Peer Review Response

Formulation and Evaluation of New Control Structures in the Great Lakes System

This document contains our response (Tolson, Razavi, Asadzadeh) to the helpful comments of two reviewers on our May 9, 2011 version of the IUGLS report entitled: **Formulation and Evaluation of New Control Structures in the Great Lakes System.**

Included in this document is a copy of the reviewer original comments.

Overall, the revised report is more substantive and contains additional analyses not included in the previous version. We also believe we have adequately addressed all the comments of the reviewers. Substantive material added/revised in this report relative to the May 9 report include:

- Better describing the LOSL 50,000 year stochastic sequence
- Improved analysis of various solutions in terms of describing their vulnerability and resilience and providing example lake level time series
- Additional solutions generated for two control point UW plans
- Additional solutions generated when Montreal was included in the objective function
- Improved description of our multi-lake rule curves
- Improved description of our risk-based (or frequency-based) objective function
- Complete documentaion on the simulation equations for the Lower St Lawrence River
- Updated multi-objective tradeoffs between cost and frequency-based objective function after more optimization (including one with Montreal in the objective function)
- Summary of Findings for Managing Montreal Water Levels (see Section 3.4)
- A more comprehensive set of conclusions and a revised list of future work tasks

REVIEWER 1 COMMENTS AND RESPONSES:

Please note that review comments are in the red font and the responses in black fonts.

B. What is the most critical aspect of the study/analysis? Why?

The problem formulations shown in equations (20) and (21) that seek to minimize relativistic changes in lake levels from a simulated baseline are the primary drivers of the results and recommendations of the report. Additionally the "50,000 year sequence" used to test the robustness of optimized results is the primary tool for trying to evaluate variation beyond what has been historically observed.

C. Which aspect of the analysis/modeling is weakest? Why? How can it be improved?

Overall the weakest points of the analysis lie in the problem formulation and 50,000 year sequence. Overall the report states that an investment of $8.5 Billion will potentially reduce extreme deviations from the baseline scenario. The challenge in these findings is it is unclear what the objective in equation (20) means? What impacts would the variation have? Are there thresholds for damage? Failures? The multi-objective approach is interesting but the problem formulation seems highly limited and distant from stakeholder metrics of concern (risk, vulnerability, reliability, resilience, impact). To be clear, I am not stating that everything must be in monetary terms but it would have been easier to evaluate the results had the performance objective been formulated in the context of more traditional vulnerability or reliability measures. There is little physical meaning the in the baseline at present. It is also unclear when or where the min-max/max-min robustness formulation in equation (21) was implemented. The work has an overall view of trying to minimize time of analysis and computational demands which seems inappropriate. Given the $8.5 Billion recommendation, I would strongly submit that a more careful and computationally explorative analysis be completed before any of these recommendations are considered.

1. The guidance we were given from the IUGLS Study Board was to emphatically steer clear of other metrics of stakeholder concern since any available methods for measuring them were completely unreliable at the extreme water levels we would be simulating. The Study Board recommended the two objectives we utilized in this report. However, as suggested by the reviewer, there were other ways we could evaluate performance (vulnerability and resilience). For example, in our revised report, for additional regulation plans we quantified performance in terms of vulnerability (magnitude of violations) and resilience (length of violations). See for example Figures 7, 8, 19, 20 for vulnerability representation and Tables 5, 13 for resilience representation. Regrettably, we didn't sufficiently clarify the formulation we developed for optimization. Otherwise, we believe the reviewer would have seen that we did explicitly account for risk

/reliability of the system in the objective function. To better clarify the details, we should highlight the following points first. We also modified the text (section 2.1) to include the clarifications on responses #2, 3, and 4 below.

2. **Definition of failure and failure thresholds:** the objective of the development on the Great Lake system (i.e., building new structures and/or excavation) was assumed to keep water levels in the historical extreme water levels everywhere. For each evaluation point (i.e., Lake Superior, Lake MH, …) at each month of the year, the maximum and minimum simulated historical levels, which are given, are considered as the failure thresholds and thus the basis for computing the reliability of not violating them. As such, for an evaluation point at a specific month, we defined failure as when the water level goes beyond one of these extremes.

