Major ions: Cluster 2 (Eastern Tributaries)



Major ions: Cluster 3 (Western Tributaries)

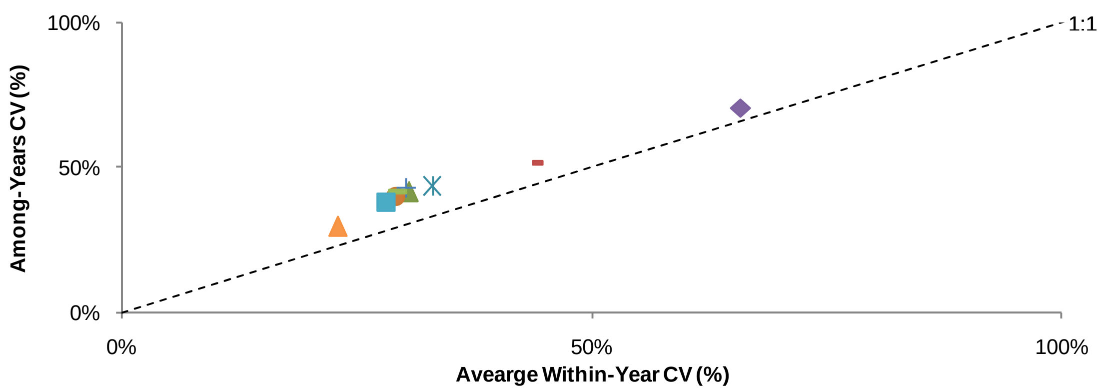
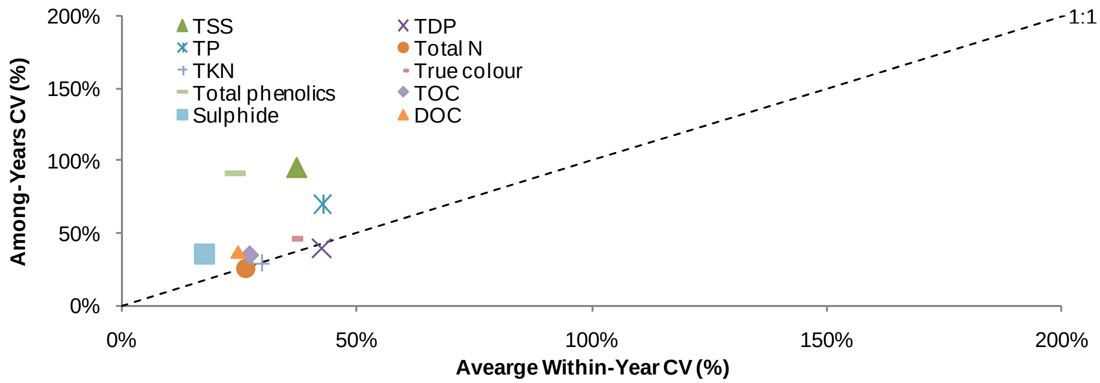
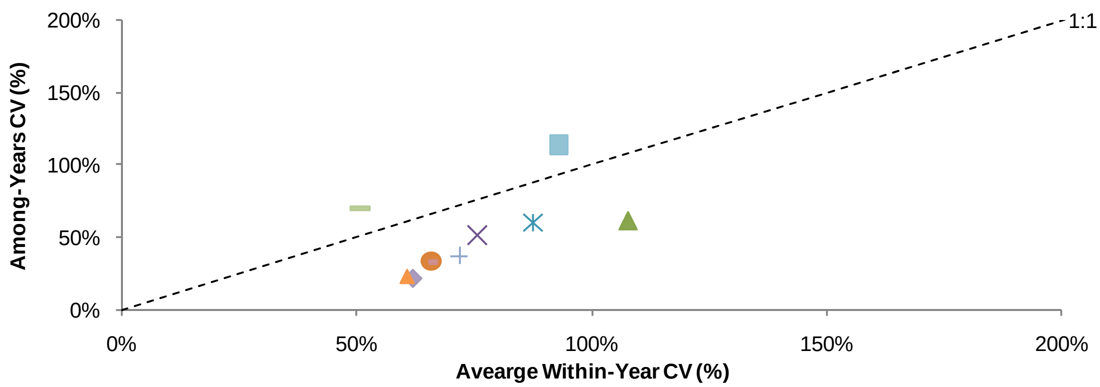


Figure A2.2-3 (Cont'd.)

Suspended solids, organic compounds and nutrients: Cluster 1 (Athabasca River)



Suspended solids, organic compounds and nutrients: Cluster 2 (Eastern Tributaries)



Suspended solids, organic compounds and nutrients: Cluster 3 (Western Tributaries)

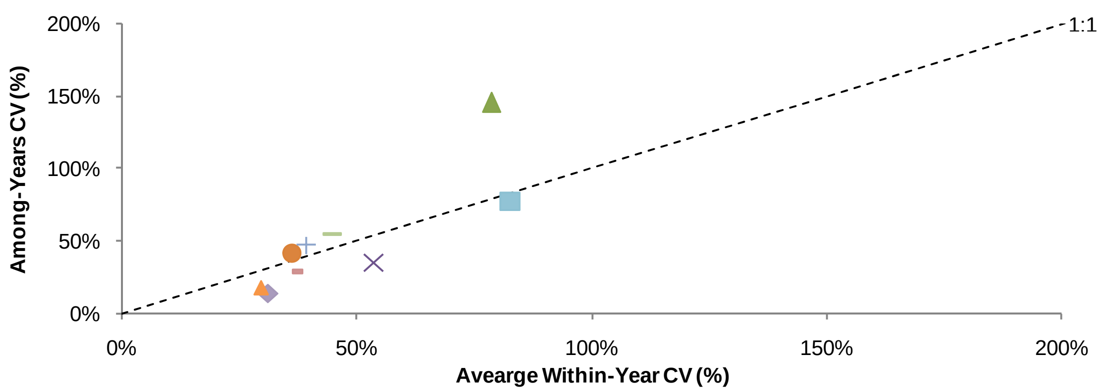
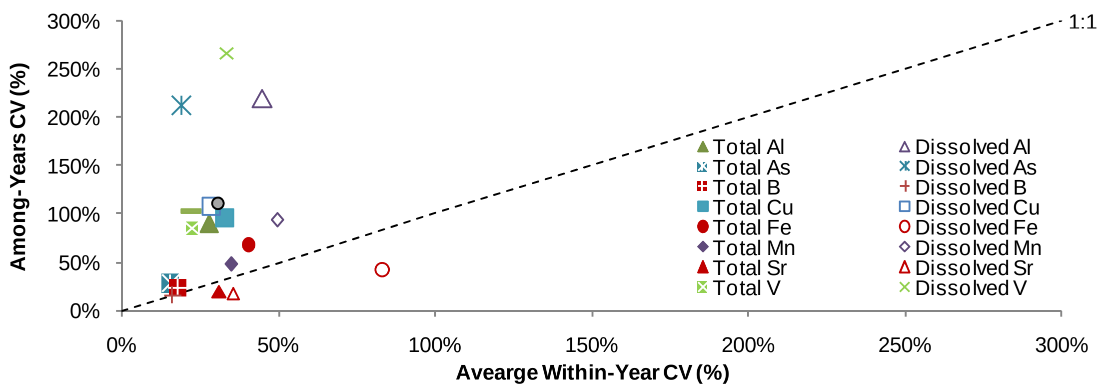
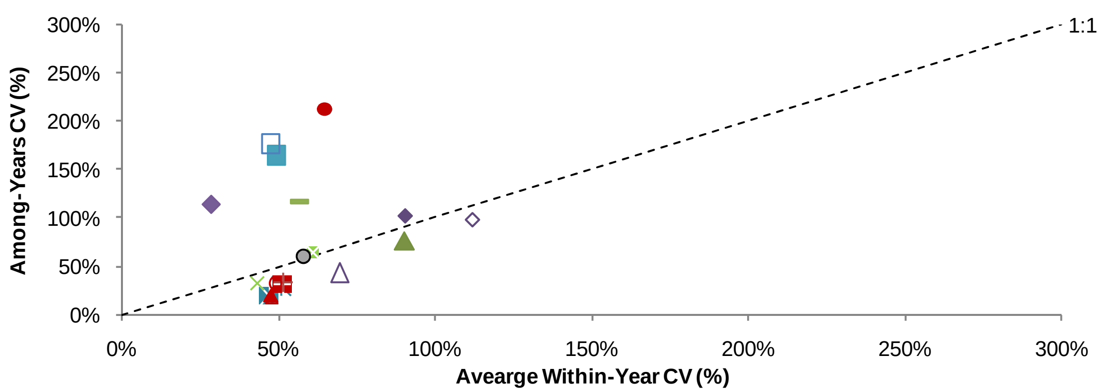


Figure A2.2-3 (Cont'd.)

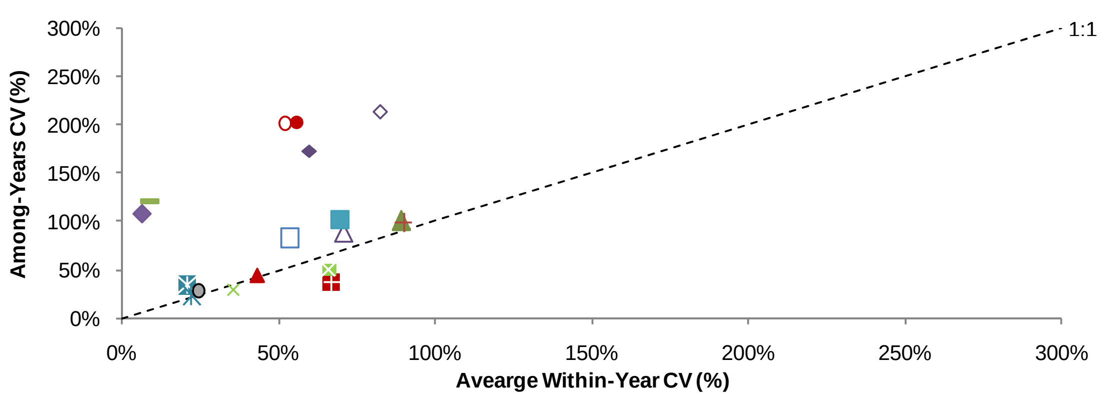
Selected total and dissolved metals: Cluster 1 (Athabasca River)



Selected total and dissolved metals: Cluster 2 (Eastern Tributaries)



Selected total and dissolved metals: Cluster 3 (Western Tributaries)



Differences among clusters and water quality variables are apparent. For major ions, within-year variability is generally similar or greater than among-year variability, particularly in eastern tributaries (consistent with the total alkalinity plot in Figure A2.2-2 mentioned previously). For suspended solids, organics and nutrients, variability was generally greater within years than among years in tributaries, but generally greater among years in the Athabasca River (this may be expected, given the strong influence of river flow on water quality in the Athabasca mainstem). For total and dissolved metals, variability was generally higher among years than within years, particularly in the Athabasca River mainstem; this is consistent with the strong influence of flow on metal concentrations, particularly in the Athabasca River. Metals present predominantly in dissolved form (i.e., B, Sr) typically showed less inter-annual variability than other metals and more similar to major ions than other metals in this regard.

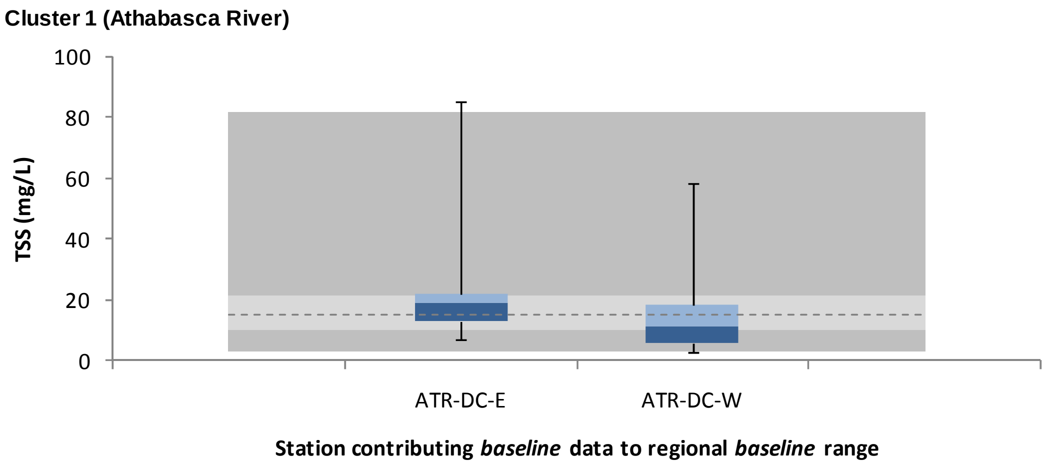
A2.2.3.3 Among-Station Variability Within Clusters

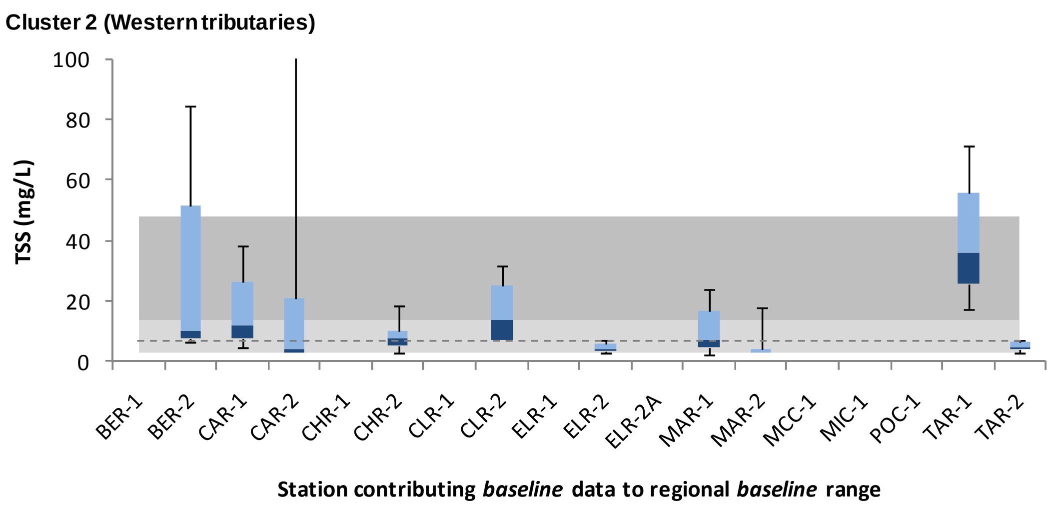
Figure A2.2-4 shows *baseline* data from individual stations that were used to generate regional *baseline* ranges for each cluster, using the same example variables as used in Section A2.2.3.2 above.

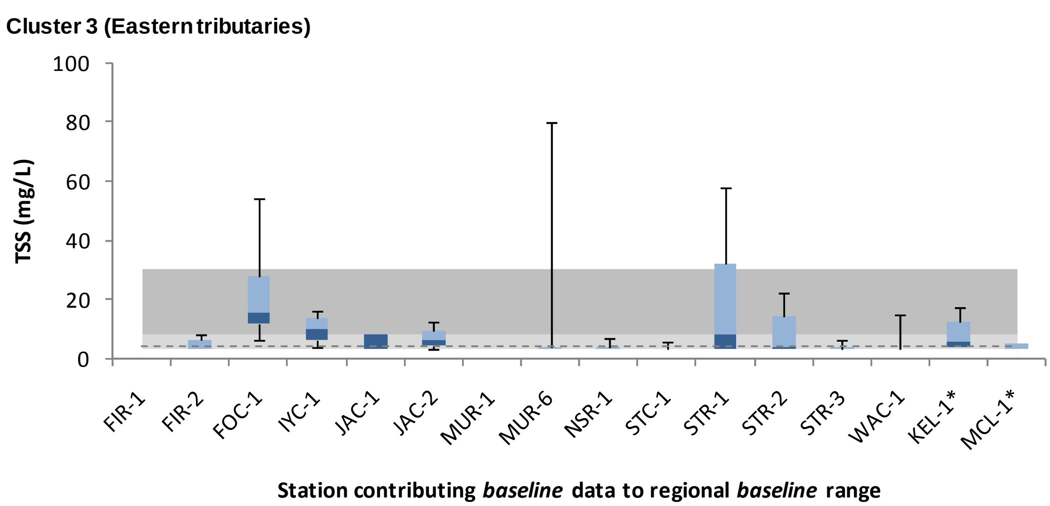
Grey background ranges in these figures correspond to regional baseline ranges as used in Section 5 of this report (i.e., 5th, 25th, 50th, 75th, and 95th percentiles); station-specific box-whisker plots correspond to similar percentiles of data within each station, with 5th and 95th percentiles represented as error bars, as previously done in Figure A2.2-2. Stations in these figures showing no data are those that use these cluster ranges for comparison, but that did not themselves contribute data to regional *baseline* ranges, typically because these stations had *test* status since RAMP sampling began at those locations. For stations that revert from baseline to test during their sampling history, only data from years of baseline status are included in regional ranges and in these graphs. For Cluster 3 (eastern tributaries), data from Kearl and McClelland lakes also are presented, although these data were not included in regional baseline ranges used for screening in the 2010 report; these lake data are discussed further in Section A2.2.3.4 below.

Although variability in water quality among stations within clusters is evident, median values for specific variables generally fall within the inter-quartile range for each cluster, with some exceptions. Some stations showed median values for specific water quality variables that fell below the inter-quartile range, particularly those in upper reaches of watersheds (e.g., FIR-2, NSR-1, IYC-1, STC-1), although this was not always the case. Generally, upper-watershed stations and those from smaller watersheds (e.g., Fort Creek Calumet River) appeared to show greater variability than those from larger watersheds. Water quality in the two stations on the Athabasca River upstream of oil-sands development (upstream of Donald Creek, west and east banks) shows consistent differences (consistent with the influence of the Clearwater River along the east bank at this location) although these are generally smaller than differences observed among tributary stations.

Figure A2.2-4 Variation within and among stations comprising regional *baseline* ranges (1997 to 2010 data).

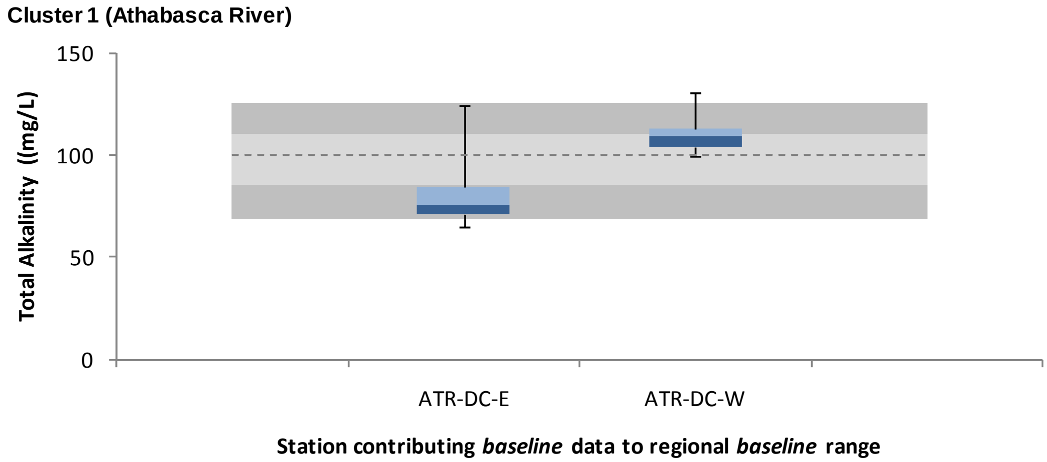


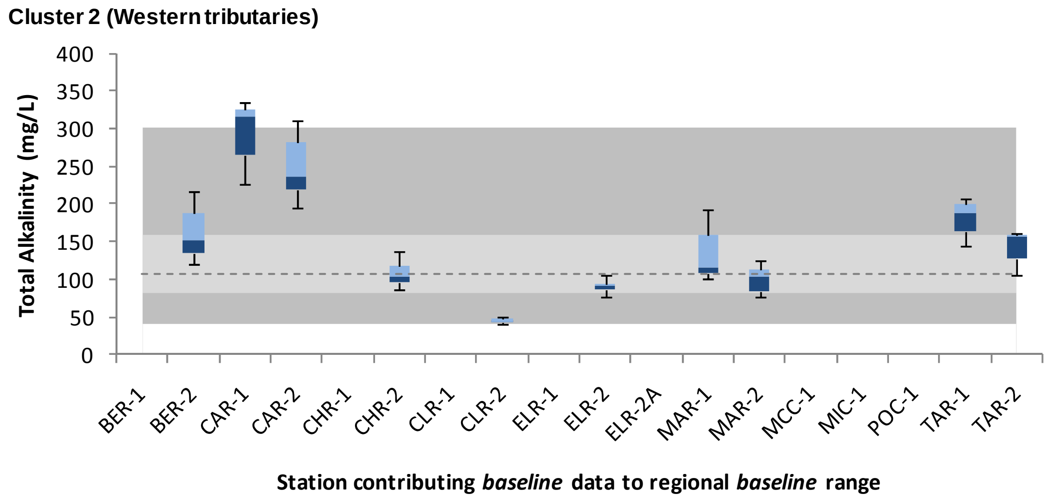


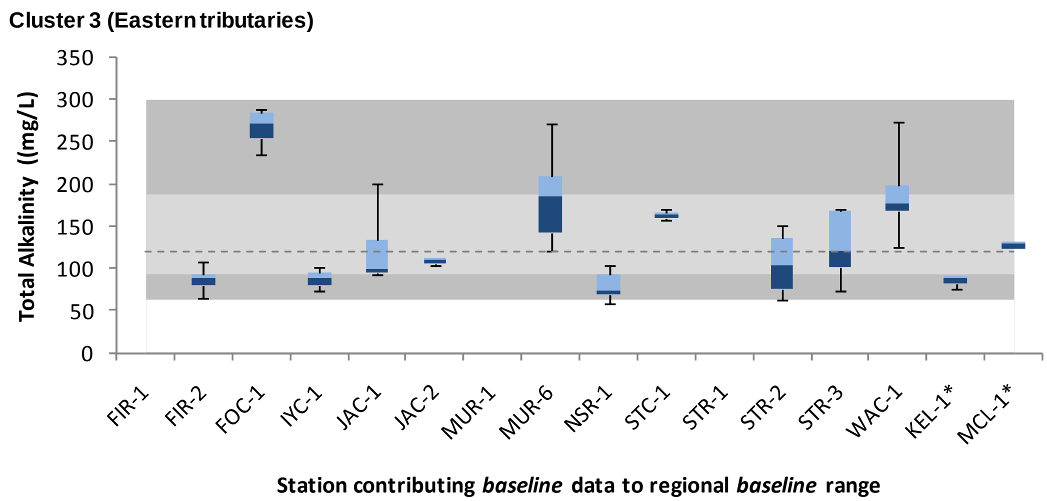


* Kearl Lake (KEL-1) and McClelland Lake (MCL-1) excluded from regional baseline calculations in 2010 (see Section 3).

Figure A2.2-4 (Cont'd.)

