Response to reviewer #3 (Kevin Turner) comments on "Use of water isotopes and chemistry to infer the type and degree of exchange between groundwater and lakes in an esker complex of northeastern Ontario, Canada" by Boreux et *al*.

In black: reviewer's comments.
In blue: our answers and/or what we added or changed in the manuscript.

We thank Kevin Turner for his comments which aided in improving a new version of the manuscript.

General comments on the manuscript:

It is evident that the authors maximized their use of conservative and non-conservative tracers to effectively characterize lake typology. They carefully considered most physical variables influencing their results and their data shows strong distinction among groupings, as backed by their statistical analyses. This paper provides important data that characterizes hydrological conditions in an area that is expected to experience more landscape (development) changes. The typology of lakes provides an important baseline for comparison to future hydrological regimes that may be altered. I am a fan of the study site description. The field measurements and water sample collection provide important details for other researchers to consider when doing this type of analysis.

251 – I'm a bit confused about their assumption that the lakes are at isotopic steady state when their data shows that recharge lakes have EI > 1.

Indeed, a significant proportion of the lakes in the study area are small closed basin lakes that receive no direct surface inflow. As such, the assumption of steady state (*i.e.* undergoing evaporation while maintaining constant volume) may not be valid for all the lakes. We thus calculated the evaporative loss fraction of the pool volume for recharge lakes and kept the E/I ratios for discharge and seepage lakes as those likely receive significant groundwater flow to satisfy the steady state assumption. E/I ratios have been used in small lakes with no direct surface inflow but significant groundwater input in similar settings (e.g. Arnoux et al., 2017a; Arnoux et al., 2017b). Details of those calculations have been added to the method section.

350 – Is the lower DOC because of dilution over a greater water volume? Would be interesting to know if catchment land cover could explain some of that difference. I see they get at that in the discussion at 453 with catch area:lake area and again at 605, but land cover has been left out. That's fine, but the possibility that land cover influences non-conservative tracers should be mentioned.

DOC tends to be lower in deeper lakes, likely due to differences in residence times and/or mixing rates. The reviewer is absolutely correct in stating that land cover affects lake trophic status and water chemistry. Many previous studies have demonstrated this. However, as mentioned in the study site description and the discussion, the study area is almost completely covered with boreal forest. As such, our study area does not offer the possibility to assess the influence of land cover on lake water tracers. We also discussed the influence of other factors as suggested by the other reviewer.

Minor edits:

95 – Strange end to the sentence.

The sentence was reworded as follows: "Nonetheless, studies that have combined chemical and isotopic approaches to investigate the connectivity between groundwater and lake water at the landscape level and for a large cluster of lakes are lacking".

164 – Change 'digitalized' to digitized. What imagery was being used in Google Earth? That is what should be mentioned.

We changed 'digitalized' to 'digitized' as suggested. The imagery date (7/26/2005) was mentioned.

171-173 – Sentence should be cleaned up.

The sentence was rephrased as follows: "Since all lakes in the study area are kettle lakes, which are characterized by steep slopes on their shore over a small distance, buffer zone of different widths were produced. The buffer width of 100 m was chosen as this distance showed the best correlation with water tracers."

Fig 2b – Typo – change to 'Local Evaporation Line'

This was adapted in the text as suggested.

213 – I assume the isotope work was done in their own lab since no other one is mentioned.

The name of the lab where the isotopes were processed (FaBRECC lab at Queen's University) was mentioned as suggested.

Table 1 – Sort rows in order of nutrients, ions, isotopes. Caption should just say that 'lower and upper elevation ranges represent the standard deviation'

Rows were sorted and the caption was adapted as suggested.

425 – ‘But those are for the most…’??
We replaced “for the most” by “mainly”.

483 – Use different choice of word/phrase for ‘supposed to be’

We deleted “supposed to be” as suggested.

634 – While it was noted that the recharge lakes are more susceptible to evaporative-drawdown during dry conditions, it could also be noted that discharge lakes may be more susceptible to contamination as development encroaches into the source water locations. The point could also be made around 669.

The point was made and developed around line 669 as suggested. We did not mention this in line 634 as we wanted to keep this paragraph solely on hydroclimatic conditions.