   Section 2.5.1 is now modified slightly (second sentence below new) to include the following statement:

   > The primary objective of the UW plan is to maintain the water level at these nine evaluation points within the range simulated for the historical period (1900-2008) when experiencing any future climate condition. A risk-based objective function was defined to keep water level in this desirable range, as going beyond the desirable range was deemed a failure.

3. **Interpretation/meaning of the objective function:** the objective function we developed <u>is</u> a risk-based objective function and we now refer to it this way at various locations in the revised report (although we also still refer to it as a frequency-based objective function in some locations). Section 2.5.2 is now revised significantly to clarify this fact. For example, the consider how we define and describe the variable y_j in Equation 20 in Section 2.5.2:

   where y_j is the number of times under release plan $\boldsymbol{X}$ that the level at evaluation point j is simulated beyond the range of the simulated historical extremes (in Table 1, page 22 of the report), and n is the number of evaluation points (i.e., Lake Superior, Lake MH, …). Similar to base case, y_j corresponds to the risk of failure of plan $\boldsymbol{X}$ at the evaluation point j, $Risk_j(\boldsymbol{X})$, such that $Risk_j(\boldsymbol{X}) = y_j/T$ where T is the total number of time steps in simulation.

   Simply minimizing the average risk of any new plan will not achieve the desired objective to improve upon or maintain risk of failures relative to the base case for all evaluation points. For example, an increased risk at one evaluation point relative to the base case is not generally acceptable to the project Study Board even if risk at one or more evaluation points is completely mitigated.

We also clearly state that our objective function is not directly interpretable. It is a proxy to score the quality of solutions. However, we do not think that is a problem – to truly interpret the numerous dimensions of a regulation plans performance, the plan must be evaluated based on the commonly formatted results in Figure 6 and Table 7. Interesting regulation plan performance needs to be evaluated using the validation analyses for example in Section 3.2.3. We respond in the revised report to this comment by stating the following in Section 2.5.2 with respect to Equation 20:

"Note that the objective function value, as formulated above, does not have an immediately interpretable meaning. To fully interpret a given objective function value, it must be disaggregated into its components (i.e., risk values at each evaluation point). The only immediate interpretation is if this value is greater than zero for new release plan ***X***, there is at least one evaluation point at which the risk of plan ***X*** is higher than the risk of base case. On the contrary, if the value is well below zero, then plan ***X*** outperforms the base case at every evaluation point."

Overall, we would say there are many ways to define a risk-based objective function for this huge system. Obviously, different formulation may yield different results and conclusions. In our opinion, the formulation developed here effectively addresses the needs of the study, while it is very simple requiring minimum subjective decisions.

4. **When and where the min-max problem was implemented?** The single-scenario formulation (Equation 20) was used in this study as a proof-of-concept. Then all results and conclusions were generated based on the multi-scenario problem which was actually based on the min-max optimization formulation (i.e., Equation 21). All formulations solved in the results section directly specify what Equation was used as an objective (Equation 21 and 22) and so this should now be clear to readers.

5. **Why computational efficiency was important?**
See response #20 below. In general, we believe the revised report is now presented in a way that does not give an overall view of trying to minimize the time of analysis.

D. Are there any other suggestions that are related to how this analysis may be used more effectively or the results explicated in a more understandable manner?

The authors failed to specify how the 50,000 year stochastic sample is generated and consequently it is unclear how it is representative of changes in extreme conditions. Does it provide variability beyond the historical observed ranges? This is a strong weakness in the report.

6. We did not include the appropriate level of detail in the first report version describing this 50,000-year sample. The revised report discusses much more detail about this 50000 year sequence in Section 2.1 (2nd thru 4th paragraphs). In addition, the end of Section 2.5.3 was modified a little along with the Appendix 2 describing the selected scenarios.