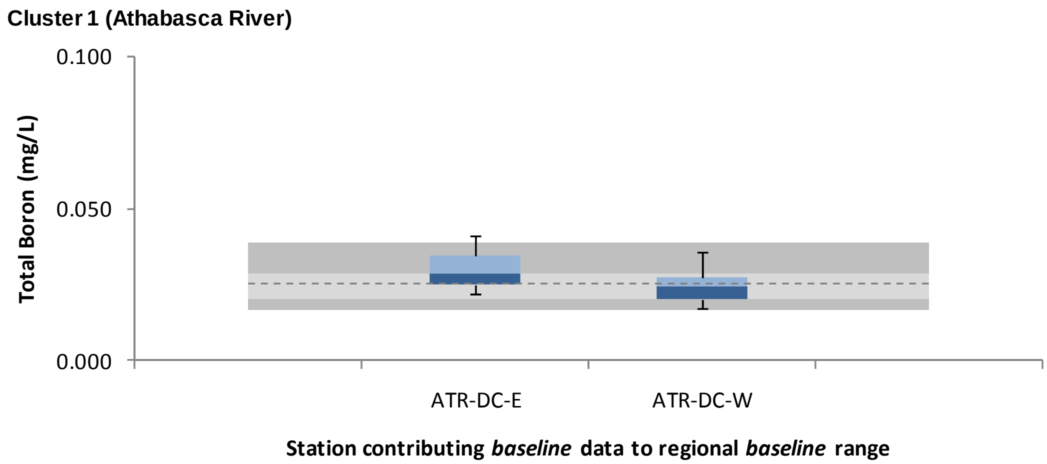


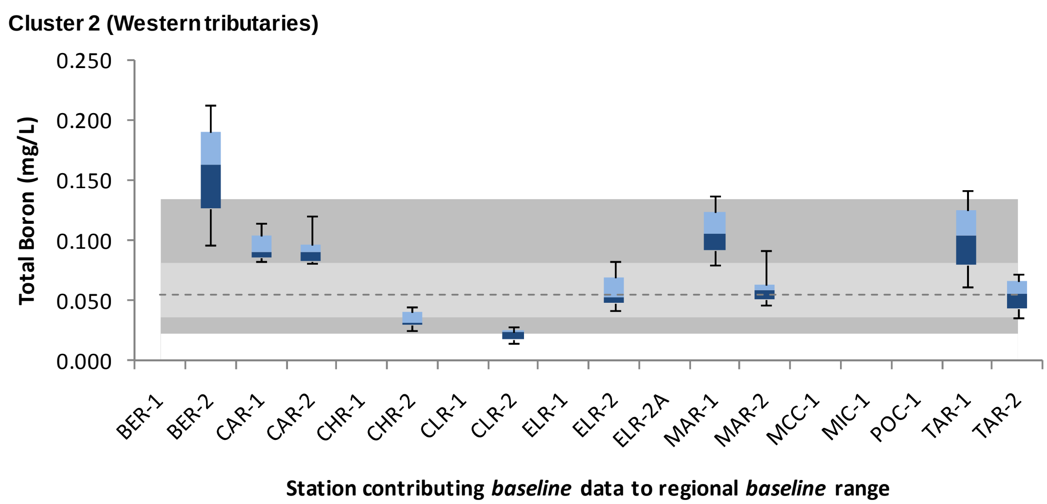


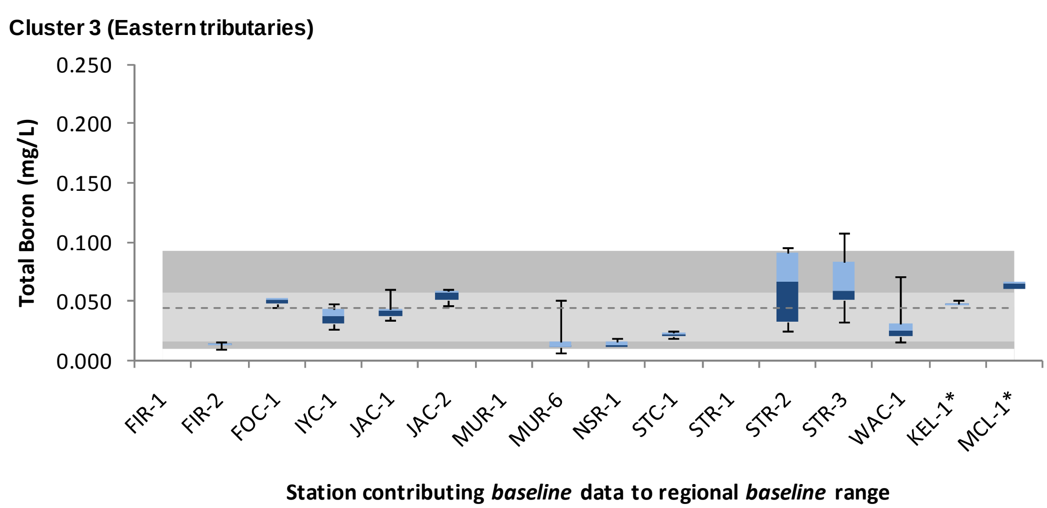


* Kearl Lake (KEL-1) and McClelland Lake (MCL-1) excluded from regional baseline calculations in 2010 (see Section 3).

Figure A2.2-4 (Cont'd.)

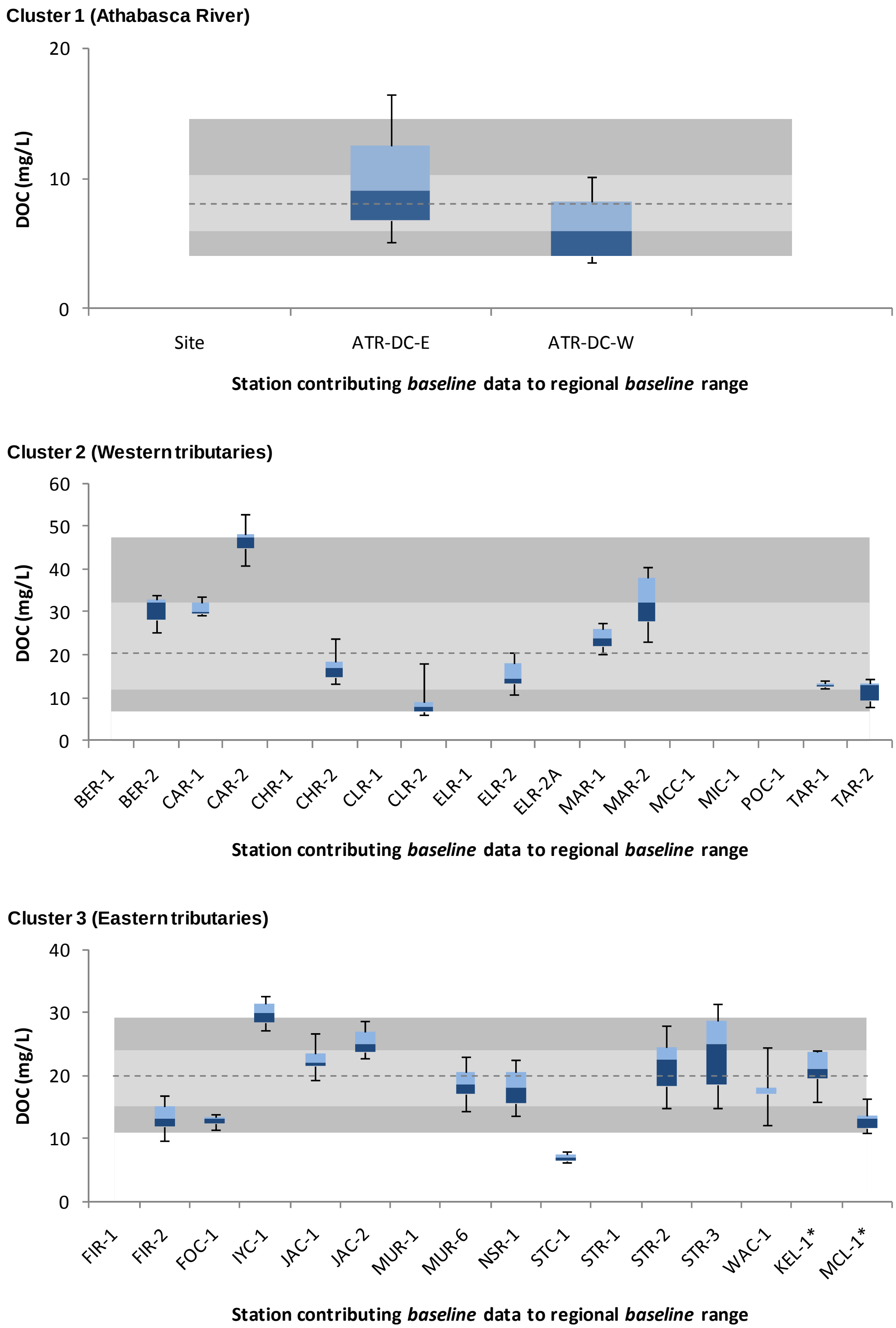






* Kearl Lake (KEL-1) and McClelland Lake (MCL-1) excluded from regional baseline calculations in 2010 (see Section 3).

Figure A2.2-4 (Cont'd.)



* Kearl Lake (KEL-1) and McClelland Lake (MCL-1) excluded from regional baseline calculations in 2010 (see Section 3).

A2.2.3.4 Similarities Between *Baseline* Data from Lakes and Streams

Figure A2.2-4 also includes *baseline* data collected from Kearl Lake (1998 to 2009) and McClelland Lake (2000 to 2009), which were included in regional *baseline* ranges calculated in RAMP Technical Reports from 2004 to 2009 (RAMP 2005, 2006, 2007, 2008, 2009a). These data were excluded from regional *baseline* ranges in the 2010 analysis due to a stated concern in the 2010 RAMP Peer Review (AITF 2011) that combining water quality data from these lakes would increase the range of regional *baseline* data used to comparison in the RAMP Technical Report, and potentially mask variability in stream water quality that was outside of background ranges of variability.

As is apparent for the variables shown in Figure A2.2-4, water quality in these shallow lakes is generally similar to water quality in streams. Generally, *baseline* water quality in these lakes fell within the inter-quartile range of regional baseline values, and/or was within the range of water quality observations of streams occurring within watersheds containing these lakes (i.e., the Firebag watershed for McClelland Lake, and the Muskeg watershed for Kearl Lake). These results suggest that inclusion of water quality data from these lakes in the regional *baseline* range did not inflate variability of these ranges to an extent that would obscure any excursions of regional *baseline* conditions in stream water quality.

A2.2.4 Future Considerations

A common factor among reference-condition approaches undertaken in RAMP and elsewhere is the aggregation of *baseline*/reference data across years. An underlying assumption of this aggregation is that conditions (water quality, benthic invertebrate community, fish community, etc.) in any given year at a *baseline* location are representative of natural conditions that are sufficient to support aquatic species that have become adapted over time to sustain their populations at this location. However, it may be possible for background conditions in a waterbody to change naturally in ways that cause significant, negative effects on resident biological communities, or that aquatic organisms in one watershed may be incapable of persisting in another nearby waterbody for some reason.

For water quality, specifically, this assumption is best tested by examining biological communities (e.g., benthos and/or fish) at corresponding locations and times with water quality; if community metrics indicate regionally normal (healthy) communities, then presumably water quality also continues to be regionally acceptable. Such effects-based assessments comprise the core of other components of RAMP, and provide a feedback mechanism between the stressor- and effects-based elements of RAMP. Further comparisons of water quality with biological endpoints at various *baseline* locations over time will help to determine the adequacy of regional water quality for maintenance of aquatic life.

The use of the regional *baseline* approach in RAMP and elsewhere is an attempt to define a range of natural variability that is considered acceptable to sustain aquatic life, so that any changes outside that range (i.e., that may threaten aquatic life) may be identified to decision-makers. Given that every sample collected in time and space may be considered unique, the key question to address in designing an analytical framework for regional analysis is: how much change is acceptable? Or, more technically, what are the effect criteria for the assessment? Such questions depend on philosophical questions of what are normal and social considerations of what is acceptable, as much as scientific questions of how these questions are may be defined and stated numerically.

The approach taken in the RAMP water quality component has successfully identified changes in water quality in one or many variables, in several watersheds, since its first implementation in 2004. However, this approach could be further refined though:

- ongoing, paired comparisons with benthic invertebrate community data in *baseline* areas; and
- more comparisons of water quality with hydrometric data and landscape variables, to better understand underlying factors that help determine water quality at a given location.

A2.2.5 Alternatives

Alternatively, screening of RAMP data to regional *baseline* ranges could be discontinued. As more data are collected at both *baseline* and (especially) *test* stations year to year, time trend analysis (using various statistical or control-charting techniques) can play a larger role in the identification of meaningful environmental change at locations monitored by RAMP.

Additionally, use of a percentile of background concentrations may be confusing to some reviewers, as, by definition, exceedances beyond these percentile ranges are expected to occur routinely (i.e., 10% of the time) at *baseline* stations. If there is a desire for more absolute, "not-to-exceed" objectives, use of objectives defined as a subset of background values is not an acceptable approach. For specific watersheds of high interest, such as the Athabasca River mainstem or larger tributaries rivers such as the Muskeg, Steepbank, Mackay, Ells, and Firebag rivers, consideration could be given to development of river-specific water quality objectives (SSWQOs), following methods outlined by CCME (2003) or others, which may incorporate direct toxicological assessments or adjustments of existing toxicological data for resident species. However, the drawback to defining SSWQOs in this way is that it would require development of separate SSWQOs for every water quality variable of interest or concern, independently for each watershed.

A2.3 COMPARISON OF NEILL-HESS AND KICK NET SAMPLING FOR BENTHIC INVERTEBRATE COMMUNITIES

A2.3.1 Introduction

Water levels were high in early September in many of the river reaches because of heavy rainfall in late August and early September. Water levels in most of the erosional reaches (i.e., MacKay River, Steepbank River, and the Firebag River) were high enough that the Neill-Hess cylinder was overtopped, effectively compromising sample integrity (overtopping of the cylinder causes organisms to be flushed from the sample). Sampling of these three rivers was; therefore, postponed until late September when water levels had potentially subsided. Water levels, even in later September had not sufficiently dropped in which case there were some stations within reaches where the Neill-Hess cylinder could not be used. At these locations, a D-framed net was used to a collect a "qualitative" kick samples using protocols from the federal CABIN methodology (Reynoldson *et al*. 2004). Given that kick net samples can be collected under many conditions and because it is possible that high water levels may compromise sampling in future, it was considered appropriate to collect kick net samples synoptically with some Neill-Hess cylinder samples for comparative purposes.

The objective of this analysis was to quantify the influence of the method of sample collection on values of measurement endpoints for benthic invertebrate communities.

A2.3.2 Methods

A2.3.2.1 Field

Kick net samples at a station (i.e., replicate sampling location within a reach) were collected by walking and kicking substrate along transects for three minutes in a zig-zag fashion, walking from the river's wetted perimeter towards the mid-channel to a maximum depth of approximately 1 m. Debris produced from kicking was collected in a D-framed net with 400 μm mesh.

Kick net samples were collected from the following reaches (Figure 3.1-4):

- *test* reaches MAR-E1 and MAR-E2 and *baseline* reach MAR-E3 on the MacKay River;
- *test* reach STR-E1 and *baseline* reach STR-E2 on the Steepbank River; and
- *baseline* reach FIR-E2 on the Firebag River.

Samples were collected synoptically with Neill-Hess cylinder samples (see Section 3.1.3.2). Two sets of synoptic samples were collected from *test* reaches MAR-E2 and STR-E1 and *baseline* reach STR-E2 and one set of synoptic samples was collected at *test* reach MAR-E1 and *baseline* reaches MAR-E3 and FIR-E2.

Collected samples were preserved in 10% buffered formalin and bottled for transport to the taxonomist.

A2.3.2.2 Laboratory

Samples were processed by Dr. Jack Zloty in a manner similar to that used for the Neill-Hess cylinder samples. Organisms were identified to lowest practical taxonomic level.

A2.3.2.3 Statistics

For each sample, the following benthic invertebrate community measurement endpoints were calculated:

- Abundance (total number of individuals/m^2);
- Taxon richness (number of distinct taxa);
- Simpson's Diversity Index (D), where

  $$D = 1 - \sum (p_i)^2$$

  and p_i is the proportion that taxon i contributes to the total number of invertebrates in a sample;

- Evenness, where

  $$Evenness = \frac{D}{D_{max}}$$

  $$D_{max} = 1 - \left(\frac{1}{S}\right)$$

  and S is the total number of taxa in the sample. In cases where S = 1 (i.e., only one taxon was identified in a sample), evenness was set to 1; and

- Percent EPT (Ephemeroptera, Plecoptera, Trichoptera).

Scatterplots were presented to visualize the effect of sample collection method on values of measurement endpoints for benthic invertebrate communities.

An analysis of variance (ANOVA) was used to test for a significant influence of the sample collection method (Table A.3-1). The data included in this analysis were from those reaches where two sets of synoptic samples were collected: the duplicate set provided a measure of within-reach variability for both methods of collection. A significant interaction between Reach and Method (i.e., R x M) would imply that the Neill-Hess and kick net samples produced different values of measurement endpoints and that the nature of the difference depended on the reach. The interaction term was tested first for each of the four indices. In the absence of a significant interaction, a significant difference in Method (M) would imply that the Neill-Hess and kick net samples produced different values of measurement endpoints and that the nature of the difference was common to all reaches.

Table A.3-1 Generic ANOVA table to test for an effect from collection method.

Source	df	F
Reach (R)	2	$\frac{MSR}{MSE}$
Method (M)	1	$\frac{MSM}{MSE}$
Reach x Method (R x M)	1	$\frac{MSR \times M}{MSE}$
Error (E)	8	

A2.3.3 Results

There was no significant difference in values of taxa richness between the CABIN kick net samples and the Neill-Hess cylinder samples (Table A2.3-2 and Figure A2.3-1). Generally, both types of samples collected between 20 and 50 taxa, depending on the reach. There were significant differences in diversity and evenness on the interaction term, implying that the differences in values between collection methods were depended on the reach. The Neill-Hess cylinder samples produced higher diversity and evenness in *test* reach STR-E1 and the kick sample produced higher diversity and evenness in *baseline* reach STR-E2, producing the significant interaction (Table A2.3-2). In all reaches, diversity and evenness were high using both types of sampling (> 0.8).

The most significant difference between the two sampling methods was in the values of percent EPT. The kick net samples consistently produced significantly higher percent EPT than the Neill-Hess cylinder samples across all reaches.

Kick net samples tended to produce lower relative abundance of some of the smaller organisms such as chironomids and naidid worms. For example, the Neill-Hess cylinder sample from *baseline* station FIR-E2 contained 45% chironomids while the kick net sample contained only 15% chironomids (Table 2.3-3). The Neill-Hess cylinder sample from the *test* reach MAR-E1 contained 34% naidid worms while the kick net sample contained only 6% naidid worms.

The kick net samples tended to contain larger organisms such as sphaeriid clams and gastropods. For example, the Neill-Hess cylinder sample from *baseline* reach FIR-E2 produced 1% bivalves (i.e., sphaeriid clams) while the kick sample produced 13% bivalves (Table A2.3-3). Clams were similarly more abundant in kick net samples from the *test* reach MAR-E2, *baseline* reach MAR-E3, and *baseline* reach STR-E2 compared to Neill-Hess cylinder samples.

Table A2.3-2 Results of ANOVA testing for differences on values of measurement endpoints for benthic invertebrate communities related to sampling method.

Variable	Source	SS	df	MS	F-Ratio	p-value
Richness	Station	0.031	2	0.015	3.855	0.067
	Method	0.001	1	0.001	0.262	0.623
	Error	0.032	8	0.004		
Simpsons	Station	0.048	2	0.024	28.193	0.001
	Method	0.002	1	0.002	2.253	0.184
	Station x Method	0.009	2	0.004	5.100	0.051
	Error	0.005	6	0.001		
Evenness	Station	0.046	2	0.023	23.703	0.001
	Method	0.002	1	0.002	2.396	0.173
	Station x Method	0.010	2	0.005	5.204	0.049
	Error	0.006	6	0.001		
EPT	Station	0.138	2	0.069	4.602	0.047
	Method	0.057	1	0.057	3.788	0.088
	Error	0.120	8	0.015		

Figure A2.3-1 Scatterplot of richness, diversity, evenness, and percent EPT in kick and Neill-Hess cylinder samples, 2010.

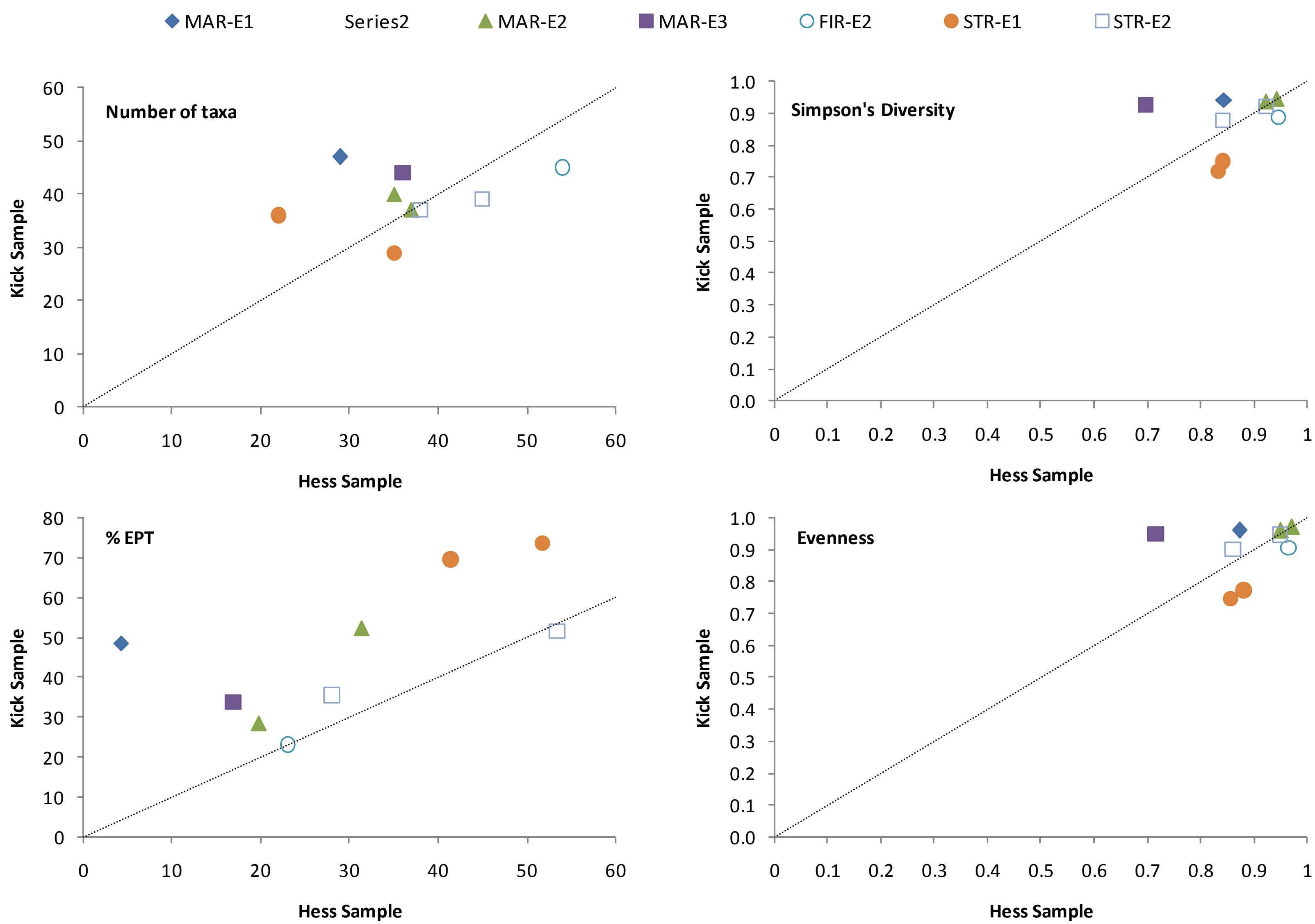


Table A2.3-3 Relative abundance of major benthic invertebrate groups in Neill-Hess cylinder and CABIN kick net samples in reaches in the RAMP FSA, 2010.