683 – Could be mentioned that paleo work could provide a reference for evaluating whether present hydrological conditions are within the range of natural variability. Furthermore this would significantly complement their baseline knowledge of hydrological conditions as development continues in the area.

We mentioned that the typology of lakes provides an important baseline for comparison to future hydrological regimes that may alter them as suggested. This was also stated in the introduction.

Notes on previous reviewer comments:

The previous reviewer had many useful comments for the authors to consider, and overall the authors responded with the necessary revisions. I agree with the authors’ responses where the reviewer comments questioned the utility of their approach. In particular, the reviewers comment about the authors’ ‘indirect’ evidence (chemistry and isotopes) of findings suggests his/her lack of confidence in the approach despite the clear evidence presented in the paper. As the authors note, the resources required to make the necessary direct measurements would be immense, but are clearly detectable using more feasible and sustainable approaches that can be applied at greater spatial scales. This point could even be showcased more in the paper.

We totally agree with the reviewer and we added the following sentence in the introduction: “The use of water tracers was preferred to direct measurements because tracers (1) have proven to be good indicators of groundwater-lake water interactions and (2) constitute a time and cost-effective approach that can be applied at a greater spatial scale for a given time.”

Response to reviewer #4 comments on "Use of water isotopes and chemistry to infer the type and degree of exchange between groundwater and lakes in an esker complex of northeastern Ontario, Canada" by Boreux et *al*.

In black: reviewer's comments.
In blue: our answers and/or what we added or changed in the manuscript.

This study examines spatial variability in conservative and non-conservative tracers across 50 "kettle lakes" in northern Ontario. Authors employ a combination of correlation, ordination, and simple water balance analyses to develop a "lake typology." The authors argue that landscape position is dominant driver of lake hydrology, suggesting lakes at higher elevations are "recharge" lakes (ie they contribute to local groundwater aquifer) and lakes at lower elevation are "discharge" lakes (ie they receive water from local groundwater). Finally, authors suggest recharge lakes are more sensitive to short term changes in hydroclimatic conditions than discharge lakes.

While this manuscript is in revision, I am reviewing this manuscript for the first time. In general, I found the manuscript quite interesting and worthy of publication. However, unfortunately, I believe both the narrative structure and analyses should be developed further before publication. In particular, more information about analyses is needed in the methods section, results associated with landscape position/morphology are overstated, and conclusions about inter-annual variability are confusing/unconvincing.

Below I provide both general and specific comments in an effort to help the authors improve their manuscript.

We thank the reviewer for his/her thorough comments. These comments helped us to improve the manuscript. We followed the majority of the reviewer's comments and suggestions when we prepared the second revised version of this manuscript.

General Comments:

1) The authors should work to further develop their narrative. In particular, authors should work to make their main points more easily accessible to readers. One approach is to structure the paper to target three different types of readers: (1) readers who will only review the abstract/conclusion/figures, (2) readers who will lightly skim discussion, and finally, (3) readers who will thoroughly read the manuscript. Because the vast majority of readers will fall into the first category, it is imperative that authors tell a coherent story with the abstract/conclusions/figures.

The abstract was adjusted to mirror the manuscript, a new section in the method was added, more results on the relation between water tracers and lake morphology was added in the result section, and efforts were made to highlight the rationale and the implication of this study in the introduction and discussion.

2) The abstract should be streamlined. (Refer to comment #1 above.) The current abstract reads like a list of facts, and provides minimal synthesis of those facts. While this is a stylistic decision, I prefer abstracts that mirror the structure of the paper [eg intro, broad objective statement, study design, 2-3 major results, and a brief conclusion that links to objective statement]. This helps the reader extract needed information, and for readers who are taking a deeper dive, it gives them a road map to refer back to if needed. Also, in general, the abstract should be constrained to one paragraph.

The abstract was constrained into one paragraph and modified according to the reviewer's comment to mirror the manuscript more.

3) The methods are incomplete. There is no description of the ordination analysis, comparative statistics, correlation, or breakpoint analyses. While it is often acceptable to omit simple analyses from a methods section, I do not believe it is appropriate in this case. Further, this will help guide readers as they pick through the results section.