Comments for Transmission to Authors

It would be useful to have both general comments and specific comments for major and minor revision. Please use additional sheets should they be required.

Major Comments:

Pages 9, Several of the simulation components have very sparse textual description. As an example the simulation of Lake Ontario simply states Plan 5800 is used based on a compiled program. This seems insufficient for a reader to understand the simulation

7. It was not within our scope to try and describe the current Plan 58DD as we were provided with a compiled executable program to simulate this plan. Instead, in our revised report we reference readers to the LOSL (2006) study for further description of this Plan 58DD (which the LOSL looked at replacing/revising).

Page 9, The justification and implications of the simplified St. Lawrence system simulation is Jacking in the report. It is unclear how well the empirical equations perform, their biases, and how systematic errors could influence the adaptive management recommendations via optimization.

8. The reviewer raised a good point. We are also concerned about that. Unfortunately, at this point, this is the only model available to us for simulating St. Lawrence system and it was beyond the scope of our work to improve upon or replace the simulation logic provided to us by Environment Canada and previously utilized in the LOSL (2006) study. As a result, we have added the following suggestion to the list of future work tasks in Section 4:

   > "Given the empirical nature of the St. Lawrence simulation equations used in this report, and their assumed decrease in simulation accuracy relative to upstream lake level simulation, future regulation studies beyond the exploratory level requiring simulation of this part of the system should strongly consider re-evaluating how well the St. Lawrence simulation equations predict Montreal

water levels. In particular, this evaluation could be conducted for the years that have passed since the empirical equations were last fitted to the data."

Page 10, There is a disparity of time scales in the simulation models (monthly, weekly, daily). The present report does a poor job of describing the necessary control time scale and the consequences of the selected time steps in its component simulations

9. The reviewer is correct. We revised the report to make sure the time steps used in each of the three simulation models are given in Section 2.1 through 2.3.

Page 11, The numerical approximation in the two point simulation of moving from hourly to daily time steps should be better supported in the authors contention that it has little impact.

10. We have compared the differences of the models with daily and hourly simulation time steps. We have already reported to the study board that the differences are negligible and we feel that adding this analysis to the report is not necessary given we reported the worst case inaccuracies (a few cm).

Section 2.3 highlights concerns on the ability of the simulation to handle extreme outflows particularly for Lake MH, Lake St. Clair, and Lake Erie. A proposed "patch" solution is described where the outflows causing the issue are zeroed in the corresponding time step. This issue was not clearly addressed or described. It has consequences for extreme hydrologic sequences of interest throughout the adaptive management framework. Does this artificial simulation constraint strongly bias management of extremes?

11. The reviewer raised a good point. We were also concerned about that and monitored a number of example optimization runs to observe when this problem occurred. In the early stages of the optimization (with poor plans), we may generate solutions causing that problem. However, after converging to better solutions, such a problem was not observed to occur in the simulation. Any solution which required this patch be invoked during the multi-objective optimization runs was assigned a very high cost value and hence the optimizer steered clear of such solutions. None of our UW plans presented in the analysis utilized this 'patch' when they were simulated and as such there is no bias in the management of extremes. Evaluating the precise role this patch played in biasing the solution trajectory of the optimizer was beyond the scope of our work.

    In response, we clarified in Section 2.3 (last paragraph) in the new report the most important point from above:

    "Note that none of the optimized solutions evaluated later in this report, exhibited behaviour requiring our strategy above to be implemented."

Overall text in Section 2.5, preceding 2.5.1 is insufficient to provide a clear summary of the decision variables in the problem. The decision variables should be defined and summarized with their ranges in a summative table. Perhaps a detailed table could be used in the appendix and a less detailed summary table could be included in the text that explicitly list the typical 10 variables used (where the text highlights the 9 variable instance). This is strongly required to improve the clarity and reproducibility of the study.

12. As requested, we have explicitly defined the decision variables optimized by DDS and PADDS in the revised report. We chose to improve the description of the text in Section 2.4.1 and 2.5.4 to describe the rule curve parameters and procedure more clearly and then added one table in new Appendix 2 to define the corresponding decision variables. Thanks to the reviewer, we are now confident our documentation on this aspect is clear enough that it can be reproduced.