Taxon	FIR-E2-1		MAR-E1-1		MAR-E2-1		MAR-E2-10		MAR-E3-10		STR-E1-1		STR-E1-10		STR-E2-1		STR-E2-10	
	Hess	Kick	Hess	Kick	Hess	Kick	Hess	Kick	Hess	Kick	Hess	Kick	Hess	Kick	Hess	Kick	Hess	Kick
Anisoptera	<1	<1		1		<1		<1	<1	1	<1	1	1	<1		<1	<1	
Athericidae												<1	<1		<1	5		<1
Bivalvia	1	13	<1	2	2	6	4	14	<1	11					1	4	3	10
Ceratopogonidae	1		<1	1			2	3	1	1	2	2	1				1	
Chironomidae	45	15	45	17	39	11	42	20	20	17	25	7	21	8	30	31	58	45
Coleoptera	2	2		1	<1	1	<1	1		1			<1				<1	<1
Empididae	<1			2		1		2	1	1	3	6	<1	1	6	2	3	5
Enchytraeidae	1		1	<1	3	1	5	5	1		4	1	1				1	
Ephemeroptera	14	15	3	28	25	28	8	5	6	18	40	64	51	70	15	23	14	18
Gastropoda	2	10		4	1	5	1	13	1	4		1	<1		1	1	<1	2
Hydracarina	6	1	1	17	14	14	9	9	5	13	21	5	15	4	5	1	2	1
Naididae	3	2	34	6	7	3	3	1	53	16	2	7	5	2	2	4		1
Nematoda	3		2		<1	3	3		1		<1		1				3	
Ostracoda	1	5			1	1		1										1
Plecoptera	3	4	<1	6	2	9	2	5	2	6	<1	3	<1	2	1	3	4	4
Simuliidae	8	27	<1			2	1		1			1	1	2	<1		2	
Tabanidae	<1		<1		<1			<1	<1							<1		
Tipulidae	<1	2				<1								<1	<1	<1	<1	<1
Trichoptera	7	5	1	14	5	15	10	18	8	10	1	3	<1	2	37	26	11	14
Tubificidae	3	1	13	1	<1		5	1		1				4				<1
Benthic Invertebrate Community Measurement Endpoints																		
Richness	54	47	29	45	35	40	37	37	36	44	22	36	35	29	38	37	45	39
Simpson's Diversity	0.95	0.94	0.84	0.89	0.92	0.94	0.94	0.95	0.70	0.93	0.84	0.75	0.83	0.72	0.92	0.92	0.84	0.88
Evenness	0.96	0.96	0.87	0.91	0.95	0.96	0.97	0.97	0.72	0.95	0.88	0.77	0.86	0.75	0.95	0.95	0.86	0.90
% EPT	23	48	4	23	31	52	20	28	17	34	41	70	52	74	53	51	28	36

A2.3.4 Discussion and Recommendations

The CABIN protocol has become (since the inception of RAMP) an alternative methodology for collecting benthos from rivers. Similar procedures are also used in lakes (David *et al.* 1998) and have been used in "acid-sensitive" lakes in the oil sands region (Parsons *et al.* 2010). The traveling kick net method is potentially suitable for use in RAMP because the equipment is robust (necessary for the field component of RAMP) and because this type of gear can collect samples under virtually any habitat conditions with the single caveat that the sample must be collected within a wadeable environment.

Given there was uncertainty whether Neill-Hess cylinder samples could be collected in September 2010 and because there is always the possibility that high water levels observed in 2010 could happen again and inhibit sampling, it was important to quantify the degree of similarity between measurement endpoints from samples collected using the CABIN kick and sweep protocol and from samples collected using the RAMP-conventional Neill-Hess Cylinder.

The CABIN kick net samples generally produced similar number and type of taxa and values of diversity and evenness compared to the Neill-Hess cylinder samples. Kick net samples, however, tended to collect more of the larger organisms such as mayflies, stoneflies, caddisflies and clams, and fewer small organisms such as chironomids and naidid worms resulting in increased percent EPT using a kick net rather than a Neill-Hess cylinder.

The discrepancy in the size of organisms collected is partly due to the difference in mesh size between the two sampling techniques. The kick net samples were collected using a 400 µm mesh, as per the recommendations in the CABIN protocol (Reynoldson *et al.* 2004) while the Neill-Hess cylinder was built with a 220 µm mesh screen.

In future, and if any reach cannot be sampled using the conventional gear (i.e., Neill-Hess cylinder), the preliminary data and assessment in this study demonstrated that some values of measurement endpoints may be comparable between the two sampling techniques (i.e., diversity, evenness, and taxa richness). Similar observations were made by Borisko *et al.* (2007) in a comparison of rapid benthic invertebrate community collection methods in the Toronto area.

The differences observed in percent EPT between the two sampling techniques (i.e., the kick net samples produced significantly higher percent EPT [~ 40% higher] than those produced by the Neill-Hess cylinder would need to be accounted for if kick net sampling was used in future sampling events. Regardless of the difference in percent EPT between the two sampling techniques, the CABIN protocol provides a reliable alternative to the Neill-Hess cylinder method that should be employed during periods of high water levels to ensure that a benthic sample is collected every year.

A2.4 PRELIMINARY ASSESSMENT OF EFFECT OF CLIMATE VARIABLES ON BENTHIC INVERTEBRATE COMMUNITY MEASUREMENT ENDPOINTS

A2.4.1 Introduction

The RAMP Benthic Invertebrate Communities component has focused on lower reaches of major tributaries to assess the effects of focal projects on benthic invertebrate communities. The lower reaches of the major tributaries are anticipated to be the most likely to respond to oil sands developments because they are at the bottom of watersheds where oil sands developments are active. In addition, the tributaries are more likely than the mainstem Athabasca River to respond to any influence from oil sands developments for at least two reasons. First, the mainstem presents a shifting sand environment that generally contains more tolerant benthic taxa than does non-shifting sands, or gravel/cobble. Second, the mainstem carries a lot of water that will dilute inputs and other stressors associated with oil sands operations. In the context of regional conclusions, if there are no effects in the areas we most expect to see them, then it is unlikely that there will be large-scale effects in a regional context.

The assessment of the condition of benthic invertebrate communities of lower reaches of major tributaries is tiered as follows:

1. An evaluation of trends over time in the lower reaches. Time trends in lower reaches are compared with time trends in *baseline* reaches found in the same watershed and upstream of oil sands developments. The assessment of time trends typically involves the use of analysis of variance, judging the significance of the observed differences relative to the variations within time periods and within reaches.

2. Where and when a lower reach is demonstrated to have produced a significant change that is consistent with an oil sands developments, variations within that lower reach are then judged relative to a range of natural variability in measurement endpoints for benthic invertebrate communities in reaches of similar habitat type, for example, an erosional reach that produces a significant difference compared to an upstream *baseline* erosional reach is then compared to the variation among other *baseline* erosional reaches.

To determine the range of variation in erosional and depositional *baseline* reaches, the lower 5^{th} and upper 95^{th} percentiles were calculated from data for all years and all reaches that, are or have been, classified as *baseline*. This lumping of reaches for the purpose of generating a range of *baseline* variation assumes generally that the composition of the benthic invertebrate communities is going to be broadly similar and that the natural influence of factors such as geology, slope, and discharge are minimal (Imhoff *et al.* 1996, Stanfield and Kilgour 2006). In the event that these influences were not minimal, there is a concern that this approach has the potential to mask effects of oil sands developments by not taking into account other natural causes of variation. Further, it is difficult to understand to the extent possible the periodic effects that are consistent with oil sands developments but that might otherwise be caused by natural variations in climatic factors.

Some potential natural causes of variation have been previously explored by RAMP (2007, Appendix E) for factors such as bankfull river width, substrate texture, chlorophyll *a* densities (for erosional reaches), etc., which explained only <5% of the variation in measurement endpoints for benthic invertebrate communities. Therefore, those variables were not used to modify the range of variation for *baseline* reaches because it would result in a trivial reduction in the size of the range of variation while complicating the overall approach to analysis.

In the recent peer review of RAMP (AITF 2011), there were some concerns raised regarding the size of the range of variability for *baseline* reaches for measurement endpoints and that some of the variation may be related to climatic variables. In the RAMP 2008 and 2009 Technical Reports, the results highlighted cyclic variations in taxa richness and percent of the fauna as EPT taxa in both *baseline* and *test* river reaches (RAMP 2009a, 2010), but did not examine the association between those variations and other regional climatic variability. A good example of the large variations that have occurred in the RAMP FSA is percent EPT in the lower Steepbank River. This reach was first sampled in RAMP in 1998 when less than 50% of the fauna consisted of mayflies, stoneflies and caddisflies (i.e., EPT taxa). Since 1998, the percent of the fauna as EPT has decreased over time relatively consistently until 2008 when approximately 15% of the fauna was species of mayflies, stoneflies and caddisflies (RAMP 2009a). In 2009 and 2010, the percent of the fauna as EPT increased to upwards of 40 to 50% of the total fauna, similar to proportions when RAMP first sampled the lower Steepbank River. To date, there have been no explanations of the causes of those variations with the exception that there were some years when the trend in percent EPT was considered to be potentially due to oil sands developments (i.e., a decrease in organisms that are more sensitive to changes in their environment).

The objective of this study is to provide a preliminary analysis of the potential influences of climatic variables on variations in measurement endpoints for benthic invertebrate communities. Variables that are considered include mean air temperature during the open-water period and annual average discharge. Mean air temperature during the open-water period is considered the most likely to influence the hatching times and frequencies of insects such as chironomids, mayflies, stoneflies, and caddisflies, which are groups that are dominant in the benthic invertebrate communities of the Athabasca River and its tributaries. Mean annual discharge is considered likely to be related to discharge events that influence the benthic invertebrate communities. This study is not exhaustive in all variables that can be explored; however, it is expected that other scientists will continue to explore discharge and climatic variables and their influence on benthic invertebrate communities.

A2.4.2 Methods

A2.4.2.1 Data Collection

The data used in this assessment were from depositional and erosional river reaches in the RAMP FSA. The following measurement endpoints were calculated using data from 1998 to 2010:

- Abundance (total number of individuals/m^2);
- Taxon richness (number of distinct taxa);
- Simpson’s Diversity Index;
- Evenness; and
- Percent EPT.

Average measurement endpoint values were calculated for each reach-year combination based on methods described in Section 3.2.3.1.

A2.4.2.2 Air Temperature

Hourly air temperature data were obtained from the weather station 719320 (CYMM) located in Fort McMurray (latitude 56.65N, longitude -111.21W, altitude 369m). Average air temperature for the open-water period between May and October of each year was used from the available data records.

A2.4.2.3 Discharge

Discharge data from the RAMP Climate and Hydrology Component database was used for locations provided in Table A2.4-1. Data were acquired from stations located furthest downstream on each river. Mean annual discharge was calculated for each river-year combination.

Table A2.4-1 Location and data from hydrology stations that were used in the study of the influence of discharge on measurement endpoints for benthic invertebrate communities.

Station Location	Station Name	Year
Athabasca River	S24	2001 to 2009
Beaver River	S39	2008, 2009
Calumet River	CR1/S16	CR1 (2001 to 2004)/S16 (2005 to 2009)
Christina River	S29	2002 to 2009
Clearwater River	S42	2009
Ells River	S14/S14A	S14 (2001 to 2007)/S14A (2008 to 2009)
Firebag River	S27	2002 to 2009
Fort Creek)	S12	2000 to 2009
Hangingstone River	S31	2002 to 2009
Jackpine Creek	S2	1998 to 2009
MacKay River	S26	2001 to 2009
Muskeg River	S7	1998 to 2009
Poplar Creek	S11	1998 to 2009
Steepbank River	S38	2009
Tar River	S15/S15A	S15 (2001 to 2006)/S15A (2007 to 2009)

A2.4.2.4 Data Analysis

Analysis of covariance (ANCOVA) was used to test for differences in the relationship between measurement endpoints for benthic invertebrate communities and mean air temperature and mean annual discharge. In the full model of the ANCOVA, the predictors of benthic invertebrate community included Reach, Air Temperature (or Discharge), and the interaction term Reach x Air Temperature (or Reach x Discharge). The interaction term tested whether there were significantly different slopes of the relationship between the measurement endpoints and the climate variable (i.e., temperature and discharge) among reaches. The difference in slopes would imply that the influence of climate differed significantly among reaches. The objective was to determine if the influence of the climate variables was approximately similar among reaches, thus demonstrating a strong influence of climate in a regional context. The magnitude of the influence of air temperature and discharge was quantified using percent of variance explained.

A2.4.3 Results and Discussion

Variations in measurement endpoints from depositional reaches were not related to variations in discharge or mean air temperature in the open-water period (Table A2.4-2, Figure A2.4-1, and Figure A2.4-2). However, total abundance in erosional reaches decreased with increasing mean annual discharge ($p<0.001$) (Figure A2.4-3). The relationship between discharge and total abundance explained 15% of the variation in total abundance. In some rivers, taxa richness and percent EPT in erosional reaches increased with increasing mean air temperature in the open-water period ($p=0.015$ and $p=0.021$, respectively) with temperature explaining 5% of the variation in taxa richness and percent EPT (Table A2.4-2, Figure A2.4-4).

Table A2.4-2 Results of analysis of covariance testing for the influence of discharge and mean air temperature on variations in measurement endpoints for benthic invertebrate communities.

Climate Variable	Measurement Endpoint	Erosional		Depositional	
		p-value	R^2	p-value	R^2
Discharge	Log Abundance	**<0.001**	**0.15**	0.428	0.01
	Log Richness	0.08	0.04	0.112	0.02
	Simpson's Diversity	0.87	0.00	0.236	0.01
	Evenness	0.73	0.00	0.379	0.01
	Log %EPT	0.01	0.00	0.685	0.00
Air Temperature	Log Abundance	0.445	0.01	0.790	0.00
	Log Richness	**0.015**	**0.05**	0.160	0.02
	Simpson's Diversity	0.227	0.02	0.934	0.00
	Evenness	0.094	0.04	0.670	0.00
	Log %EPT	**0.021**	**0.05**	0.101	0.02

Figure A2.4-1 Scatterplot of measurement endpoints for benthic invertebrate communities in depositional reaches in relation to the mean annual discharge.

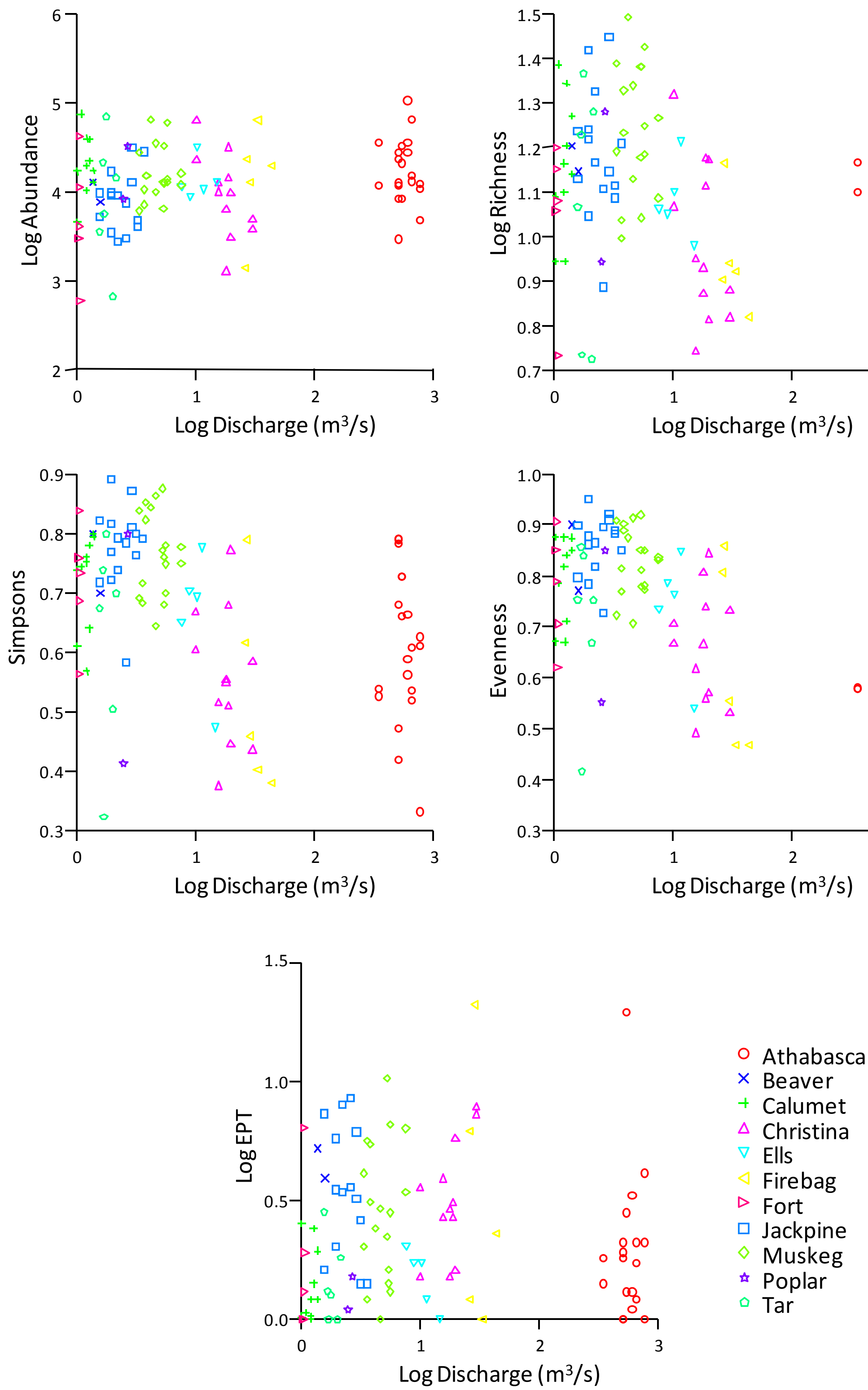


Figure A2.4-2 Scatterplot of measurement endpoints for benthic invertebrate communities in depositional reaches in relation to mean air temperature during the open-water period.

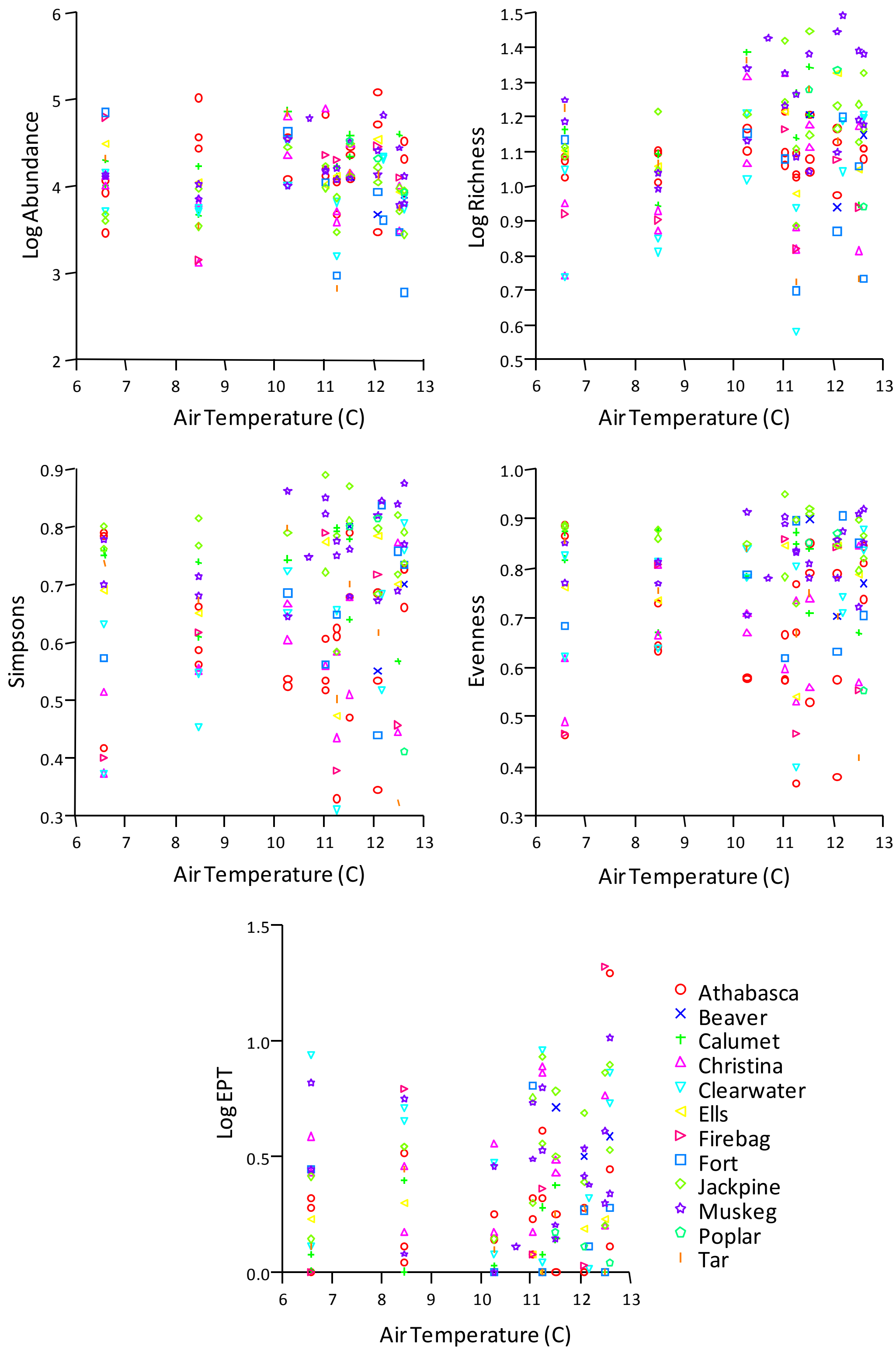


Figure A2.4-3 Scatterplot of measurement endpoints for benthic invertebrate communities in erosional reaches in relation to the mean annual discharge.