We added a section in the methods named "Statistical analysis" as suggested as follows: "Linear regressions were used to assess the degree of co-variability between quantitative variables while logistic regressions were utilized to assess the relations between binary variables and quantitative variables at the 0.05 level. Breakpoint analysis or segmented regression was used to detect any change of trends in water tracers along an elevation gradient and to produce subsequent higher level groupings of lakes. Breakpoints that were significant at the 0.05 were averaged to obtain the elevation of the "breakpoint line". A non-metric multidimensional scaling (NMDS) was run to assess the differences among lake types in a 2 dimensions ordination space using non-scaled values of electrical conductance, Ca, $\delta^{18}O$ and $\delta^{2}H$ as input variables and Euclidean distance as dissimilarity measure; no rotation was applied. A Wilcoxon signed-rank test was subsequently applied as a post-hoc analysis for all lake-water variables that were above detection limits to determine if differences among the different types of lakes were statistically significant at the 0.05 level as most of the Shapiro-Wilk test for normality revealed that most variables were not normally distributed. An analysis of similarity (ANOSIM) was also carried out as a complement to determine if within group similarity was significantly greater than in-between group similarity at the 0.05 level. All statistical analysis were performed in *R* on the data from the August 2014 campaign as it was the one with the most samples".

4) The description of the sampling design is ambiguous. For the synoptic sampling, 50 lakes were sampled across three sampling campaigns. Were all lakes sampled during each campaign, or were 50 total lakes sampled and only some sampled multiple times?

We agree with the reviewer that the sampling design should be more clearly described. We reworded the sentence at line 178-179 as follows: "50 lakes were sampled (29, 28 and 50 lakes during the June 2013, June 2014 and August 2014 campaign respectively), as well as a number of streams (lake outlets and lake inlets) and groundwater springs".

5) Authors should discuss the limitations of the temporal component of their sampling design. Throughout the document, the authors suggest the groundwater and precipitation observations are too similar to differentiate. I'm not sure if I agree with this statement, as it vastly oversimplifies associated isotopic fractionation and solute mixing processes. (Specifically for water isotopes, see recent review paper by Sprenger et al., 2016 in Reviews of Geophysics. Note, the detail presented in the Sprenger review is admittedly beyond the scope of this study.) While I think the sampling design is adequate for the given inference, it may be worth discussing the benefits of greater temporal resolution sampling.

We touched on the resemblance between the isotopic composition of precipitation and groundwater as suggested by one previous reviewer. Indeed, mean precipitation and groundwater often display similar isotopic signatures as they both tend to retain their original isotopic composition because they undergo little to no evaporation. This is shown by the data: groundwater springs have isotopic values similar to mean annual precipitation, respectively -14.7‰ to -13.1‰ for $\delta^{18}O$ and 105.5‰ to -96.0‰ for $\delta^{2}H$. One of the previous reviewer was concerned that the spatial extent of the study area would be large enough to have significant differences in precipitation (and their isotopic signature), which could introduce a bias in the interpretation. We do agree with the reviewer that the isotopic signature of precipitation (Fig. 2a) and groundwater vary on an annual basis and although they have similar values, those are different: groundwater values were more depleted than the ones of precipitations during the sampling campaigns that took place in the growing season (and those would be likely more enriched during the winter as snow is more depleted). We specified this briefly in our results section.

The authors acknowledge the benefits of having a greater temporal resolution of sampling. When comparing the variability between sampling campaigns, we highlighted briefly the fact that a greater temporal resolution of sampling would reinforce positively our interpretations in the discussion.

6) I am concerned about the assumption dV=0 in the water balance analysis. In large lakes with consistent inflow and outflow channels, this assumption is likely appropriate. However, in small lakes without surface connections, this assumption is absolutely not valid. [As noted in the authors

discussion, ephemeral lakes are actually defined by their seasonal variation in volume.] It appears the lakes sampled in this study are in-between these two endmembers. Therefore, the authors should caution readers about the assumption dV=0, and further, cite other papers that use a similar approach to develop I/E ratios to justify their methods.