The text support for the objective function in equation (20) is unclear. The 4 priorities claimed in this section do not make sense. There are two priorities that are mixed. Both of the formulations in equation 20 and 21 differentiate instances of violation of extremes by minimizing squared deviations when performance is worse than baseline and maximizing improvements relative to the baseline using the binary Z variable. In (21) the min-max and max-min behavior is used to proxy robustness across tested conditions. It is also unclear how then case mixtures influence optimization. It would be informative to provide cumulative distributions for the typical objective components (i.e., a separate CDF for each term qf equations 20 and 21). A Latin Hypercube sampling of the simulation would provide this and help to support the degree of variability in the components. Also is there a homogeneous impact from deviations? Across locations are deviations from the baseline equally important and with similar economic impacts? How sensitive is the system overall relative to the baseline? It unclear how much the decision variables impact baseline deviations? Should there be physical constraints beyond the simulation?

13. See response #3 above re Equation 20. Combined with the improved description of objective function we believe the four priorities outlined now make sense. They are not meant as strict priorities and they are included to help us justify the logic behind how we constructed the objective function. Note the binary Z variable is not a decision variable - this was unclear in our original report and we clearly state now in the description of Equation 20 that these are not decision variables. The n case mixtures comment is not clear.

14. The idea of generating and investigating CDFs seems interesting. However, we are unclear how it would benefit our final regulation plan and due to time constraints did not perform this analysis.
15. There was no information available to the authors to quantify the economic impacts of deviations from historical simulated levels. As such, we have assumed the impacts for the same deviation are similar. See also response #1.
16. To measure the sensitivity of the deviations to the decision variables require extra work which is currently beyond the scope of the study.
17. For the existing structures we have considered all physical constraints in the simulation code (i.e., they are never violated). For the synthetic structures, no hard physical constraints were assumed and instead they were embedded in the second objective (cost) function. We are not sure what other physical constraints the reviewer is referring to.

In Section 3.1 and overall in the report, it is unclear what purpose or advantage the formulation in equation 20 provides.

18. Please see the responses #1-#4. In brief, the objective function is designed to measure how close a regulation plan is to improving/maintaining the system performance over the base case at all evaluation points.

The first paragraph of Section 3.1 provides important methodological assumptions that would be more appropriate in the Methods. The validity of the underlying assumptions for the Lower St. Lawrence need to be better supported.

19. The results section was re-organized but the general comment of the reviewer still applies to some other sections in Chapter 3. We decided to keep some brief methodological details in the results section since we believe it improved readability.

Page 24, The authors give some specifics on the solution algorithm (DDS) and computational demands in Section 3.I but fail to place their work in the broader context of tools and computational frameworks. It is stated that a 2.7 minute simulation time requires the use of the DDS algorithm without any discussion of other tools that exist. Also it is unclear if the authors have made any effort to free their analysis to broader analysis using high performance computing which is used commonly in major water resources planning efforts by the US Army Corps and other groups (e.g., coastal hurricane analysis). Given the severe costs of the recommendations, why is a minimal analysis justified?

20. There are a wide variety of options (including optimization algorithms) to optimize our system. When it comes to selecting one for a computationally intensive job, three important factors should be taken into account: 1- global optimization capability, 2-

computational efficiency (see response #6), and 3- familiarity of the algorithm. The algorithm we used (DDS) could satisfy all three factors. DDS is an algorithm developed in our research group, and as such, we all have a good insight and understating of what is happening in the course of our experiment, and this significantly helped us generate better results.

As described in Section 1.2 of the report, this work was meant to be an 'exploratory analysis' based on guidance from the IJC. As such, we designed our analysis to simply inform the IJC about possible levels of system improvements and their roughly estimated costs. We did not utilize high performance computing for this analysis – but as the reviewer noted – if we were searching for a solution to actually recommend for implementation (which this study is not intended to do), we would certainly agree that computational efforts should be increased by a few orders of magnitude by utilizing high performance computing resources.