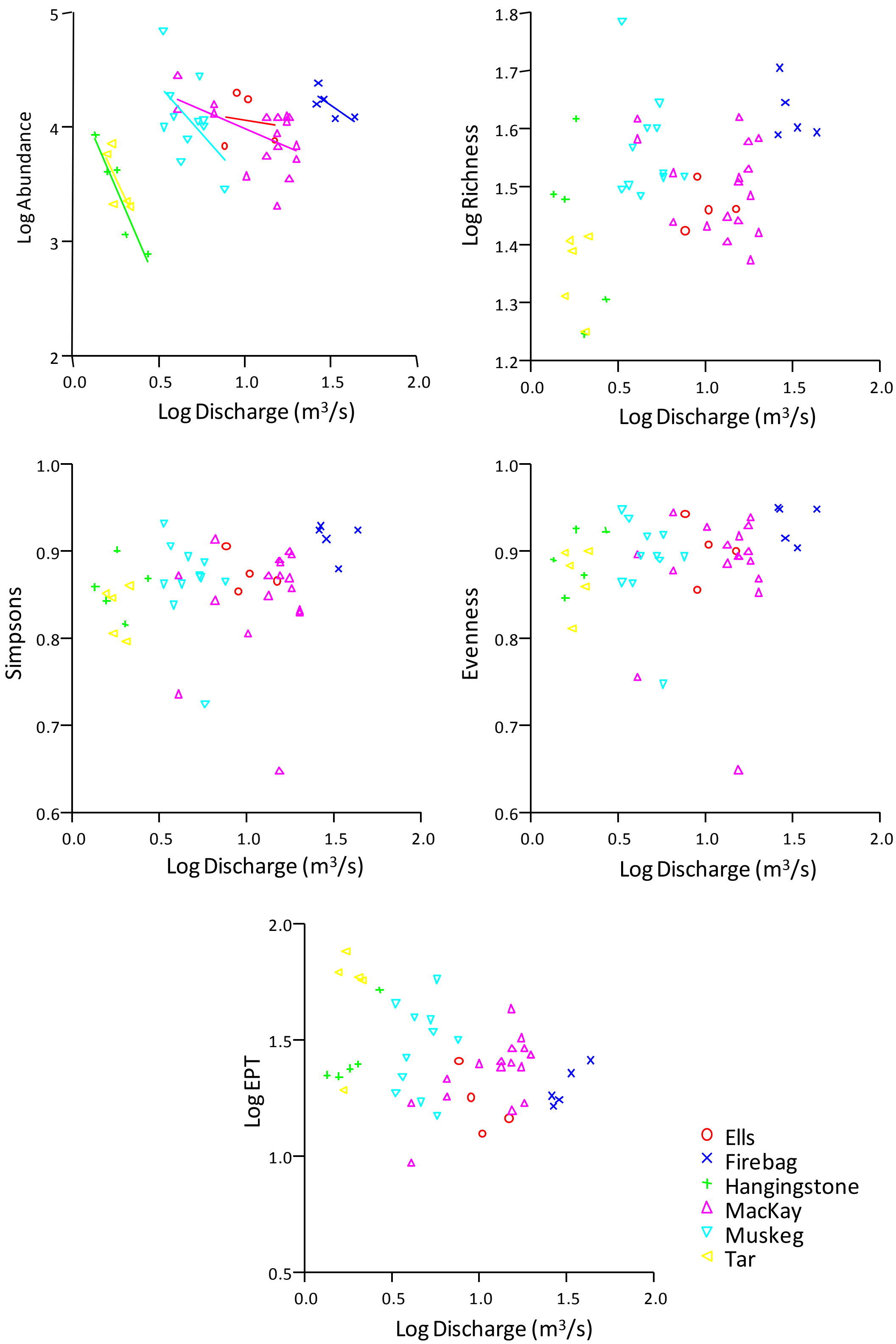
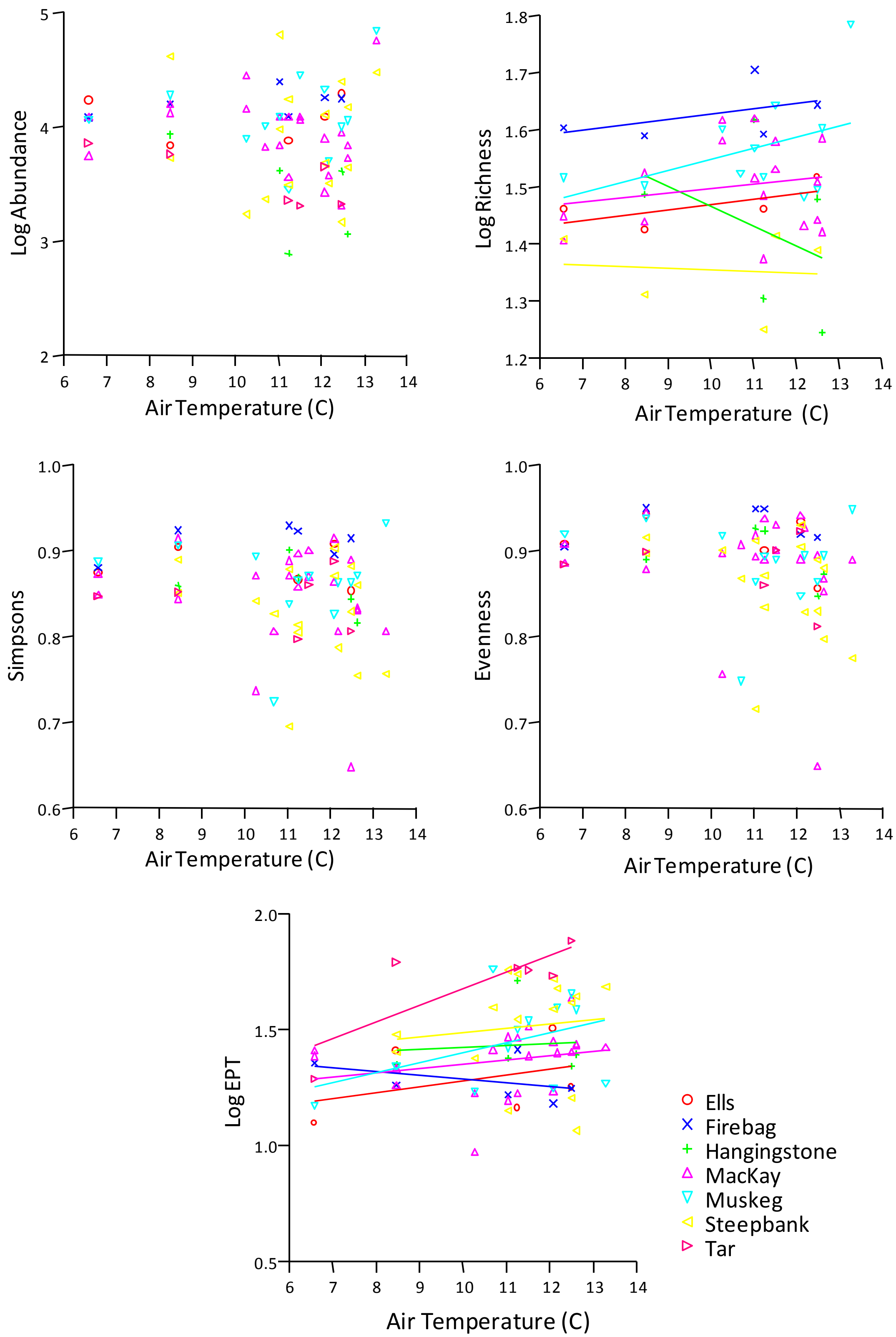


Figure A2.4-4 Scatterplot of measurement endpoints for benthic invertebrate communities in erosional reaches in relation to mean air temperature during the open-water period.



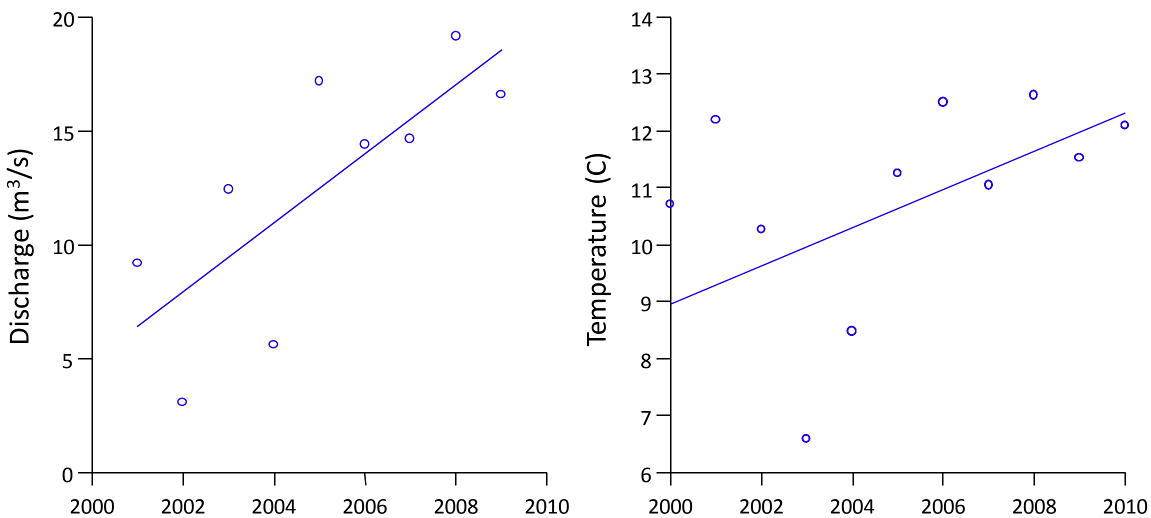
A2.4.4 Discussion and Recommendations

The measurement endpoints for benthic invertebrate communities in depositional reaches do not appear to be influenced by climatic variability on a regional scale. The *baseline* range of variation, as calculated by RAMP can; therefore, be considered to be broadly applicable on a regional scale for depositional reaches.

The influence of mean annual discharge and the influence of mean air temperature during the open-water period should be taken into account for the baseline range of variation for taxa richness in erosional reaches. In addition, time trends observed for total abundance, taxa richness, or percent EPT should also consider the potential influence of variations in discharge (climate or operations related) and mean temperature during the open-water period.

Of the erosional *test* reaches sampled by RAMP, there were no significant time trends in abundance, taxa richness or percent EPT in the lower Muskeg River (*test* reach MUR-E1, see Section 5.2) or lower Steepbank River (*test* reach STR-E1, see Section 5.3). There was, however, a significant decreasing trend over time in abundance and an increasing trend over time in taxa richness and percent EPT in the lower MacKay River (*test* reach MAR-E1). These differences did not imply a change associated with oil sands developments because of the nature of the change was not negative. However the trends were consistent with increasing discharge and air temperature over time (Figure A2.4-5).

Figure A2.4-5 Scatterplot of mean annual discharge in the MacKay River and mean air temperature.



A2.5 FISH ASSEMBLAGE MONITORING PILOT STUDY

A2.5.1 Introduction

In an effort to harmonize the monitoring activities under RAMP, a fish assemblage monitoring (FAM) pilot study was initiated in 2009 and continued in 2010 at stations/reaches where the Water Quality, and Benthic Invertebrate Communities and Sediment Quality components conduct sampling. The objective of the fish assemblage monitoring program is to assess the health of fish populations in tributaries that are potentially influenced by oil sands activities similarly to monitoring objectives of other components in RAMP.

A2.5.1.1 Study Design Considerations

In 2009, sampling of fish assemblages was conducted at 11 locations including the Beaver, Dunkirk, Horse, MacKay, Muskeg, Steepbank and Tar rivers and Jackpine and Poplar creeks (RAMP 2010). The 2009 analysis was primarily designed to evaluate the ability to assess fish assemblage metrics between *test* and *baseline* reaches following methods developed by the United States Environmental Protection Agency (EPA) Environmental Monitoring and Assessment Program (EMAP) for stream monitoring programs throughout the United States (Peck *et al.* 2006). The analyses also examined variations in measures of community composition including an Assemblage Tolerance Index (ATI) (Whittier *et al.* 2007a) multivariate ordination axis scores, and a modified Index of Biotic Integrity (IBI) (Karr 1981).

Given the limited number of fish species in the lower Athabasca region and the low abundance in small tributaries to the Athabasca River, the 2009 survey did not produce adequate sample sizes to compare the metrics established by the USEPA protocols. The USEPA protocols indicate that with adequate fishing effort, 30 times as many individual fish as the expected number of species should be captured to reduce the effect of rare species on the Index of Biotic Integrity (IBI) and other metric scores (Hughes and Peck 2008, Kanno *et al.* 2009, Dußling *et al.* 2004). These protocols were derived from USEPA fish assemblage studies in the northwestern US, where fish assemblages can contain upwards of 60 to 70 fish species per reach with a high abundance of each individual species. The rivers in this region are generally more productive because of higher water temperatures and higher nutrient loads.

The recommended reach length to be surveyed should be at least 40 times the wetted width with a minimum length of 150 m (Peck *et al.* 2006) and if that distance were to not yield the expected number of individuals, the distance should be increased or the fishing effort increased within a reach. The rationale supporting the requirement that the minimum length of a reach should be 150 m (or 40 x the wetted width) or that the total catch of individual fish should exceed 30 times the number of species is based on a presumed desire to document 95% of the species available in a river reach.

Taking into account this expected level of effort and assuming that there are 12 common fish species in the RAMP FSA (RAMP 2009b) that are expected at almost any reach, an adequate tributary sample should contain a minimum of 360 individuals; an adequate large river sample, with presence of large-bodied species, which are less frequent in smaller tributaries, should contain a minimum 480 individuals. In many of the smaller tributaries in the RAMP FSA, to achieve these sample sizes, the level of effort that would be required could not be completed in one day of sampling. Therefore, for the 2010 fish assemblage monitoring program, an alternative method was derived to determine the adequate level of fishing effort required to characterize the fish assemblage in a river reach.

A2.5.1.2 Design and Objectives of the 2010 FAM Program

The recommended method to estimate sampling effort for benthic invertebrate community surveys under the Canadian Environmental Effects Monitoring (EEM) programs for metal mines and pulpmills has a fundamentally different approach, based on signal-to-noise ratios, and the desire to statistically "detect" differences in composition between reaches (Environment Canada 2010). For benthic invertebrate community surveys, individual samples (e.g., Neill-Hess cylinders, Ekman grabs, Ponar grabs, etc.) are collected within stations or sub-reach, with stations/sub-reaches considered to be a random sample, and the unit of replication. The variation among stations/sub-reaches (or replications) is then used to judge the significance of variations between or among reaches (i.e., *test* vs. *baseline*). Within the EEM program, it is considered important to know that the variation in estimates of measurement endpoint values is measured with minimal variance. Therefore, the EEM program dictate a level of sampling effort that would ensure that measurements endpoints within a station/sub-reach are measured to within ±20% of the true (but yet unknown) average value. The technical guidance document for the EEM program recommend that pilot studies be carried out to determine the required sampling effort to ensure that the within-reach variance of measurement endpoints is within ±20% of the true (but yet unknown) average value (Environment Canada 2010).

The 2010 RAMP fish assemblage monitoring survey was designed taking into account the EEM recommendation for pilot studies with the objective of determining the level of effort that would be required in order to estimate conventional (and ecologically fundamental) measurement endpoints for fish assemblages. The objective of this pilot study was to determine the number of sub-reaches that would be required in order to produce estimates of measurement endpoints that were within some acceptable level of precision (i.e., ± 20% of the average sub-reach value).

A secondary objective of this pilot study was to document differences in measurement endpoints of fish assemblages in reaches that have been sampled two years in a row (2009 and 2010) including *test* reach JAC-F1, *test* reach MUR-F1, and *test* reach STR-F1.

A2.5.2 Field Methods

A2.5.2.1 Fish Sampling and Handling

The methods used to develop a FAM pilot study for RAMP were adopted from the United States Environmental Protection Agency (USEPA) Environmental Monitoring and Assessment Program (EMAP) for stream monitoring programs throughout the United States (Peck *et al.* 2006). The procedures described were modified to include appropriate indicators related to the RAMP FSA. The EMAP protocols outline the collection of physical habitat, fish, water and sediment chemistry, and benthic invertebrate variables.

The FAM pilot study was conducted from September 14 to September 19, 2010 to assess changes in the fish assemblage of rivers that may potentially be related to focal projects. The study included sampling at six reaches on tributaries of the Athabasca River within the RAMP FSA where Water Quality, Benthic Invertebrate Communities and Sediment Quality components conducted sampling in 2010 (Figure A2.5-1 and Table A2.5-1). Four of these reaches are designated as *test*: the lower Steepbank (STR-F1), lower Muskeg (MUR-F1), lower Ells River (ELR-F1) and lower Jackpine Creek (JAC-F1), while the remaining reaches are designated as *baseline*: the upper Ells River (ELR-F2) and upper Jackpine Creek (JAC-F2) (Table A2.5-1). Two of the reaches were in depositional habitat and four were in erosional habitat. Average wetted widths of reaches ranged from 6 to 28 m, with two reaches ≤ 10 m. The depositional reaches were all <u><</u>10 m wide and the erosional reaches were all <u>></u>20 m wide. The FAM pilot study included reaches of varying stream order and size, upstream and downstream of focal projects, across representative set of watercourses in the RAMP FSA.

Five of the six reaches were separated into 10 sub-reaches to assess variance and error to determine the number of sub-reaches required to capture the expected fish assemblage within a reach. *Baseline* reach JAC-D2 was not separated into 10 sub-reaches because the depth made it difficult to wade continuously through the entire reach. The wadeable, near-shore area of each sub-reach was electrofished with intensities that varied between 4 and 19 seconds per lineal meter. The catch per sub-reach was standardized by the length of the sub-reach. The width of the electrofishing pass was approximately 2 to 3 m, or from the river bank to a point mid-river based on what the electrofisher operator could reach. Fish from each sub-reach were kept in a holding bucket until the completion of all fishing. For each sub-reach, captured fish were measured for length (± 0.01 mm) and weight (± 0.01 g) and an external assessment was conducted for general health.

Table A2.5-1 Location and designation of fish assemblage monitoring reaches, 2010.

Watershed	Reach	Habitat Type	Reach Designation	Water Quality Station/Benthic Invertebrate Reach	Effort (sec)	Reach Length (m)	Average Wetted Width (m)
Jackpine Creek	JAC-F1	despositional	*test*	JAC-1/JAC-D1	3,863	300	6
Jackpine Creek	JAC-F2*	despositional	*baseline*	JAC-2/JAC-D2	5,161	502	10
Ells River	ELR-F1	erosional	*test*	ELR-1/ELR-D1	4,694	500	23
Ells River	ELR-F2	erosional	*baseline*	ELR-2a/ELR-E2	3,959	500	25
Muskeg River	MUR-F1	erosional	*test*	MUR-1/MUR-E1	2,491	500	28
Steepbank River	STR-F1	erosional	*test*	STR-1/STR-E1	4,997	500	20

* Reach was not separated into subreaches.

Figure A2.5-1 Locations of fish assemblage monitoring reaches in the RAMP Focus Study Area, as part of a two year pilot study, 2009 to 2010.

V:\Data\Projects\RAMP1565 (2010)\GIS4Review\H_TechRpt\RAMP1565_P1_FishAssemblage_20110218.pdf

letter

A2.5.2.2 Fish Habitat Assessments

Habitat assessments were completed at three transects at the downstream, upstream and midpoints of each reach. Habitat assessment methods involved recording a range of variables relating to channel morphology, substrate, water quality, and stream cover similar to that outlined in RAMP (2009b) and Peck *et al.* (2006). The following information was collected at each transect:

- Habitat type (Table A2.5-2);
- Wetted width (m);
- Maximum depth (m);
- Velocity and depth (m/sec) (at 25%, 50%, and 75% of the wetted width);
- Overhead and instream cover (%) (Table A2.5-3);
- Substrate (dominant and subdominant particle size) (Table A2.5-4);
- Bank slope (°);
- Bank height (m); and
- Large and small woody debris (count of debris in length/size classes).

In situ water quality variables including temperature, dissolved oxygen, and conductivity were measured using a Hanna hand-held probe (temperature, conductivity, pH) and a LaMotte Winkler titration kit (dissolved oxygen) and collected at the upstream, middle, and downstream transects of each reach.

Table A2.5-2 Habitat type and code for the fish assemblage monitoring pilot study (adapted from Peck *et al.* 2006).

Habitat Type (code)	Description
Plunge pool (PP)	Pool at base of plunging cascade or falls
Trench pool (PT)	Pool-like trench in the centre of the stream
Lateral Scour Pool (PL)	Pool scoured along a bank
Backwater Pool (PB)	Pool separated from main flow off the side of the channel (large enough to offer refuge to small fishes). Includes sloughs (backwater with vegetation), and alcoves (a deeper area off a wide and shallow main channel).
Impoundment Pool (PD)	Pool formed by impoundment above dam or constriction
Pool (P)	Pool (unspecified type)
Glide (GL)	Water moving slowly, with a smooth, unbroken surface. Low turbulence.
Riffle (RI)	Water moving, with small ripples, waves and eddies-waves not broken, surface tension not broken. Sound: babbling, gurgling
Rapid (RA)	Water movement rapid and turbulent, surface with intermittent white water with breaking waves. Sound: continuous rushing, but not as loud as cascade.
Cascade (CA)	Water movement rapid and very turbulent over steep channel bottom. Much of the water surface is broken in short, irregular plunges, mostly whitewater. Sound: roaring.
Falls (FA)	Free falling water over a vertical or near vertical drop into plunge, water turbulent and white over high falls. Sound: splash to roar.
Dry Channel (DR)	No water in the channel or flow is submerged under the substrate.

Table A2.5-3 Percent cover rating for instream and overhead cover at each transect for the fish assemblage monitoring pilot study (adapted from Peck *et al.* 2006).

Code	Percent Cover
0	absent, zero cover
1	sparse, <10%
2	moderate, 10-40%
3	heavy, 40-75%
4	very heavy, >75%

Table A2.5-4 Substrate size class codes for the fish assemblage monitoring pilot study (adapted from Peck *et al.* 2006).

Code	Description
RS	bedrock (smooth) - larger than a car
RR	bedrock (rough) - larger than a car
RC	asphalt/concrete
XB	large boulder (1000-4000 mm) - metre stick to a car
SB	small boulder (250-1000 mm) - basketball to a metre stick
CB	cobble (64-250 mm) - tennis ball to basketball
GC	coarse gravel (16-64 mm) - marble to tennis ball
GF	fine gravel (2-16 mm) - ladybug to marble
SA	sand (0.06 to 2 mm) - gritty, up to ladybug size
FN	silt/clay/muck - not gritty
HP	hardpan - firm consolidated fine substrate
WD	wood - any size

A2.5.3 Analytical Approach

A2.5.3.1 Measurement Endpoints

Several conventional measurement endpoints of fish assemblage composition were calculated using the fish data:

- Total Abundance – the total number of fish caught in the reach, divided by the lineal length of the reach (# of fish/m);
- Richness (S) – the total number of fish species collected per reach. Higher richness values are typically used to infer a "healthier" fish assemblage;
- Diversity – this metric was computed for each reach following the calculation for Simpson's Diversity (D), calculated as:

$$D = 1 - \sum (p_i)^2$$, where

p_i is the proportion of the total abundance accounted for by species i

Higher diversity values are typically used to infer a "healthier" fish assemblage;

- Evenness – this metric was computed for each reach following the calculation for evenness (E) as per the EEM guidance documents (Environment Canada 2005), calculated as:

$$E = \frac{1}{\frac{\sum (p_i)^2}{S}}$$

With this index, lower values imply that the fish assemblage is more evenly distributed and healthier, and not dominated by one or a few species; and

- Assemblage Tolerance Index (ATI) - The Assemblage Tolerance Index was developed by Whittier *et al.* (2007a) for stream and river fish assemblages in the western United States to quantify a species' tolerance to an overall human disturbance gradient (Table A2.5-5). For species captured in the RAMP FSA but not assessed by Whittier *et al.* (2007a), a number was assigned based on species similarity to those with calculated values, as per RAMP (2010). With this index, lower tolerance values imply a species that is more sensitive to disturbance.

Table A2.5-5 Tolerance values for fish collected during the 2009 to 2010 fish assemblage monitoring surveys (adapted from Whittier *et al.* 2007a).

Common Name	Species Code	Tolerance Value
Arctic grayling	ARGR	2.0
brook stickleback*	BRST	9.4
burbot	BURB	2.0[1]
finescale dace*	FNDC	7.0
fathead minnow*	FTMN	8.3
lake chub*	LKCH	5.5
longnose dace*	LNDC	6.2
longnose sucker*	LNSC	4.6
northern redbelly dace*	NRDC	7.0[1]
northern pike	NRPK	7.8
pearl dace*	PRDC	6.7
slimy sculpin*	SLSC	3.0[1]
spoonhead sculpin	SPSC	3.0[1]
spottail shiner*	SPSH	7.7
trout-perch*	TRPR	8.4
walleye	WALL	8.7
white sucker*	WHSC	7.6
yellow perch	YLPR	7.4

* Commonly caught fish species of Athabasca River tributaries in the Alberta oil sands region.

[1] Judgment-based score from value for similar species.

A2.5.3.2 Precision

The number of sub-reaches required to obtain estimates of measurement endpoints that are within ± 20% of the reach mean was calculated as (from Elliott 1977):

$$n = \frac{s^2}{D^2 \overline{X}^2}$$

Where,

s is the within-reach standard deviation;

$\overline{X}$ is the reach-average index value; and

D is the proposed required precision, here 20% (or D=0.2).

A2.5.4 Results

A2.5.4.1 Fish Count and Species Composition

Table A2.5-1 provides a summary of the length and width of a watercourse that was sampled at each reach. A total of 12 fish species were collected during the FAM pilot study in 2010, compared to 16 species captured in 2009, although more reaches were sampled in 2009 than in 2010. Fish species richness per reach ranged from five (*baseline* reach JAC-F2) to ten (*test* reach MUR-F1) and number of individuals captured ranged from 64 (*baseline* reach JAC-F2) to 317 (*baseline* reach ELR-F2) (Table A2.5-6). An unknown sucker species was collected at *test* reach ELR-F1 and was; therefore, not included in the total species count. There was no clear pattern in the number of fish captured and number of species between *test* and *baseline* reaches but there was generally higher number of fish captured at erosional reaches compared to depositional reaches.

Table A2.5-6 Number of fish captured by species at each reach for the FAM pilot study, 2010.