We modified the report at the end of Section 1.2 to say:

> "The report is not intended to recommend a new Great Lakes control strategy for implementation. Instead, this report is designed to inform the IJC about the range of possible water level management improvements and the corresponding estimated costs for two additional control structures on the Great Lakes."

We also added a new bullet about additional analyses that should be considered in any future work:

> "Apply additional computational efforts to improve solution quality through further optimization experiments. Better solutions almost certainly exist but as more and more search is conducted, the probability that a near-optimal solution has been achieved increases. High performance computing facilities should be considered in any future work and should definitely be utilized for any future work that is seeking to recommend a new 3- or 4-point multi-lake regulation strategy for implementation."

Page 24, The authors make reference to "pre*emption" without sufficiently describing the approach for attaining savings.

21. Unfortunately, we felt there was no room in the report to describe how it works and we only referred to the relevant paper. However, we will explain that in future publications on this project.

In Section 3.2 it is unclear how the 50,000 year stochastic sample is generated and consequently it is unclear how it is representative of changes in extreme conditions. Does it provide variability beyond the historical observed ranges? This is a strong weakness in the report.

22. Please see response #6 above.

Overall the report states that an investment of $8.5 Billion will potentially reduce extreme deviations from the baseline scenario. The challenge in these findings is it is unclear what the objective means? What impacts would the variation have? Are there thresholds for damage? Failures? The multiobjective approach is interesting but the problem formulation seems highly limited and distant from stakeholder metrics of concern. To be clear, I am not stating that everything must be in monetary terms but it would have been easier to evaluate the results had the performance objective been formulated in the context of more tmditional vulnembility or reliability measures. 11tere is little physical meaning the in the baseline at present.

23. Please see response #1 - #4 above.

Minor Comments:

Page 7, 50,000 year sequence is awkward phrasing.

We make sure the new report at least says “50,000 year NBS sequence”

Page 7, "...the robustness of the UW plan for water supply sequences the was not optimized for." This phrasing is awkward and confusing.

Statement eliminated in new report.

Page 11, An overview illustration of components, their time scales, and the overall assumptions in the simulations would be beneficial.

We did not add a complete overview of this but we did do something along these lines to better describe the St. Lawrence River simulation (see new map in Figure 1 and new material at end of Appendix 1 for source code for complete documentation).

Page 13, The term quarter monthly seems awkward. Why not use weekly? Weekly is less ambiguous as a time step measure given the non-standard duration of months.

Page 13, How important is the effect of the monthly time step for Lake Superior following Plan 77A on adaptive options?

The quarter monthly time step phrase is used because this is the terminology used in the routing model documentation and in previous IUGLS studies like LOSL (2006).

The question re Superior time step is unclear. We proposed evaluating a shorter regulation time step (quarter monthly) but it was not viewed as a high priority analysis.

Page 14, Figure 1 needs improved labeling on y-axis representation of releases. It may also help to re-label "e" and "f' given that they are not functional slopes. You could improve your notation to have the thresholds more uniquely and explicitly labeled.

We felt like Figure 1 (now Figure 2 in revised report) was clear enough the way it was.

Page 18, 1st text after (19), change "that level" to "the level"

The author in Section 3.1 state six trial optimization runs were used to avoid poor results. It would be beneficial to report the variance in the results attained. Also the quality of the optimization is justified by the failure of the pattern search by Torczon, 1997 to improve results. This seems like a weak baseline justification. The authors could have demonstrated search dynamics for at least one case for a larger number of evaluations and potentially baselined DDS stat of-the-art tools outside of the water literature.

We only wanted to focus on the best solution found over multiple optimization trials and felt like in this report, reporting on variance of optimizer results or convergence behaviour of the optimizer would be distracting from the important results. We did not intend to justify the quality of our optimization with this statement and as such, we deleted the statement from the revised manuscript.