Species	Reach					
	JAC-F1	**JAC-F2**	**ELR-F1**	**ELR-F2**	**MUR-F1**	**STR-F1**
	Depositional	Depositional	Depositional	Erosional	Erosional	Erosional
brook stickleback	19	32	-	-	6	-
finescale dace	75	12	36	161	26	8
lake chub	-	10	-	-	8	-
longnose dace	-	-	4	11	20	63
longnose sucker	3	-	-	13	10	-
northern pike	1	-	-	-	-	-
pearl dace	21	9	49	82	58	64
slimy sculpin	23	-	-	-	19	60
spoonhead sculpin	-	-	-	-	4	3
trout-perch	9	-	1	4	-	7
white sucker	16	1	13	46	5	4
yellow perch	-	-	15	-	1	1
unknown sucker	-	-	1*	-	-	-
Total Fish Captured	**167**	**64**	**118**	**317**	**157**	**210**
Total No. Species	**8**	**5**	**6**	**6**	**10**	**8**

* not included in total species count

A2.5.4.2 Temporal Trends

For *test* reach JAC-F1, *test* reach MUR-F1 and *test* reach STR-F1, where sampling was conducted in 2009 and 2010, differences in values of measurement endpoints are presented in Figure A2.5-2 to Figure A2.5-4. Annual within-reach variations of measurement endpoints for fish assemblages in Jackpine Creek, Muskeg River and Steepbank River could not be determined because reaches were not divided into sub-reaches in 2009. The number of fish captured was generally higher in 2010 than 2009 in all three reaches. In 2010, the abundance varied between 0.2 fish per metre in *test* reach MUR-F1 to 0.4 fish per metre in *test* reach JAC-F1 and *test* reach STR-F1 (Figure A2.5-4).

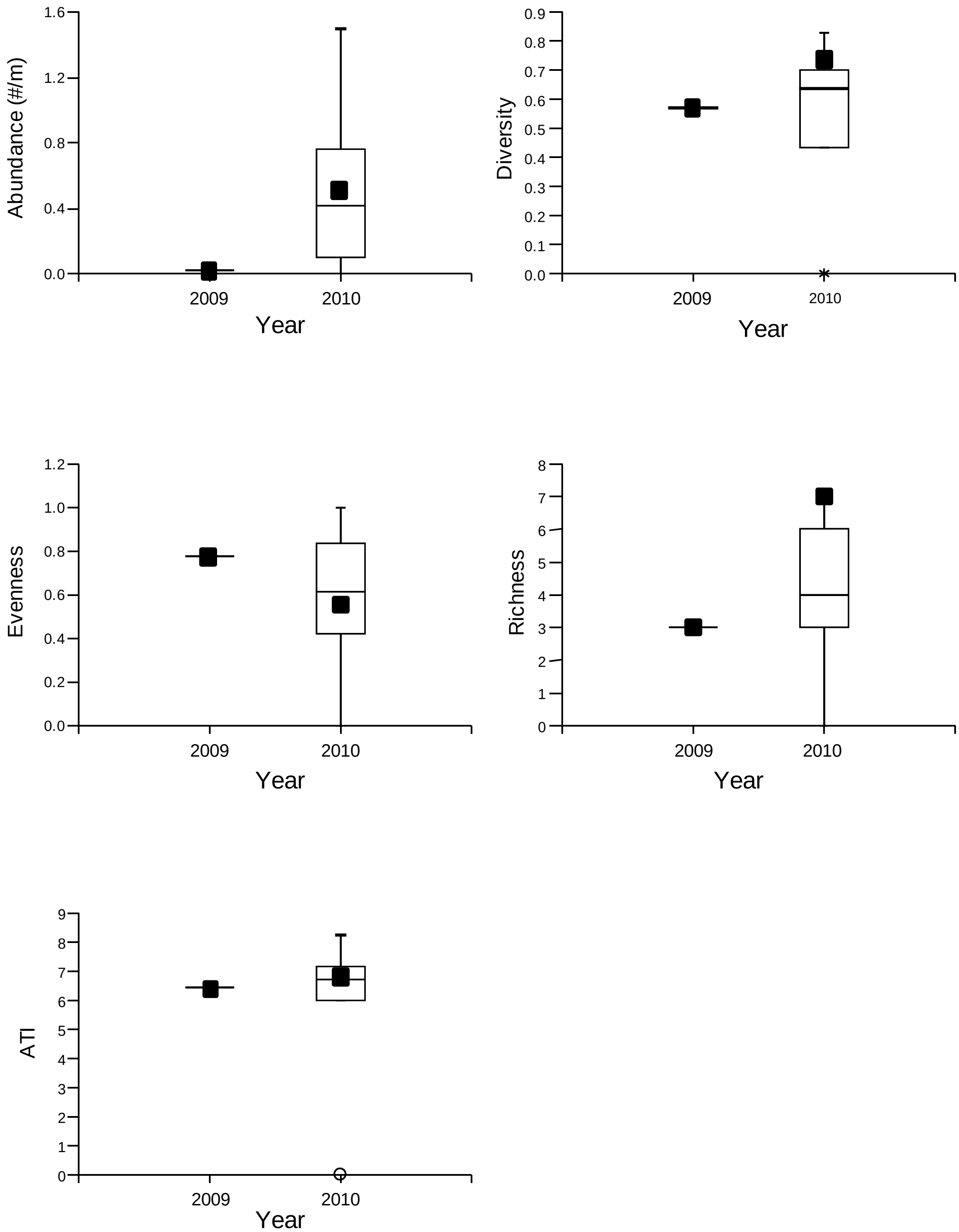
Test reach MUR-F1 produced higher ATI values in 2010 (>6) than in 2009 (<4) implying that the fish assemblage was dominated by more tolerant species in 2010. The Muskeg River fish assemblage was dominated numerically in 2009 by slimy sculpin (74% of the total catch), a species that is considered quite sensitive (ATI value of 3); however, the relative abundance of slimy sculpin was reduced to approximately 12% in 2010 (Table A2.5-7). Pearl dace, with an ATI of 6.7, was not found in the Muskeg River in 2009, and was the most numerically dominant species (38% of the total catch) in 2010. The reduced relative abundance of slimy sculpin and increased relative abundance of pearl dace in 2010 contributed to the higher ATI value in 2010. Continued monitoring at *test* reach MUR-F1 will confirm the stability of the temporal change in the fish assemblage.

Test reach STR-F1 produced an ATI of approximately 6.5 in 2009, while values varied between about 4.5 and 6.5 between sub-reaches in 2010. The Steepbank River fish community was dominated numerically in 2009 by northern redbelly dace (48%), a species which was absent from the catch in 2010. Longsnose dace, pearl dace and slimy sculpin were sub-dominant numerically in the Steepbank River fish assemblage in 2010. The reduction in the ATI value from 2009 to 2010 was partially due to the loss of northern redbelly dace (ATI value of 7) in 2010, and the increase in relative abundance of slimy sculpin (ATI value of 3).

A2.5.4.3 Spatial Comparisons for Within-Reach Variation

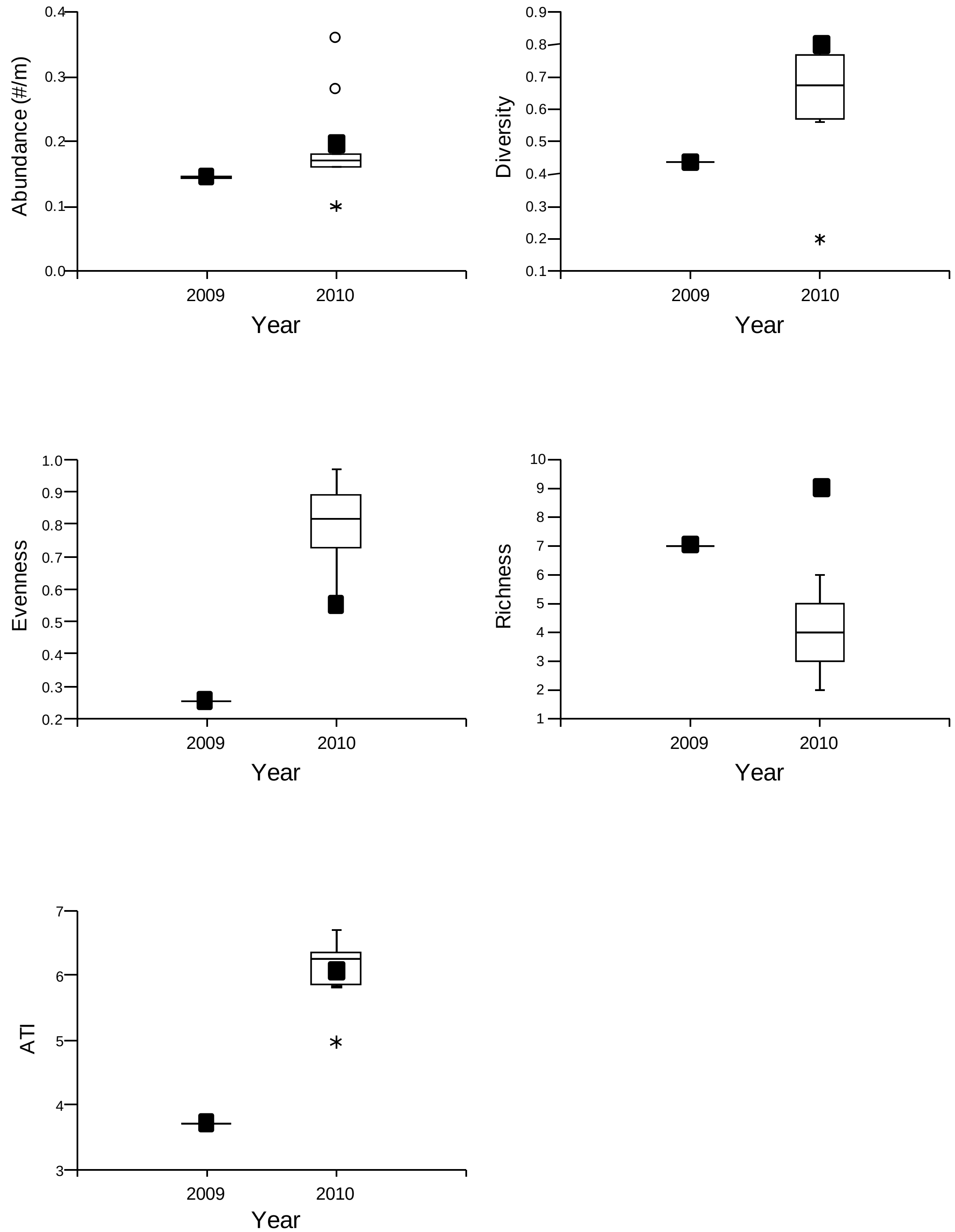
The variation in measurement endpoints across sub-reaches within each reach are provided in Table A2.5-8. Total abundance was the most variable measurement endpoint (within a reach) in 2010; the coefficient of variation (CV) for catch per metre varied between 42 and 99% with greater variability in depositional reaches than erosional reaches (Table A2.5-8). The Assemblage Tolerance Index (ATI) was the most precise measurement endpoint with the coefficient of variation between 3 and 54% with no clear pattern in variability between erosional and depositional reaches (Table A2.5-8).

Figure A2.5-2 Within-reach variation in values of measurement endpoints for fish assemblages in *test* reach JAC-F1, 2009 and 2010.



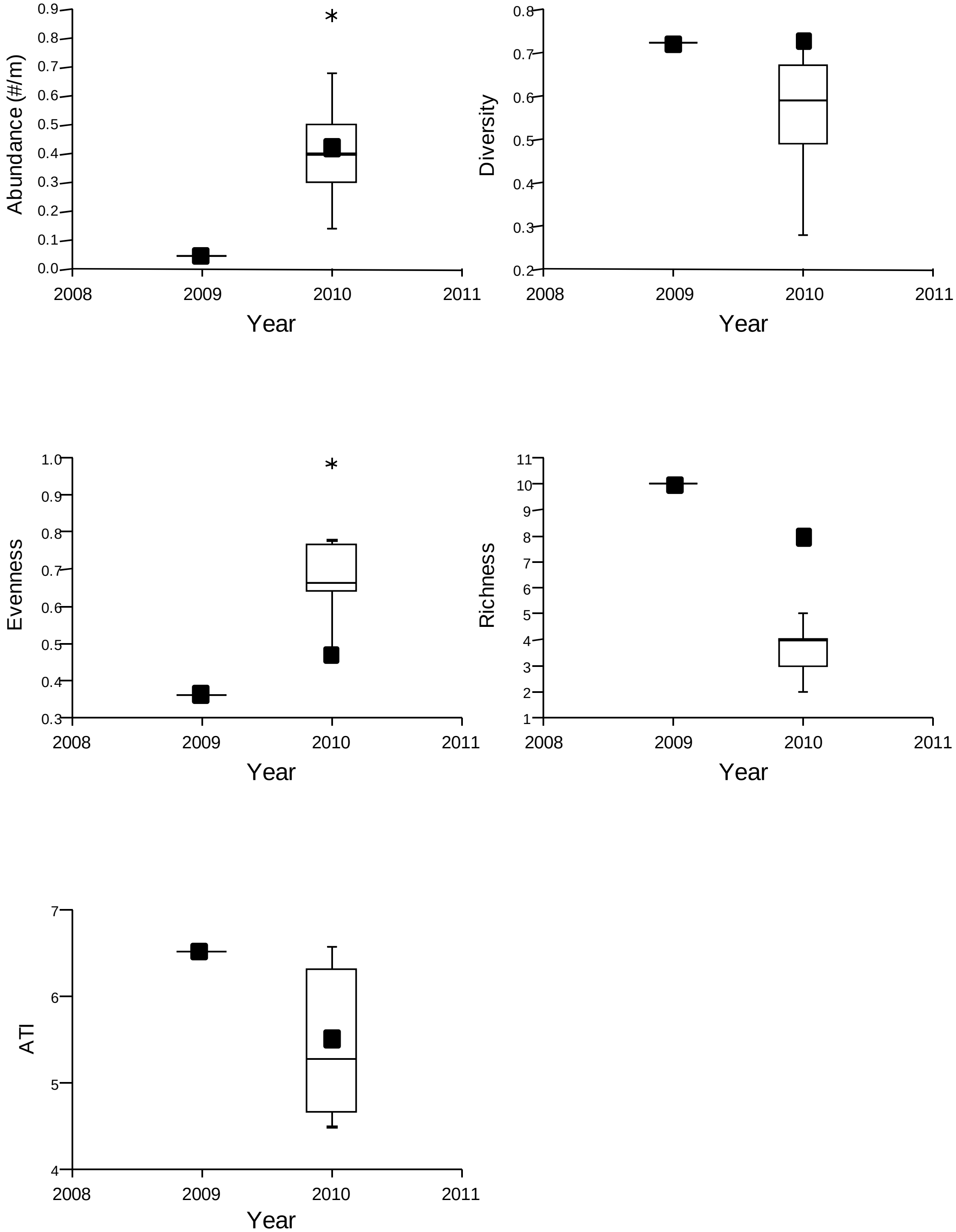
Note: Variations among sub-reaches in 2010 are illustrated using box plots. Black squares denote reach-wide means in 2009 and 2010.

Figure A2.5-3 Within-reach variation in values of measurement endpoints for fish assemblages in *test* reach MUR-F1, 2009 and 2010.



Note: Variations among sub-reaches in 2010 are illustrated using box plots. Black squares denote reach-wide means in 2009 and 2010.

Figure A2.5-4 Within-reach variation in values of measurement endpoints for fish assemblages in *test* reach STR-F1, 2009 and 2010.



Note: Variations among sub-reaches in 2010 are illustrated using box plots. Black squares denote reach-wide means in 2009 and 2010.

Table A2.5-7 Percent of total catch of each fish species in three reaches with two years of data, 2009 to 2010.

Common Name	Code	Tolerance Value	Jackpine Creek *Test* reach JAC-F1		Muskeg River *Test* reach MUR-F1		Steepbank River *Test* reach STR-F1	
			2009	2010	2009	2010	2009	2010
Arctic grayling	ARGR	2.0	0	0	0	0	0	0
brook stickleback	BRST	9.4	0	12.0	5.2	5.4	0	0
burbot	BURB	2.0	0	0	1.7	0	0	0
finescale dace	FNDC	7.0	0	44.5	0	16.1	0	3.8
fathead minnow	FTMN	8.3	0	0	0	0	0	0
lake chub	LKCH	5.5	14.3	0	6.9	8.6	6.1	0
longnose dace	LNDC	6.2	0	0	0	10.8	3.0	30.0
longnose sucker	LNSC	4.6	28.6	1.8	8.6	4.3	6.1	0
northern redbelly dace	NRDC	7.0	0	0	0	0	48.5	0
northern pike	NRPK	7.8	0	1	0	0	0	0
pearl dace	PRDC	6.7	0	12.6	0	37.6	6.1	30.5
slimy sculpin	SLSC	3.0	0	13.5	74.1	11.8	6.1	28.6
spoonhead sculpin	SPSC	3.0	0	0	1.7	3.2	0	1
spottail shiner	SPSH	7.7	0	0	0	0	0	0
sucker unidentified	-	7.6	0	0	1.7	0	0	0
trout perch	TRPR	8.4	0	5.3	0	0	3.0	3.3
unknown	UNK	-	0	0	0	0	15.2	0
walleye	WALL	8.7	0	0	0	0	3.0	0
white sucker	WHSC	7.6	57.1	9.4	0	2.2	3.0	1.9
yellow perch	YLPR	7.4	0	0	0	0	0	<1
Total Number of Species			**3**	**7**	**7**	**9**	**10**	**8**

Table A2.5-8 Coefficient of variation for measurement endpoints for FAM reaches, 2010.

Reach		Coefficient of Variation (%)				
		Abundance	Richness	Diversity	Evenness	ATI
ELR-F1	Depositional	68	34	20	19	3
ELR-F2	Erosional	43	19	20	26	3
JAC-F1	Depositional	99	61	56	62	54
MUR-F1	Erosional	42	33	27	17	8
STR-F1	Erosional	54	26	23	21	15

Note: JAC-F2 was not included because fishing effort was not separated into sub-reaches at this reach.

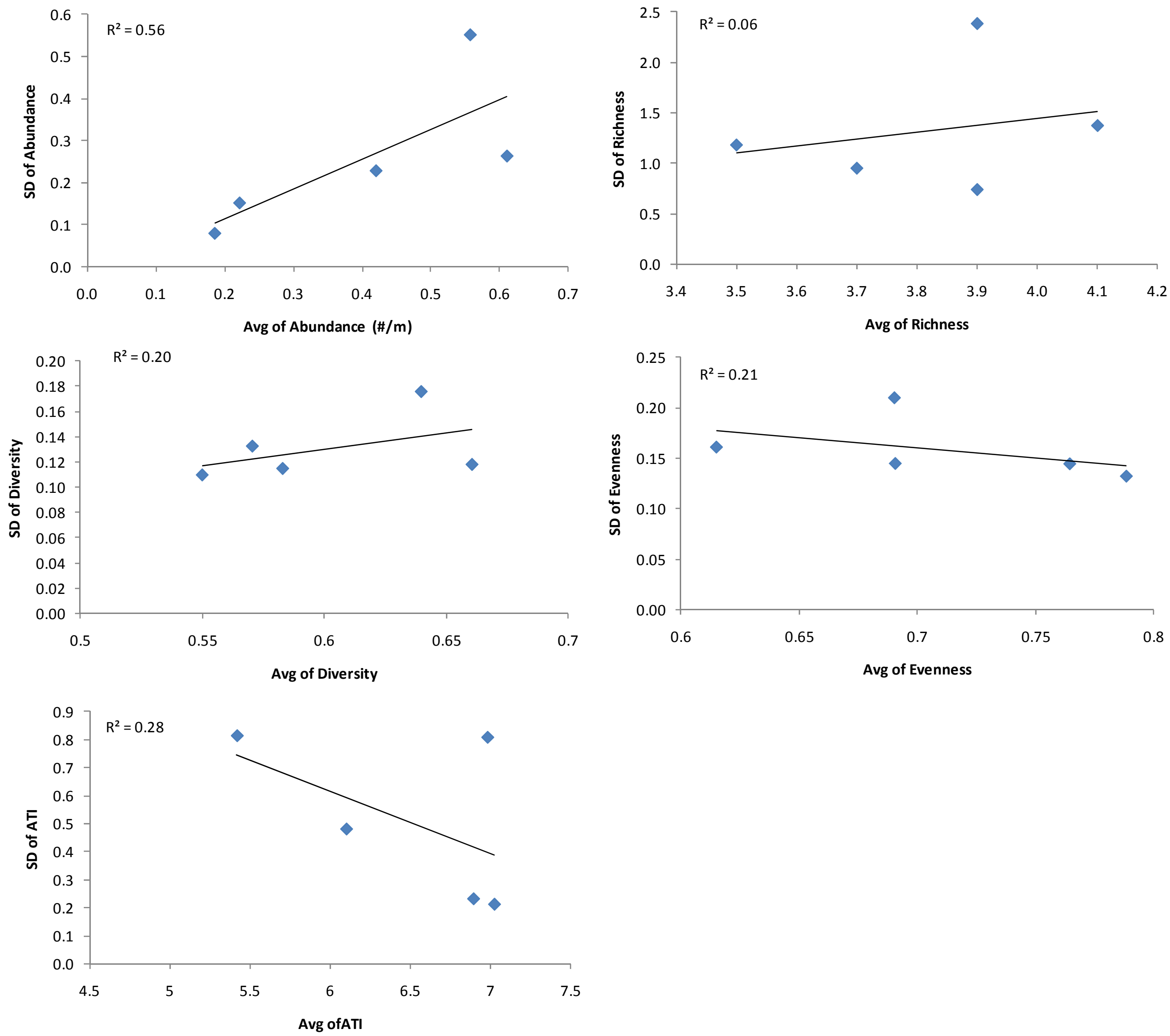
A2.5.4.4 Sample Size Requirements for Fish Assemblage Monitoring

The within-reach variation for each measurement endpoint across all reaches is provided in Table A2.5-9. The variance (standard deviation) in abundance, richness and diversity increased with increasing reach average and variance in evenness and ATI decreased with increasing reach average (Figure A2.5-5). Given the differing trends in variability across measurement endpoints, sample size requirements were calculated for the maximum and minimum measurement endpoint values (Table A2.5-9). The number of sub-reaches that would be needed to produce measurement endpoint values that were within ± 20% of the true reach average varied between one and 12, depending on the measurement endpoint (Table A2.5-9). Total abundance was the most variable measurement endpoint, requiring upwards of 12 sub-reaches to be sampled in reaches where abundance is high (0.6 fish per metre), or as few as seven sub-reaches when abundance was lower (0.2 fish per metre). ATI was the most precise measurement endpoint requiring only a single sub-reach to estimate the sub-reach mean. The precision requirement (±20% of the true average of sub-reaches) was met with four sub-reaches for species richness, three sub-reaches for evenness and two sub-reaches for diversity.

Table A2.5-9 Sample size of sub-reaches required for measurement endpoints for fish assemblages to obtain 20% of the true sub-reach average.

Community Index	Minimum/ Maximum Value	Standard Deviation (SD)	Sample Size (n)
Total Abundance (# fish per m)	0.2	0.1	7
	0.6	0.4	12
Richness	3.5	1.2	3
	4.1	1.5	4
Simpson's Diversity	0.55	0.115	2
	0.65	0.140	2
Evenness	0.61	0.175	3
	0.80	0.150	2
Assemblage Tolerance Index (ATI)	5.6	0.75	1
	7.0	0.40	1

Figure A2.5-5 Scatterplot of variance (standard deviation) in relation to average values of measurement endpoints for all FAM sampling reaches, 2010.



Note: each point represents a sampling reach.

A2.5.5 Habitat Assessments

Habitat data has been collected in both years of the pilot study and supporting information from the Water Quality and Benthic Invertebrate Communities components is available for further comparison between reaches. Given the objective of the pilot study was to determine if measurement endpoints could be developed to look at differences in fish assemblages between reaches and across years, the supporting data were not evaluated in 2010. The supporting data is primarily collected so that if a change was observed, a more thorough analysis could be conducted by interpreting all environmental characteristics of a reach.

A2.5.6 Historical Data

Historical data from the FWMIS (Fisheries and Wildlife Management Information System) database have been collected from various sources to identify the species composition in the vicinity of the fish assemblage monitoring reaches. Table A2.5-10 provides catch per unit effort (a measure of relative abundance) for species captured in other studies in the vicinity of the FAM reaches (i.e., within 500 m of the reach) compared to CPUE of fish species captured in 2009 and 2010 in the RAMP FAM pilot study. Species richness and presence of species is generally the same or higher in 2009 and 2010 compared to previous sampling years at the same location. Data available from historical years can provide a guide of the type of assemblage that should be present in each reach, although keeping in mind that sampling are conducted differently across studies with differing objectives.

Table A2.5-10 CPUE of species captured within and in the vicinity of RAMP FAM reaches (within 500 m), 1999 to 2010.

Watercourse	Reach	Year	No. Species	Brook stickleback	Burbot	Finescale dace	Longnose sucker	Lake chub	Longnose dace	Northern pike	Pearl dace	Slimy sculpin	Spoonhead sculpin	Trout-perch	Walleye	White sucker	Yellow perch
Steepbank River	STR-F1	1999	7	-	0.070	-	0.349	0.233	0.581	-	1.628	1.372	-	1.279	-	-	-
		2000	4	-	-	-	-	0.577	2.309	-	-	4.906	2.597	-	-	-	-
		2004	2	-	-	-	-	-	-	-	-	1.709	1.352	-	-	-	-
		2009	10	-	-	-	0.055	0.055	0.027	-	0.055	0.055	-	0.027	0.027	0.027	-
		2010	8	-	-	0.202	-	-	1.591	-	1.617	1.516	0.076	0.177	-	0.101	0.025
Muskeg River	MUR-F1	1999	4	-	0.097	-	0.645	-	-	-	1.290	1.451	-	-	-	-	-
		2000	4	-	-	-	-	0.833	2.917	-	0.833	15.833	-	-	-	-	-
		2004	2	-	-	-	-	-	-	-	-	0.066	2.831	-	-	-	-
		2009	7	0.146	0.049	-	0.244	0.195	-	-	-	2.097	0.049	-	-	-	-
		2010	10	0.130	-	0.562	0.216	0.173	0.434	-	1.254	0.411	0.087	-	-	0.108	0.022
Jackpine Creek	JAC-F1	1997	8	-	-	-	0.062	0.326	0.139	0.062	0.062	-	0.278	-	0.062	0.685	-
		2000	1	-	-	-	1.379	-	-	-	-	-	-	-	-	-	-
		2009	3	-	-	-	0.090	0.045	-	-	-	-	-	-	-	0.180	-
		2010	8	0.492	-	1.941	0.078	-	-	0.026	0.543	0.595	-	0.233	-	0.414	-
Ells River	ELR-F2	2002	2	-	-	-	-	1.513	0.757	-	-	-	-	-	-	-	-
		2010	5	-	-	4.041	0.328	-	-	-	2.071	-	-	0.101	-	1.162	-

A2.5.7 Discussion and Recommendations

The fish assemblage pilot study in 2010 demonstrated that generally, the collection of fish from four sub-reaches would adequately characterize the average sub-reach measurement endpoint values. Total abundance of fish per lineal metre was the most variable measurement endpoint and would require up to 12 sub-reaches in order to produce estimates that were within ±20% of the true mean of sub-reach value. The assemblage tolerance index (ATI) was the most precise index, requiring data from a single sub-reach to achieve the same level of precision.

The measurement endpoint that explains tolerance of species is generally less variable than those that describe abundance or richness because of redundancies among taxa. Northern redbelly dace and finescale dace, for example, are similar species, as are slimy sculpin and spoonhead sculpin. These species may vary in abundance from time to time, and may replace each other, or co-exist. Abundance will; therefore, be variable, while the niches that they occupy will remain occupied by similar species. The result is that generally, the average taxonomic tolerance is generally more stable than the actual count of fish. Measurement endpoints such as ATI are; therefore, excellent measures that can be used for the detection of meaningful trends in taxonomic composition. If the required sample size is based on the requirement to obtain precision in the ATI value, then it is adequate to conclude that a single sub-reach with a catch of approximately 50 fish would be adequate for future monitoring of oil sands development.

The influence of sample size on estimates of species richness, diversity and evenness is that shorter reaches produced fewer species, lower diversity and higher evenness. Thus, any comparison among or within reaches must consider the length of the reach for these key conventional metrics, whereas for measurement endpoints such as abundance and assemblage tolerance index (ATI), a standardized reach length is not as important as there was little variability with reach length.

The measurement endpoints used to make assessments for the Fish Assemblage Monitoring pilot study is not a complete list and more can be evaluated, if this methodology continues to be used as a monitoring tool.

The two year pilot FAM study has helped to determine the level of fishing effort and catch required to provide statistically robust measurement endpoints that can be used to assess potential changes due to oil sands development. Measurement endpoints have been developed based on Canadian EEM protocols, which can be compared across time and space, if the FAM program continues as a monitoring tool under RAMP.

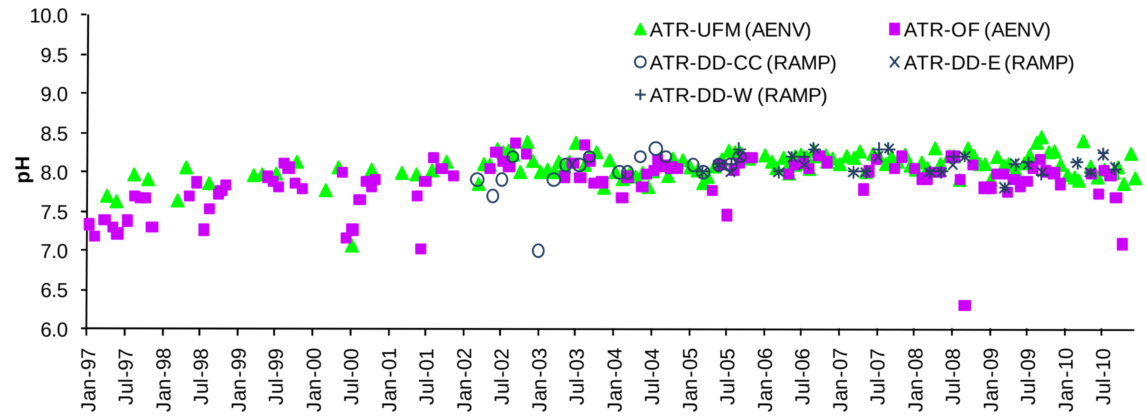
A2.6 COMPARISON BETWEEN AENV AND RAMP WATER QUALITY DATA

This section provides supporting information to Recommendation #60 related to the temporal comparisons of long-term datasets.

Figure A2.6-1 Water quality measurement endpoints, 1997 to 2010 AENV and RAMP data for the Athabasca River mainstem.

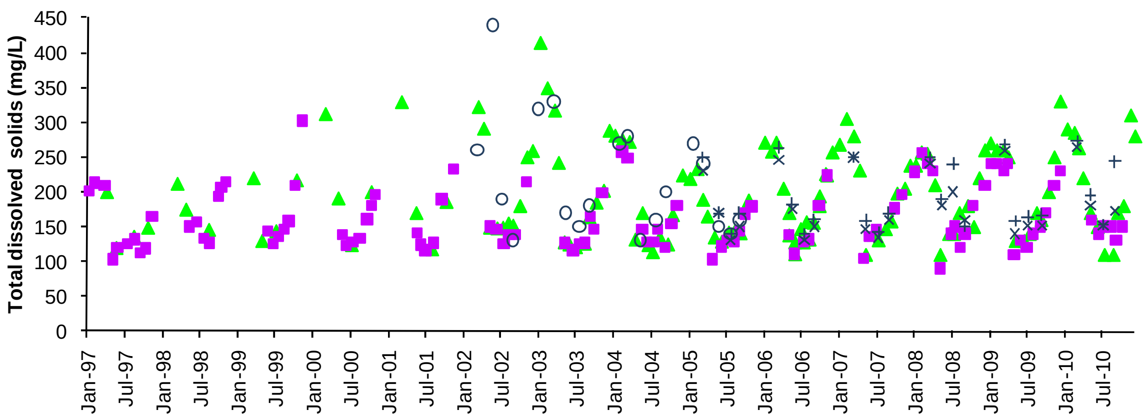
pH

Trend at ATR-UFM: up Trend at ATR-OF: up



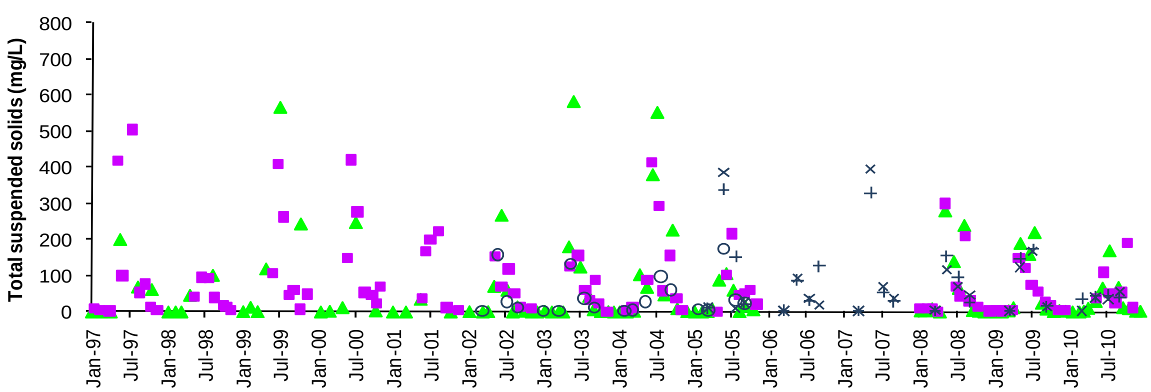
Total dissolved solids

Trend at ATR-UFM: none Trend at ATR-OF: none



Total suspended solids

Trend at ATR-UFM: none Trend at ATR-OF: none

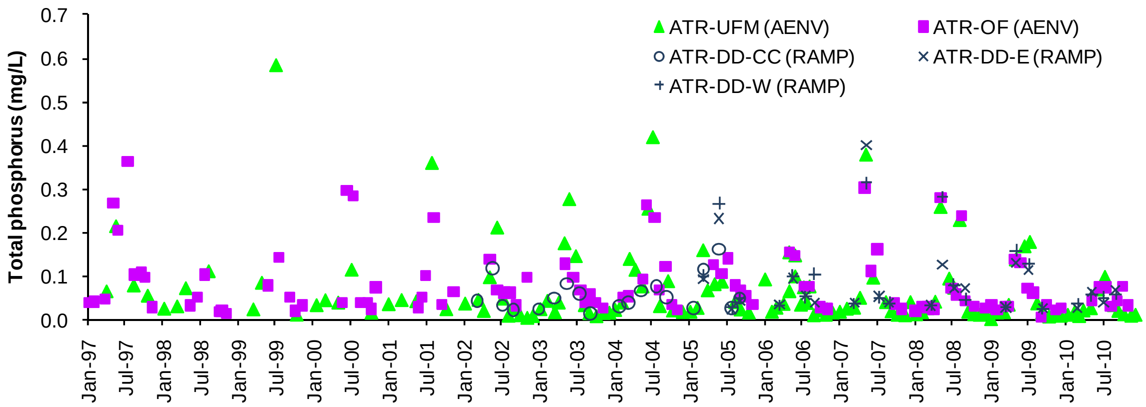


Non-detectable results are shown at the detection limit.

Figure A2.6-1 (Cont'd.)

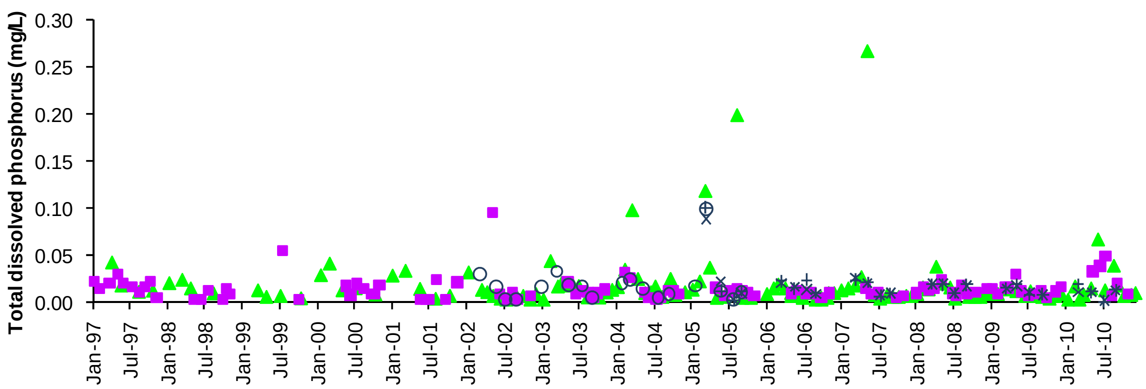
Total phosphorus

Trend at ATR-UFM: down Trend at ATR-OF: down



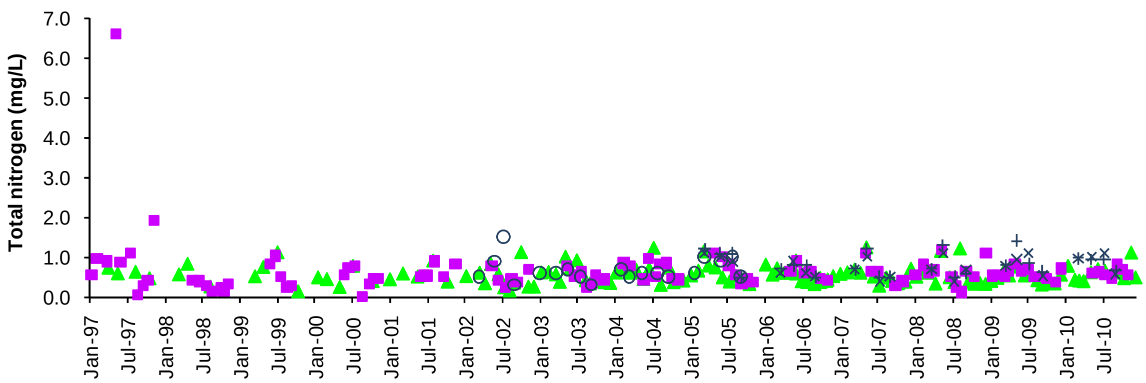
Total dissolved phosphorus

Trend at ATR-UFM: none Trend at ATR-OF: up



Total nitrogen

Trend at ATR-UFM: none Trend at ATR-OF: up

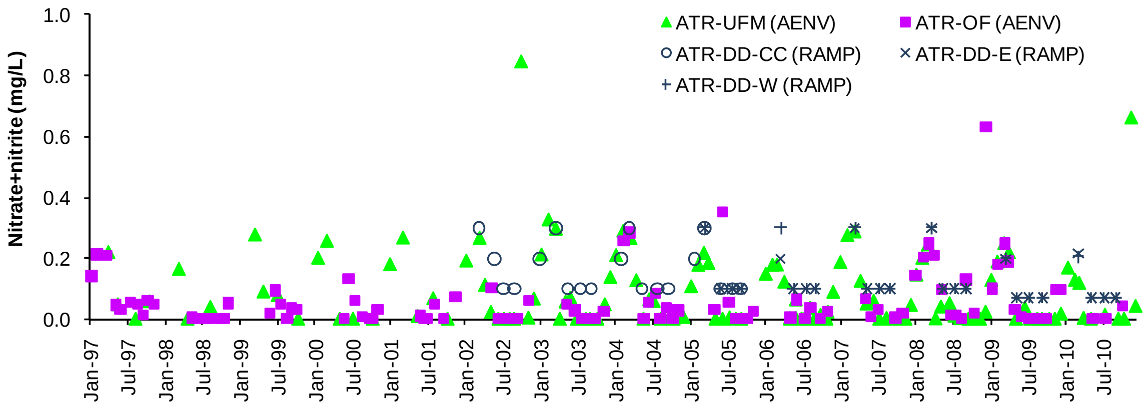


Non-detectable results are shown at the detection limit.

Figure A2.6-1 (Cont'd.)

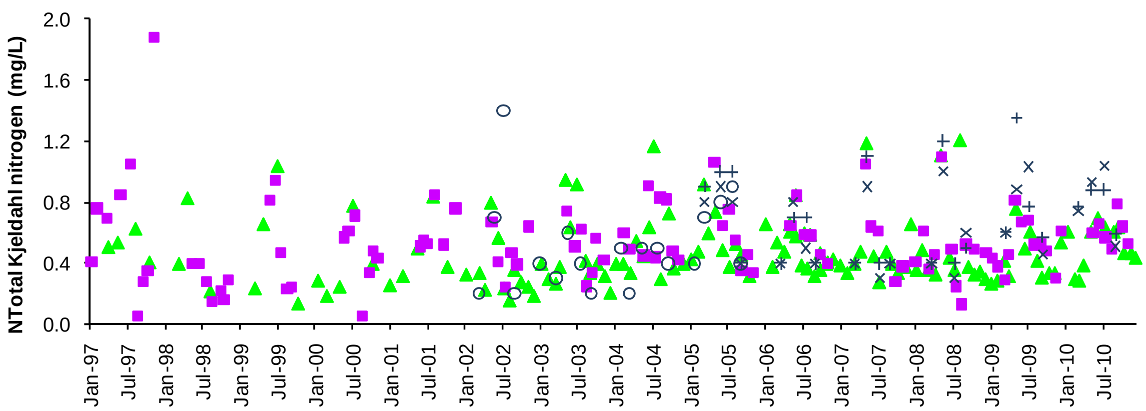
Nitrate + Nitrite

Trend at ATR-UFM: down Trend at ATR-OF: none



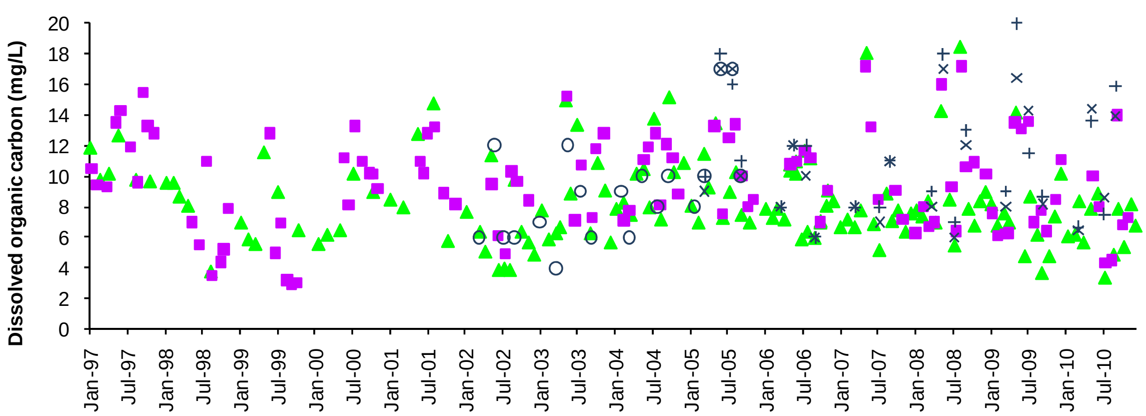
Total Kjeldahl nitrogen

Trend at ATR-UFM: none Trend at ATR-OF: up



Dissolved organic carbon

Trend at ATR-UFM: none Trend at ATR-OF: none

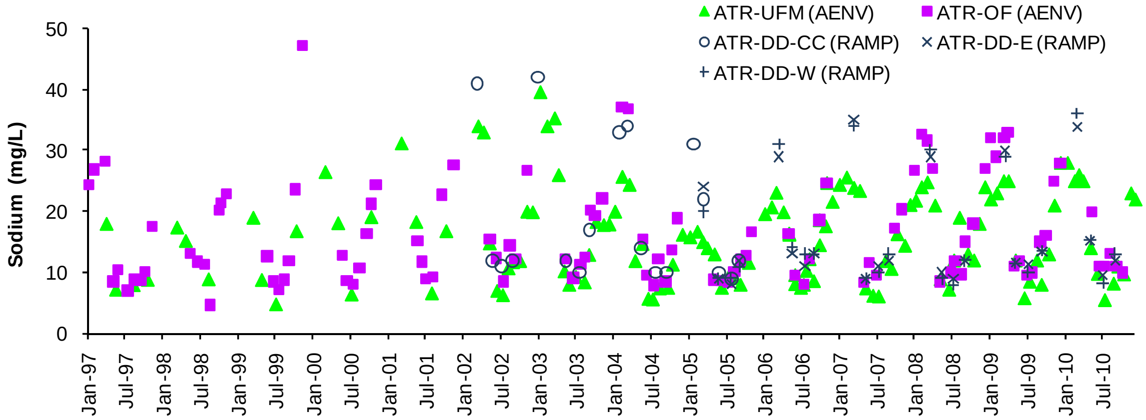


Non-detectable results are shown at the detection limit.

Figure A2.6-1 (Cont’d.)

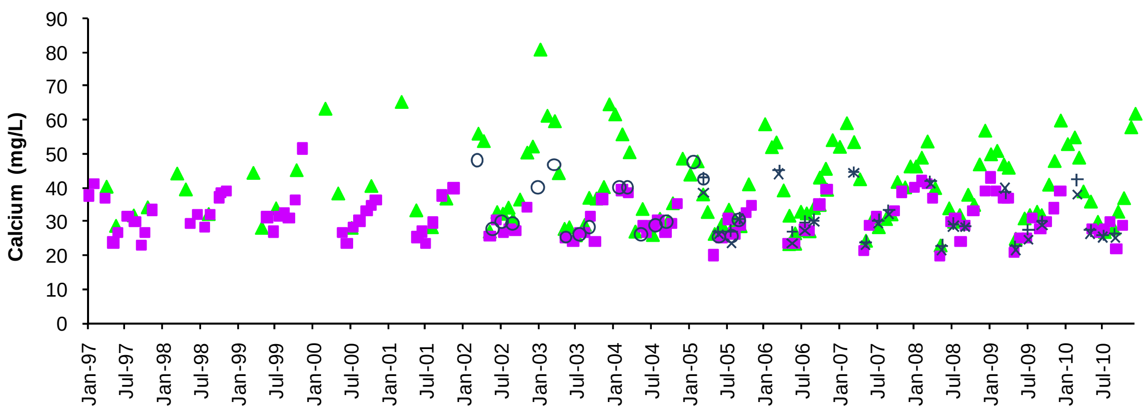
Sodium

Trend at ATR-UFM: none Trend at ATR-OF: none



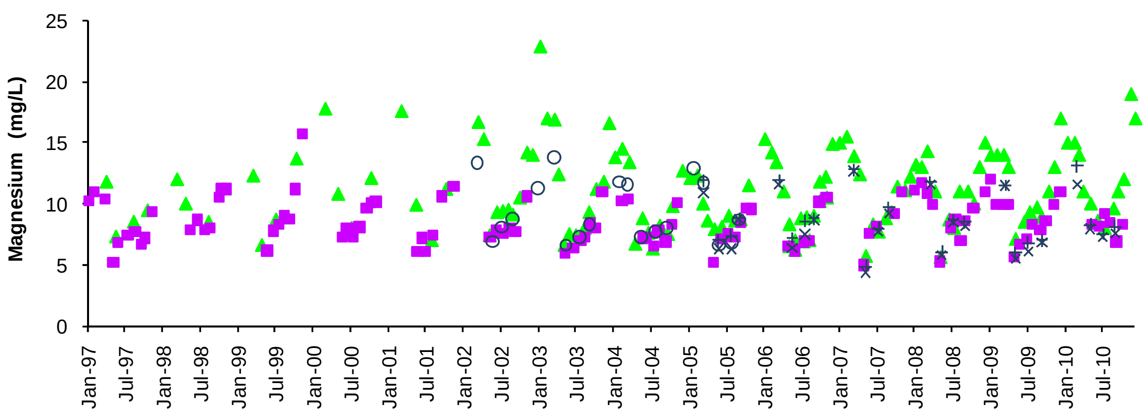
Calcium

Trend at ATR-UFM: none Trend at ATR-OF: none



Magnesium

Trend at ATR-UFM: none Trend at ATR-OF: none

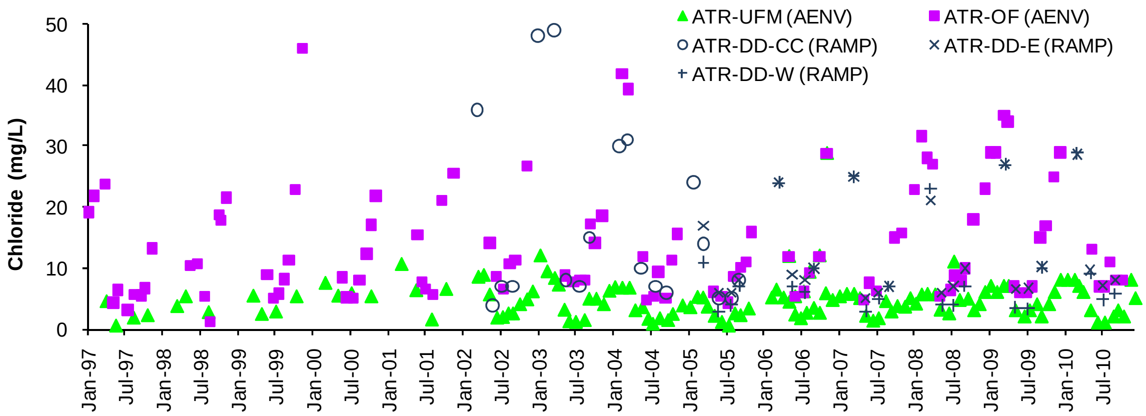


Non-detectable results are shown at the detection limit.