REVIEWER 2 COMMENTS AND RESPONSES:

A. What is the best/most unique part of the analysis?

The main strength of their method is the development of three simulation models used to assess the multi-lake system performance under various regulating rule curves. Extensive simulation analysis and detailed formulations of rule curves for each lake are the best part of this work. Another advantage of their method is due to directly optimizing rule curve coefficients by which they avoid inferring rule curves from a previously found optimal sequence of releases using regression analysis. Furthermore, the method includes sufficient details about the system constraints and flow information to produce practically useful and viable results.

Nice. Thank you.

B. What is the most critical aspect of the study/analysis? Why?

The optimization method uses only 8 scenarios of 70 years from 50,000 available scenarios. It is also not clear how they incorporate those 8 scenarios during the course of optimization. They could use all available scenarios to produce a robust baseline for target releases using a stochastic programming or robust optimization method. Such methods would consider both probability of occurrence and magnitude of extreme flows among all 50000 scenarios and, as a result, would produce more reliable baseline solution.

Please see response to specific comments 1, 2 and 6 below.

C. Which aspect of the analysis/modeling is weakest? Why? How can it be improved?

The method failed to generate a trade-off solution between 29.6 and 8.5 billion clusters of solutions. It would be desirable to find a solution that significantly improves the frequency based objective at some reasonable more cost. The 8.5 billion UW plan is the best solution found here but still not satisfactory in terms of greatly improving the operation over the base case. We would expect much improvement over the base case and almost similar (1%) solution is not good enough since the base case itself is far from ideal and questionable.

Please see response to specific comments 7 and 8 below.

D. Are there any other suggestions that are related to how this analysis may be used more effectively or the results explicated in a more understandable manner?

There is a major question on UW plan and that is: when it is actually violating the base case in any of the lakes or points, how much is the average magnitude of the flow that is violating the extremes? These magnitudes should have been considered in the objective function using a Conditional Value at Risk (CVaR) measure of risk which is popular in Finance. The objective function is connected to decision variables X through Y which is an embedded function defined by UW plan based on simulation. That is why the model becomes an embedded-simulation optimization model. The convergence of solution method for such model is seriously questionable and has to been taken care of by means of a mechanism to ensure convergence. This is lacking and could be the main reason for the algorithm to take so much time and also fail to converge to better expected results.

Please see response to specific comments 2 and 6 below. Note that although we do not assess algorithm convergence as requested by the reviewer, we believe that the results provide the Study Board with a sufficient amount of detail for an 'exploratory level'. More importantly, we believe the results (even if they are not optimal or far from optimal), demonstrate a substantial improvement in system performance and hence provide the Study Board with sufficient information to make a recommendation on future multi-lake regulation studies.

Specific comments numbered 1 – 8 below. For each numbered comment below:

- First pgh in blue text was copied from our original report by the reviewer.
- Any paragraphs in read text are the reviewer comment.
- The last pgh or two in black text is our response to the comment.

1) This study focuses on developing multi-lake regulation strategies to mitigate extreme climates, or more specifically, extreme water supply scenarios.

They did not specifically test extreme climates. The extremes are lost among 50000 scenarios and the deviations averaged over all 50000 scenarios could be misleading.

We did a poor job in the original report describing the 50000 year NBS scenario and how the extreme NBS scenarios were derived from it. The revised report discusses much more detail about this 50000 year sequence in Section 2.1 (2nd thru 4th pghs). In addition, the end of Section 2.5.3 was modified a little along with the Appendix 2 describing the selected scenarios. The most relevant of these changes (last pgh of 2.5.3) in the new report is as follows:

> "A total of eight different NBS scenarios, selected from the 50,000 year LOSL NBS sequence described in Section 2.1, were identified by David Fay of Environment Canada. These scenarios were selected to be diverse in terms of generating a range of high and low lake levels overall, as well as differentially across, the Great Lakes when the base case regulation strategy was simulated. These eight NBS scenarios are described in more detail in Appendix 2. Seven of these scenarios were the most extreme periods out of the 50,000 year NBS sequence with respect to different characteristics of simulated lake levels (one was randomly selected). Therefore, these eight scenarios were thought of as representative of possible extreme future NBS sequences under climate change."

2) The comparison with the base case could also be a problem since the base case itself violates the extremes (bounds) frequently. They use monthly flow information and monthly average lake levels but the target releases found are not monthly. Rather, they have found release targets for only two regulation seasons as mentioned in Section 2.4.2. The results given in Table A4-1 may not be helpful for guiding the actual monthly operation.

Although the base case violates the historical simulated extremes in future extreme NBS scenarios, comparison with the base case is not a problem – we have to compare since we want to improve upon the base case everywhere.

Our original presentation inadequately described our procedure. In the new report, we improved rule curve descriptions in Section 2.4.1 and we revised the report to make sure the time steps used in each of the three simulation models, as well as the regulation time steps, are given in Section 2.1 through 2.3. We have two sets of rule curves for each season and for a *given set* of beginning of period levels and flows, our rule curves generate a target release. Lake Superior uses a monthly regulation time step, while all other control points have a quarter-monthly regulation time step. Our revised introductory paragraph in Section 2.4.2 now makes it clear to readers how target releases are computed and how they vary with season:

> "Two regulation seasons in a year were considered for all control points in rule curve development. This means that we optimized two different sets of rule curves – one for each season. Rule curves therefore generate a target release that is a function of season as well as beginning of regulation time step levels and flows across the system."

3) Overall, the objective function is constructed to try to identify a plan that improves upon the base case regulation performance (reduced frequency of going beyond extremes) at every evaluation location for every NBS scenario considered. Thus, a multi NBS scenario formulation was developed to optimize the rule curve parameters over multiple NBS scenarios with significantly different behaviors simultaneously. The resulting rule curves would be expected to be robust and more reliable when facing unpredictable future climate conditions.

This method does not necessarily create a robust baseline since the probability of occurrence is not considered explicitly in the optimization process. Therefore, the solution found may damage regular operation for the sake of improbable scenarios. An alternative approach would be an advance method to create robust baseline for operating the system using the key idea of uncertainty margin budgeting.

It was not possible to consider alternative approaches for the optimization given the time constraints. However, it is important to point out that:

- Our plans through Equation 21 essentially minimize the worst case probability of extreme lake level violation under extreme climates.
- The reviewer is correct in that we must be sacrificing some performance under regular or non-extreme NBS conditions. We are unable to evaluate this degree of sacrifice.
- Most critically, our validation of plan performances under the full 50,000 yr sequence shows that overall performance (against long periods of normal conditions as well as less frequent extreme conditions) demonstrate that our plans do improve overall performance over the base case and as such, regular operation of the system must also

improve. Tables 8 and 12 summarize the improved overall validation performance (not just in extremes) over the base case.

4) The single objective function to optimize the rule curve parameters for any single NBS scenario. This formulation causes difficulties for the optimizer as it involves binary variables that may not get binary values during the course of optimization but that is not the case here since they used a genetic algorithm. I think they made the problem unconstrained (except for bounds) because they wanted to use Genetic algorithm which has difficulty with constrained problems.

The optimization algorithm used in this study was DDS not a genetic algorithm. Moreover, there are no binary decision variables in the problem formulation. This was unclear in our original report and we clearly state now in the description of Equation 20 that these are not decision variables.

5) The authors claim that: initial results showed that pattern search could not noticeably enhance the quality of DDS solutions suggesting that the solutions found with DDS were all very close to local minima.

It is not known how good the found solution is in comparison with other possibly optimal solution since nothing is reported on the convergence of optimization method and the solution found is only compared with the base case which is itself questionable.

We eliminated the statements comparing DDS to pattern search in the revised report. We also explicitly state now that it is not clear how good the solutions are relative to the better but unknown solutions. See paragraph 1 in Section 3.4 which says:

> "However, given the number of decision variables (78) and non-linearity of this problem, it is possible that our solutions are poor approximates of the global minimum or the true trade-off and thus there is the potential that there are solutions that would improve performance everywhere in the system for all NBS scenarios even with Montreal."