Figure A2.6-1 (Cont'd.)

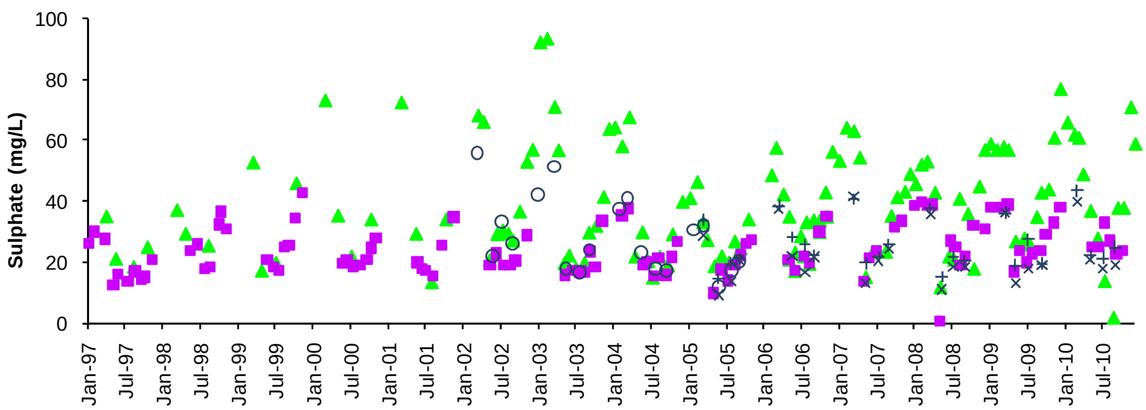
Chloride

Trend at ATR-UFM: none Trend at ATR-OF: none



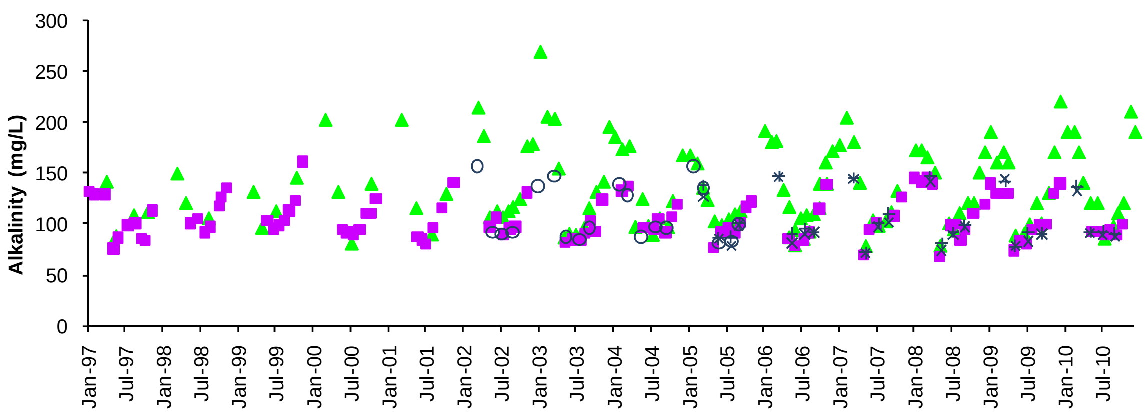
Sulphate

Trend at ATR-UFM: none Trend at ATR-OF: up



Alkalinity (as $CaCO_3$)

Trend at ATR-UFM: none Trend at ATR-OF: none

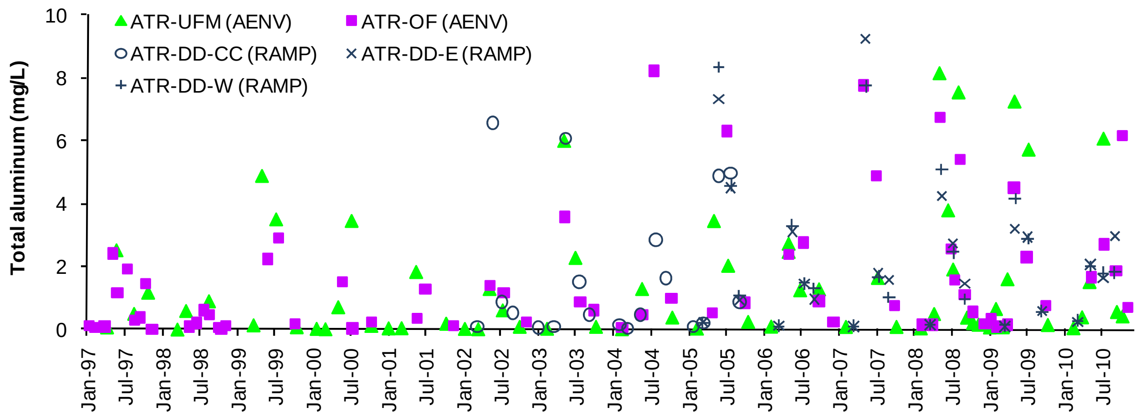


Non-detectable results are shown at the detection limit.

Figure A2.6-1 (Cont'd.)

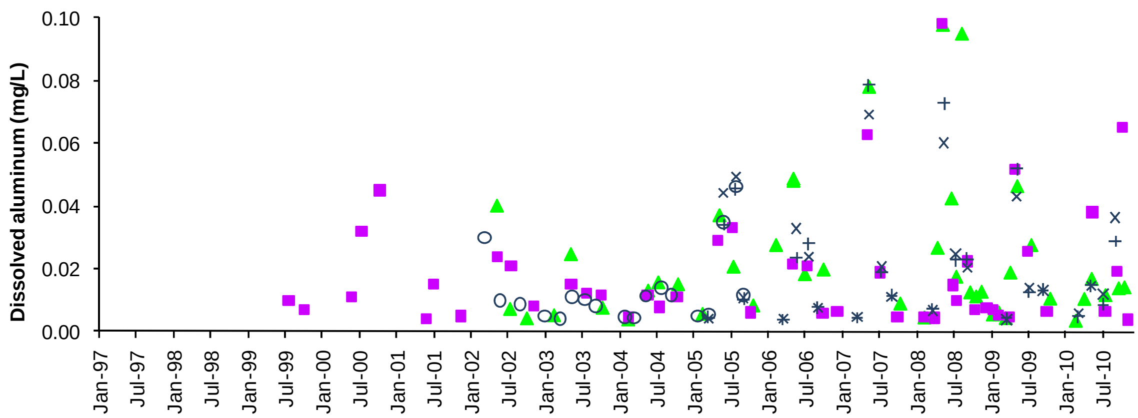
Total aluminum

Trend at ATR-UFM: up Trend at ATR-OF: up



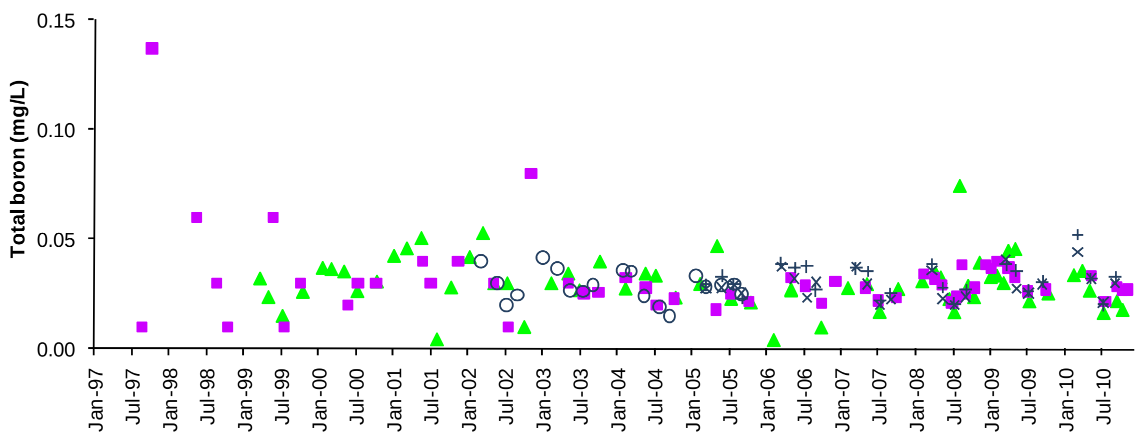
Dissolved aluminum

Trend at ATR-UFM: up Trend at ATR-OF: none



Total boron

Trend at ATR-UFM: none Trend at ATR-OF: none

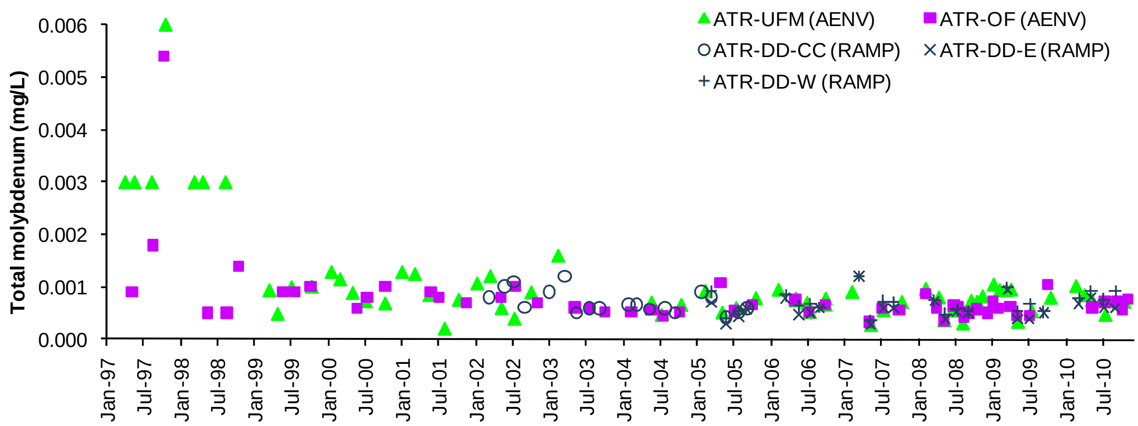


Non-detectable results are shown at the detection limit.

Figure A2.6-1 (Cont'd.)

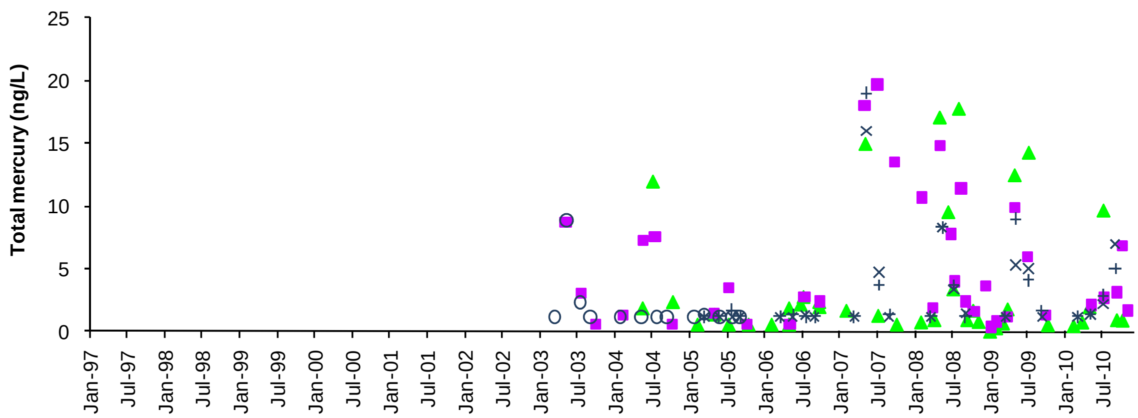
Total molybdenum

Trend at ATR-UFM: none   Trend at ATR-OF: down



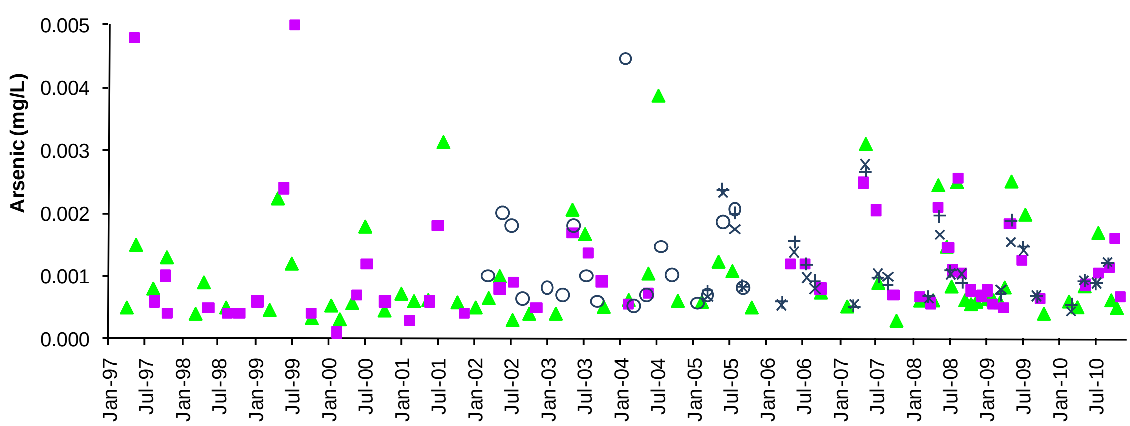
Total mercury (ultra-trace)

Trend at ATR-UFM: none   Trend at ATR-OF: none



Total Arsenic

Trend at ATR-UFM: up   Trend at ATR-OF: none



Non-detectable results are shown at the detection limit.

A2.7 ATHABASCA RIVER FISH INVENTORY RESULTS

This section provides supporting information to Recommendation #110 related to the presentation of results from the RAMP Athabasca River fish inventory program.

Figure A2.7-1 Mean CPUE (± 1SD) of large-bodied KIR fish species combined in spring, summer and fall from 1987 to 2010.

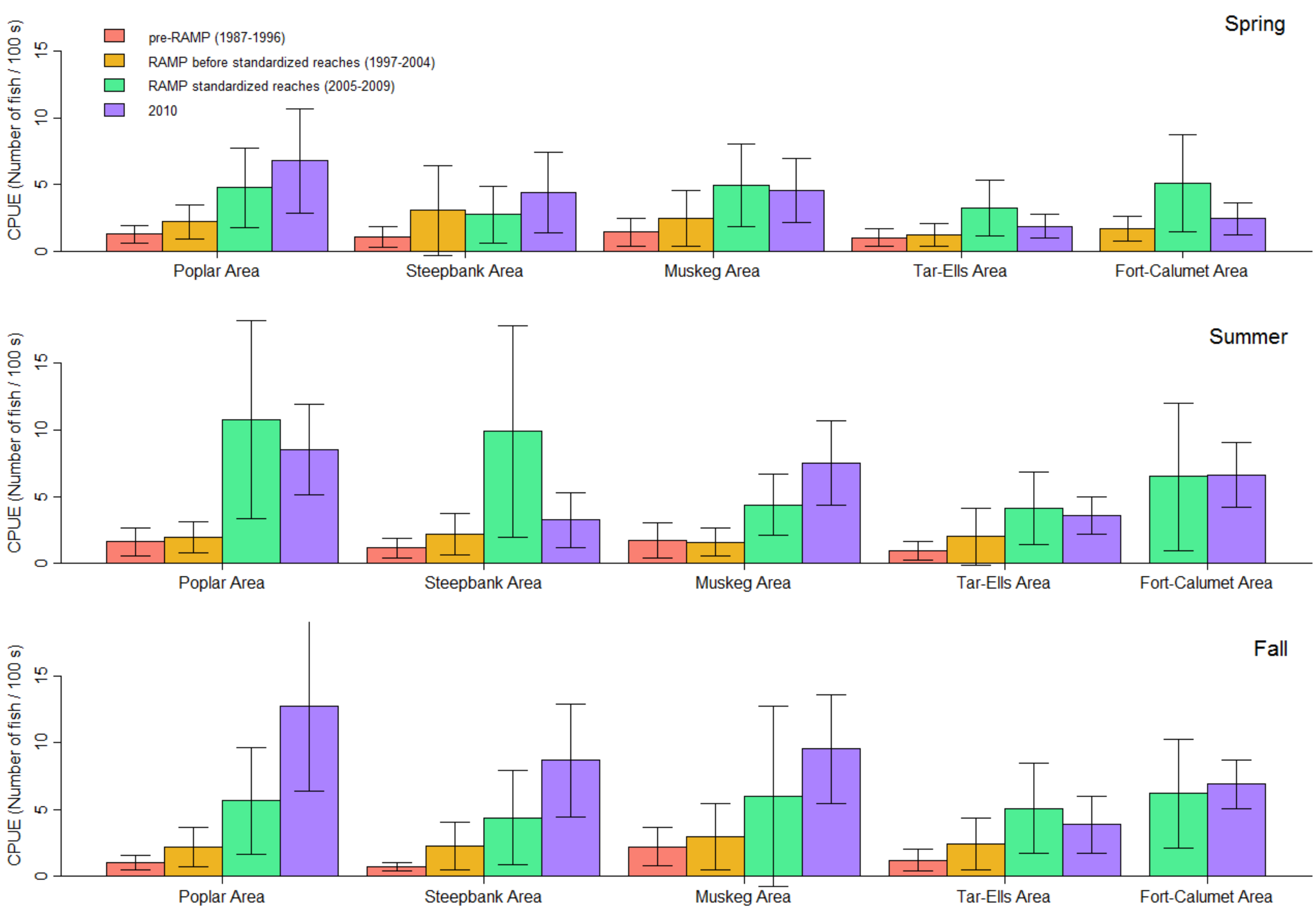


Figure A2.7-2 Spatial comparisons of mean CPUE (± 1SD) of large-bodied KIR fish species in spring, summer and fall 2010 in the Athabasca River.

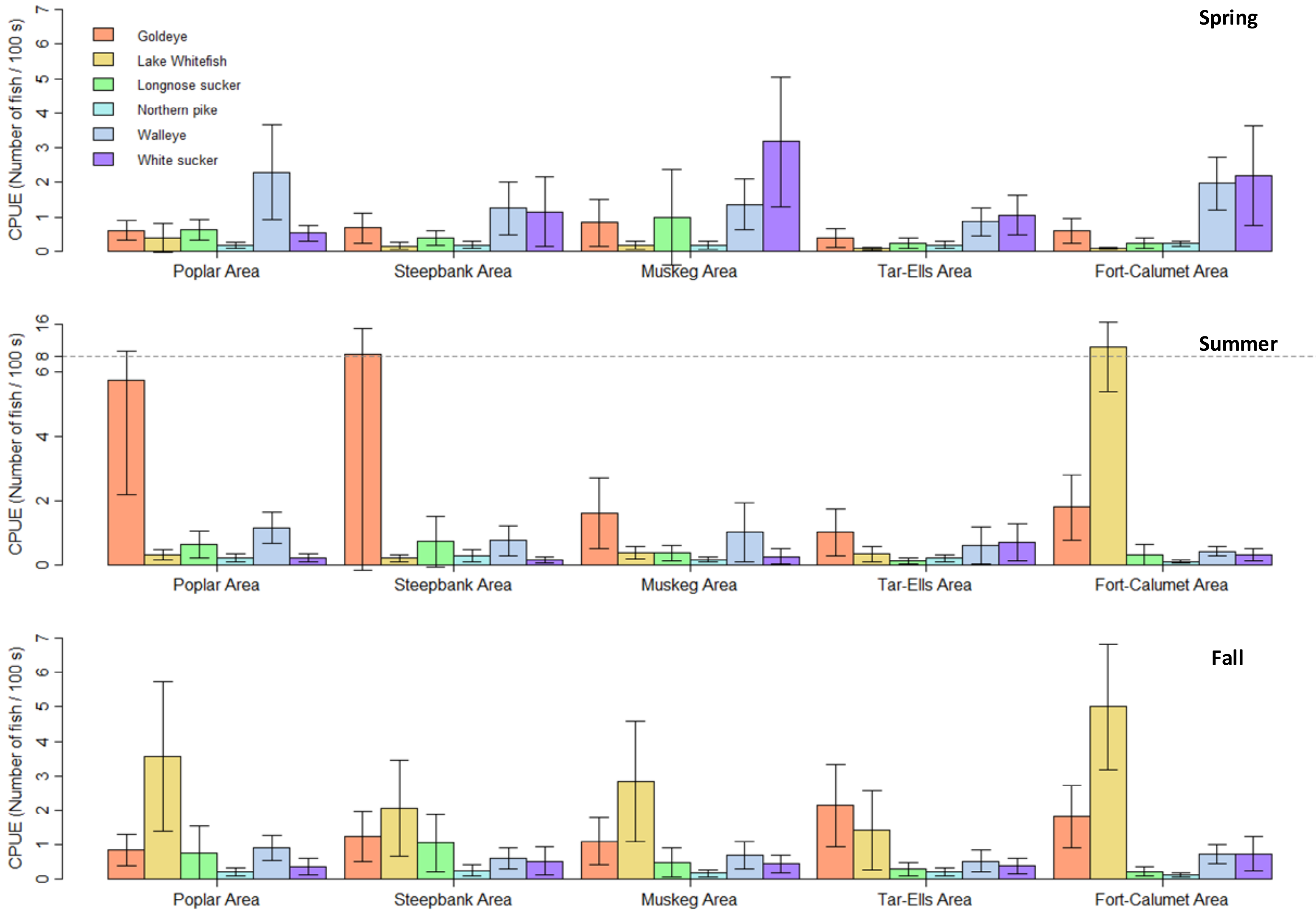


Figure A2.7-3 Relative length-frequency distributions for goldeye captured in the Athabasca River in 2010 (n=298) compared to the average from 1997 to 2009 (period of RAMP sampling sands), and the average from 1987 to 1996 (pre-RAMP); 50 mm length classes.

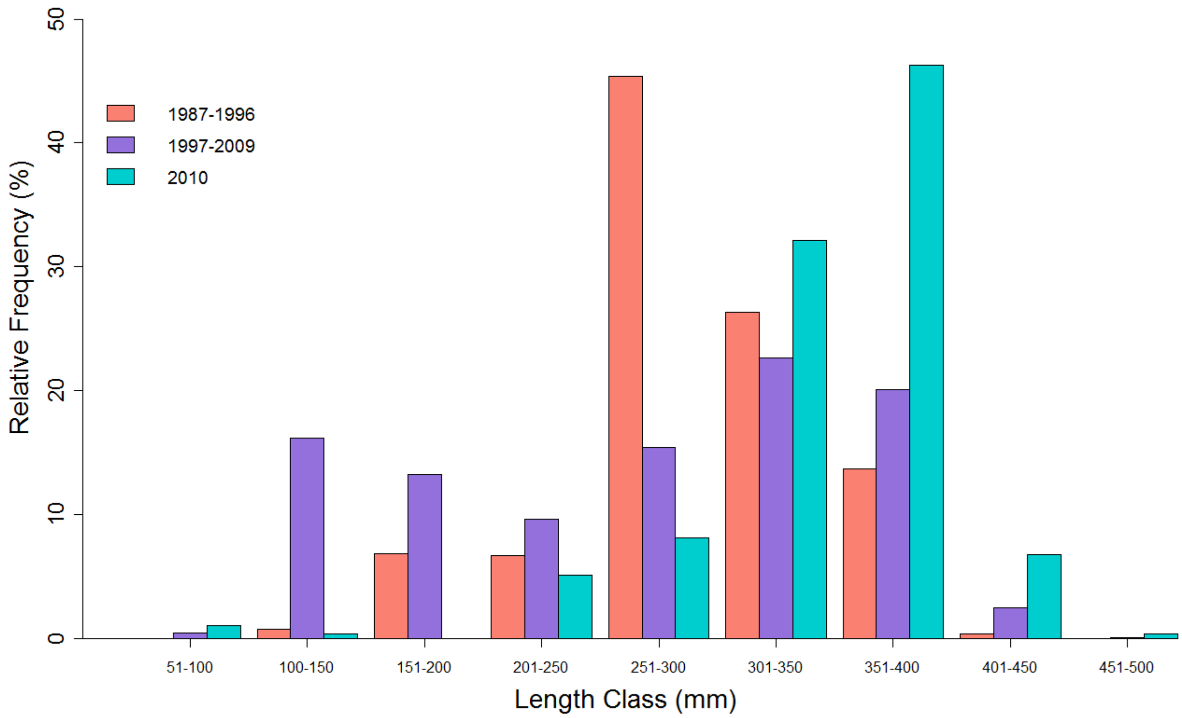


Figure A2.7-4 Relative length-frequency distributions for longnose sucker captured in the Athabasca River in 2010 (n=117) compared to the average from 1997 to 2009 (RAMP sampling period) and from 1987 to 1996 (pre-RAMP); 50 mm length classes.

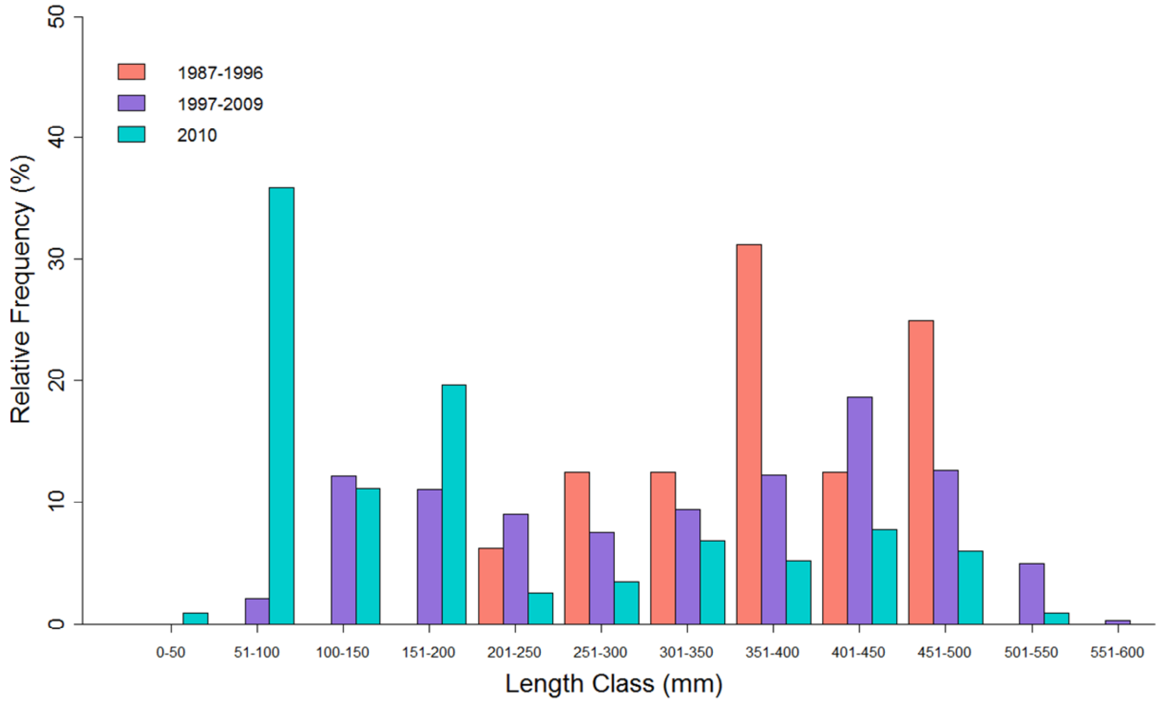


Figure A2.7-5 Relative length-frequency distributions for northern pike captured in the Athabasca River in 2010 (n=86) compared to the average from 1997 to 2009 (RAMP sampling period), and the average from 1987 to 1996 (pre-RAMP); 50 mm length classes.

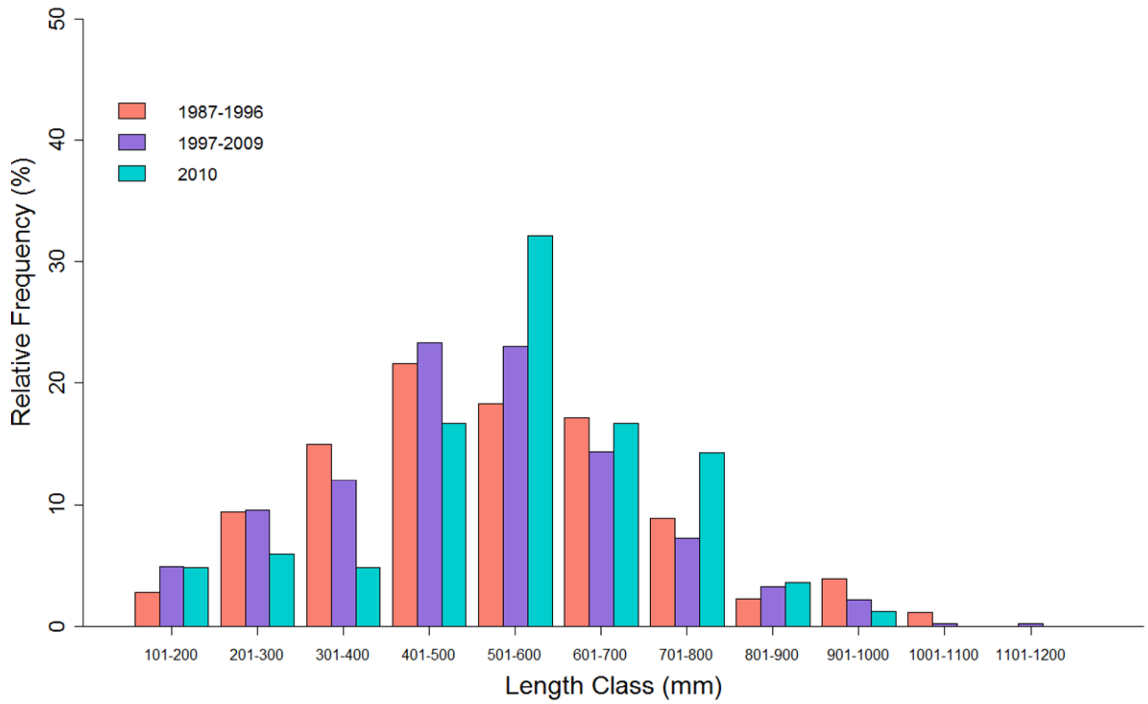


Figure A2.7-6 Relative length-frequency distributions for walleye captured in the Athabasca River in 2010 (n=572) compared to the average from 1997 to 2009 (RAMP sampling period), and the average from 1987 to 1996 (pre-RAMP); 50 mm length classes.

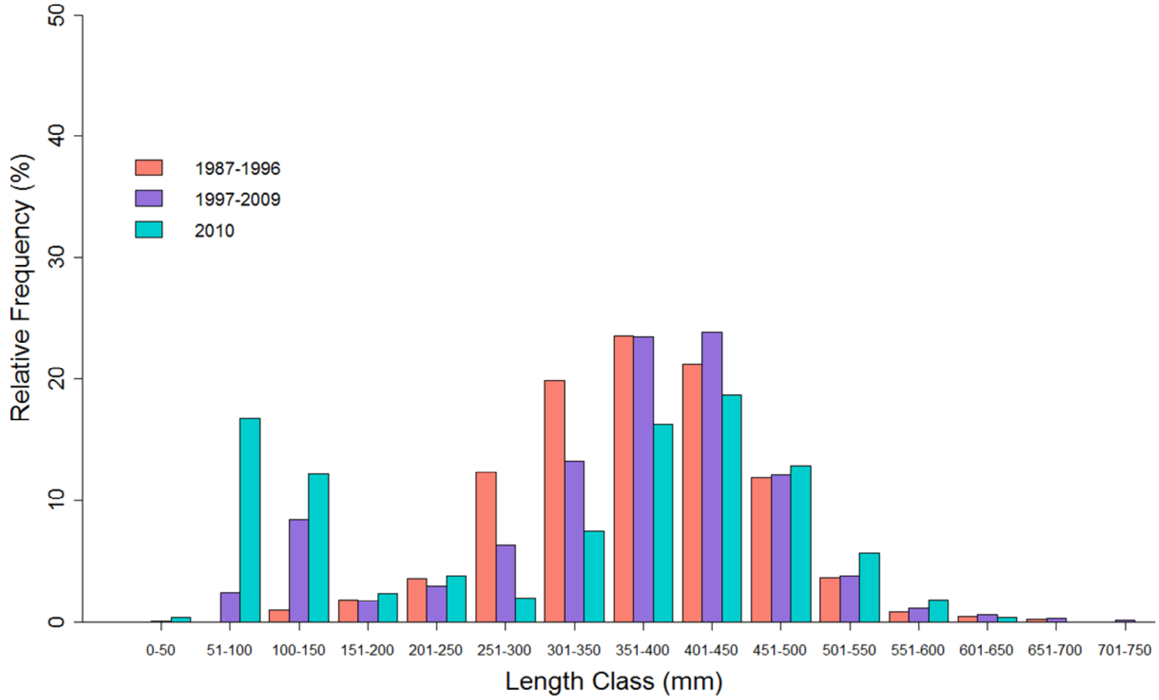


Figure A2.7-7 Relative length-frequency distributions for white sucker captured in the Athabasca River in 2010 (n=235) compared to the average from 1997 to 2009 (RAMP sampling period), and the average from 1987 to 1996 (pre-RAMP); 50 mm length classes.

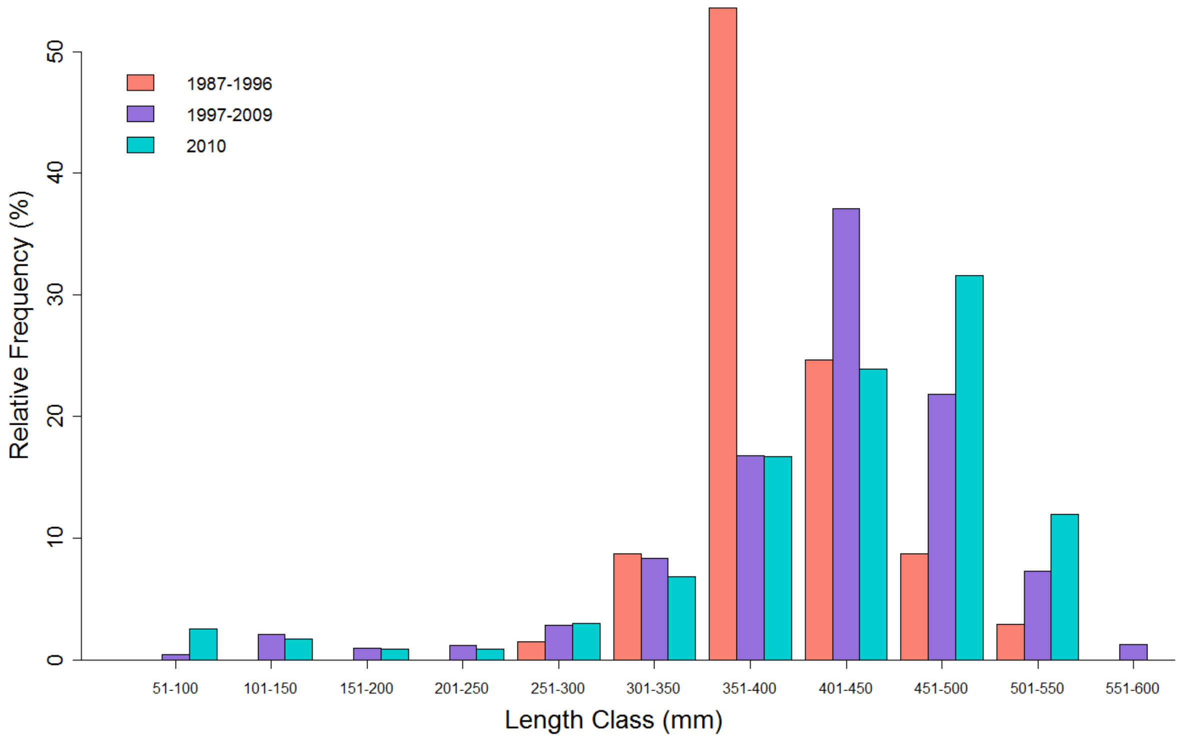


Figure A2.7-8 Mean condition (± 1SE) of goldeye captured during the spring, summer, and fall inventories from 1997 to 2010 in the Athabasca River, relative to pre-RAMP values (1987 to 1996).

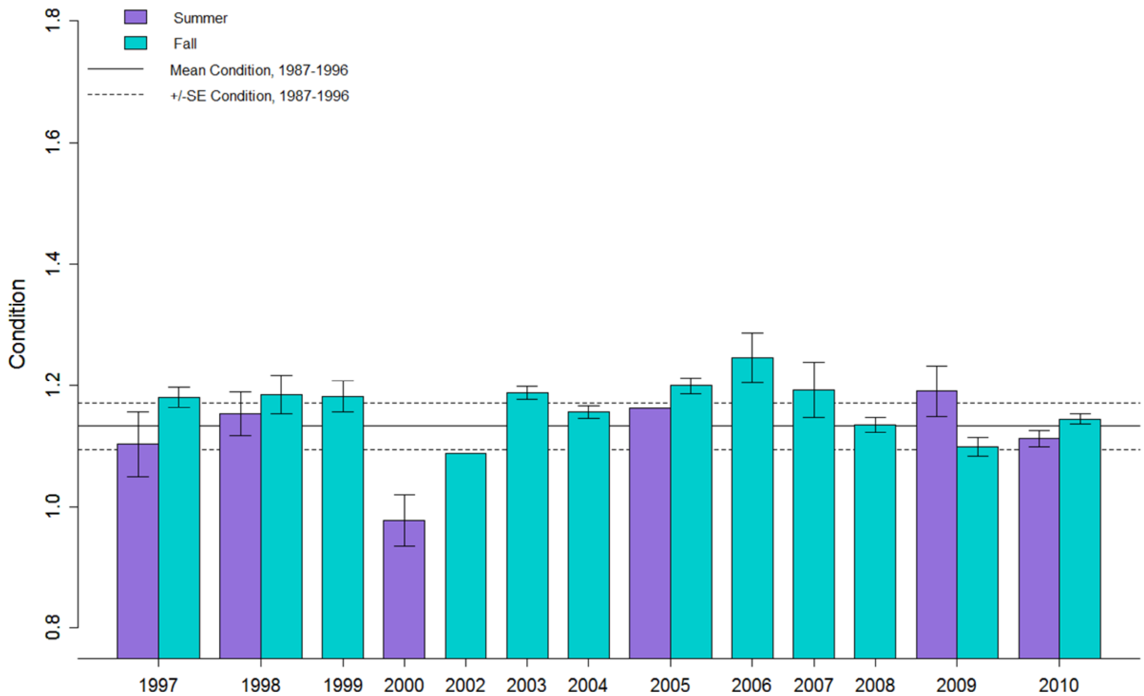


Figure A2.7-9 Mean condition (± 1SE) of longnose sucker captured during the spring, summer, and fall inventories from 1997 to 2010 in the Athabasca River, relative to pre-RAMP values (1987 to 1996).

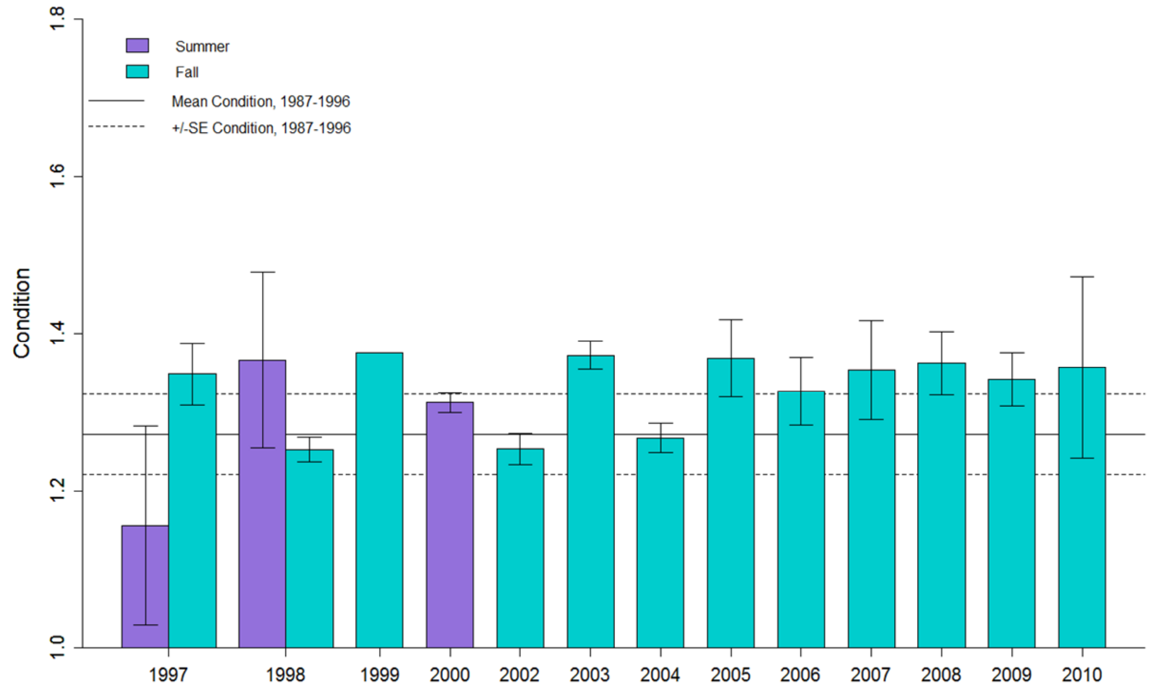


Figure A2.7-10 Mean condition (± 1SE) of northern pike captured during the spring, summer, and fall inventories from 1997 to 2010 in the Athabasca River, relative to pre-RAMP values (1987 to 1996).

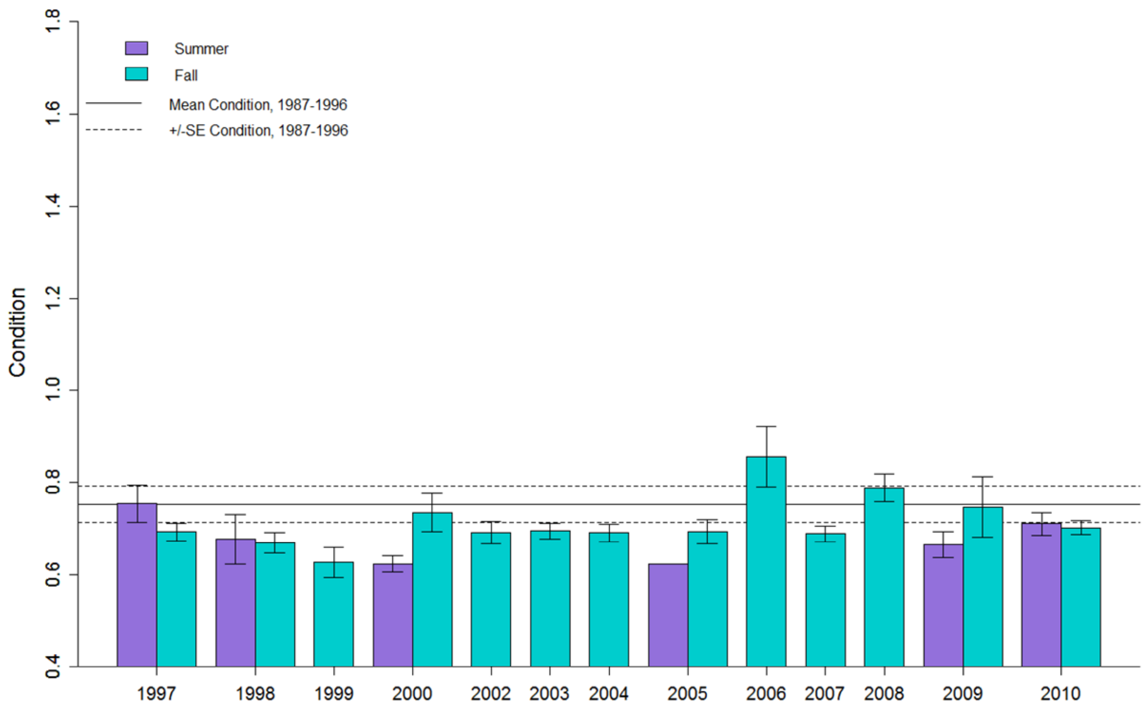


Figure A2.7-11 Mean condition (± 1SE) of walleye captured during the spring, summer, and fall inventories from 1997 to 2010 in the Athabasca River, relative to pre-RAMP values (1987 to 1996).

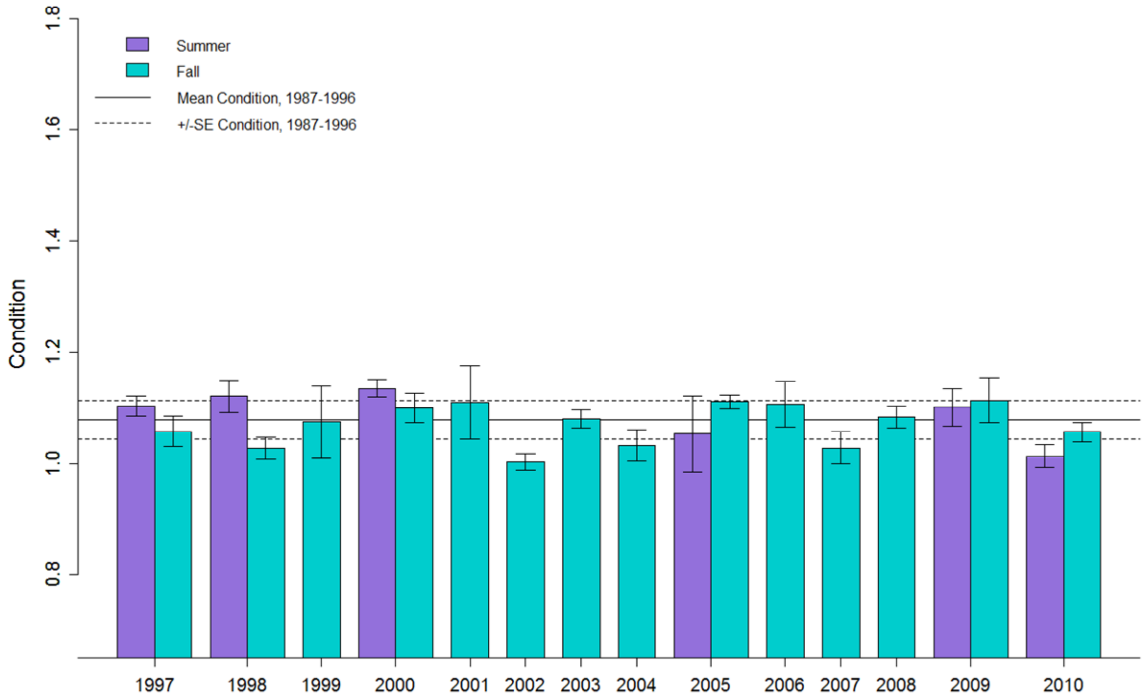


Figure A2.7-12 Mean condition (± 1SE) of white sucker captured during the spring, summer, and fall inventories from 1997 to 2010 in the Athabasca River, relative to pre-RAMP values (1987 to 1996).