We did not choose to report on the convergence of DDS in the report. One reason is that convergence behaviour of DDS is conditional to the user specified computational budget (see Tolson and Shoemaker, 2007).

6) Validation performance of UW plan and Base Case over the full 50,000 year.

There is a major question on UW plan and that is: when it is actually violating the base case in any of the lakes or points, how much is the average magnitude of the flow that is violating the extremes? These magnitudes should have been considered in the objective function.

In brief, the analysis we did was based on a reliability-based optimization formulation which was tested later in terms of vulnerability (i.e., the magnitude of violations) in validation. All the histogram plots of violation magnitudes in the text (Figures 7, 8, 13, 18, 19, and 20) and supporting discussions aimed to evaluate the system in terms of vulnerability. These results show that despite not including magnitudes in the objective function, the magnitudes violating the extremes were generally reduced by our UW plans.

7) There are two distinct clusters of solutions on the trade-off separated by a very large difference in regulation costs (a difference of approximately $20 billion). It is not currently clear why there is such a distinct difference in cost or whether this difference actually exists in the true set of trade-off solutions (remember, PA-DDS is heuristic and can only be expected to approximate the true trade-off like all other applicable multi-objective optimization algorithms).

This shows that two sets of single objective optimal solutions have been found and a trade-off solution between the two clusters have not been established. The method failed to generate a trade-off solution between 29.6 and 8.5 billion. It would be desired to find a solution that significantly improves the frequency based objective at some reasonable more cost. The 8.5 billion dollar solution is the best solution found here but still not satisfactory in terms of greatly improving the operation over the base case. We would expect much improvement over the base case and almost similar (1%) solution is not good enough.

First of all, the multi-objective results were updated in the new report based on further optimization results. However, the reviewers comment is still valid given the spread of solutions still evident in new Figure 14. To address this comment, we added the following discussion to Section 3.3.1 to help describe why such a gap in the trade-off could exist:

> "The trade-off in Figure 14 shows a substantial discontinuity at a frequency based objective of -20. There are at least two potential causes of this discontinuity. First of all, the PADDS algorithm only approximates the true set of trade-off solutions and there is no guarantee the algorithm will find all trade-off solutions (in particular with such a high dimensional multi-objective optimization problem). As such, many solutions could exist that span this discontinuity and simply have not been discovered by the algorithm. In addition, the nature of the optimization problem could be the second cause of this discontinuity. The extreme increase in costs is driven by the need to excavate drastically more material in order to pass required flows down the St. Clair River. Compare the maximum increases in St. Clair River flow over the current conveyance regime for the green (1751 m^3/s) and blue (6520 m^3/s) solutions. Note the maximum increases on the Niagara are much closer (5467 m^3/s versus 6520 m^3/s). The form of the frequency based objective function (Equation 21), which focuses on improving the worst performing NBS scenario relative to the base case, is important in understanding the discontinuity. Equation 21 implies that for a given lake, performance is measured/controlled by performance in only the critical NBS scenario and the non-critical performance in the other seven NBS scenarios is ignored. Thus, as the frequency based objective is improved, there will be points (i.e. the discontinuity) at which improving the objective can only accomplished by improving performance in a new

critical NBS scenario. Improved performance in such a new critical NBS scenario could require drastically more dredging (large increase in costs) before a corresponding improvement in the frequency-based performance measure is observed."

8) The comparison of the various solutions in Table 7 and Table 8 (and Table 4) highlights that a variety of reasonable quality solutions (multi-lake regulation plans) exist for a range of estimated costs.

This is exactly what is missing in the report. We do not see a variety of tradeoff solutions between the two clusters of (single objective) solutions found therein

See the response to comment 7 above. The new Figure 14 shows 'n' increase in variety of trade-off solutions relative to the original report. The results are sufficient to inform decision makers about the range of performance and costs possible with 4-point multi-lake regulation.