As for the two to three "ephemeral lakes" in our study area those correspond to dry kettles as specified in the manuscript and those are not considered lakes per se or even ponds. Those areas refer to small kettles depressions that become muddy during moist conditions. But, indeed, their presence is indicative of the variation of the water table in the study area and changes in volumes of lakes deprived of surface connections. As such, we agree with the reviewer than the hypothesis of steady state may not hold for specific lakes. As mentioned to the other reviewer, we thus calculated the evaporative loss fraction of the pool volume for recharge lakes and kept the E/I ratios for discharge and seepage lakes as those likely receive significant groundwater flow to satisfy the steady state assumption. E/I ratios have been used in small lakes with no direct surface inflow but significant groundwater input in similar settings (e.g. Arnoux et al., 2017a; Arnoux et al., 2017b). Details of those calculations have been added to the method section.

7) I encourage the authors to use non-parametric statistics. (eg Use a Wilcoxon rank-sum test instead of an ANOVA.) I would wager that their data likely violate normality assumptions.

Some of our variables (but not all) indeed do fail the normality test (tested with a Shapiro-Wilk test), although most of the variables that do fail the normality test are near normal as evidenced by QQ plots. In order to improve the statistical analysis of the manuscript we switched our post-hoc analysis from an ANOVA to a Wilcoxon signed-rank test as suggested by the reviewer. Tab A2 and A3 were adjusted in the appendix. The outcome of the test was very similar to the ANOVA performed (except for a limited amount of variables).

8) I would like more information about the NMDS analysis.

In the methods section, the following information should be provided [at a minimum]: What was the input data? [Was it from across all 50 sites and all three synoptic sampling events?] Was the data scaled before running the analysis? What was the dissimilatory measure? How many iterations were used to fit the model? Were the final results rotated so the dominant gradient varies along the primary axis? What statistical package did the authors use (R, SAS, PC-ORD?).

In the results section: How many axes were utilized, and why was that decision made? Did the model converge, and if so, after how many iterations? What was the resulting stress?

> What was the input data? [Was it from across all 50 sites and all three synoptic sampling events?] All 50 lakes from the August 2014 campaign as it was the one with the most samples.

> Was the data scaled before running the analysis?
No scaling was applied to the data prior to the calculation of the dissimilarity measure (we tested the raw data and the data transformed with a z-standardisation. The latter gave different coordinates values but the same relative distances in ordination space and the exact same polygons as displayed in Fig. 6).

> What was the dissimilatory measure?
Euclidean distance

> How many iterations were used to fit the model?
35 iterations.

> Were the final results rotated so the dominant gradient varies along the primary axis?
No rotation was used.

> What statistical package did the authors use (R, SAS, PC-ORD?).
The software R was used.

> How many axes were utilized, and why was that decision made?
Two axes were utilized to visualize the data in two dimensions.

> Did the model converge, and if so, after how many iterations?
The model converged after 35 iterations.

> What was the resulting stress?
Stress = 0.023

A summary of those details was added in the method section as suggested.

9) I appreciate the authors attempt to classify lakes in Figure 7. This is one of the more compelling parts of the manuscript. To further clarify the discussion surrounding the differences between higher level groups, I encourage the authors to add an additional plot that displays differences between the higher level groups. Personally, I love boxplots because of the amount of information they provide! Also, authors may consider showing differences across groups in either ordination or dataspace. (A multivariate, nonparametric test like MRPP or NP-MANOVA may be appropriate.) Further, cluster or CART analyses could be used to differentiate groups. However, just to be clear, the listed analyses are simply suggestions for improvement, and authors should not feel obligated to use them!

We really appreciate the reviewer suggestions, especially for guiding us to specific statistical analysis. An ANOVA (now changed to a Wilcoxon signed-rank test) and NMDS analysis were processed as a post-hoc analysis: the first one in the appendix and the second one as a plot and a commentary in the discussion. While we totally agree with the reviewer that boxplot are effective in showing differences among groups, a lot of figures would need to be included while a Wilcoxon signed-rank test table show those differences in a single table that is more synthetic. The NMDS showed the differences among groups for the upper two higher level groups. Plus, an ANOSIM was also done as a complementary analysis of the NMDS. Post-hoc analysis used in this study all converged to the same conclusions. As such, we think including more post-hoc analysis would be overwhelming for the reader.

As for additional clustering analysis, clusters were made using different dissimilarity measures and those provided similar groups as the method we used (while sometimes slightly different groups). Following this, we thought that including another method of classification would be redundant and confusing for the reader.

10) The authors' conclusions about landscape vs local drivers of hydrology are likely overstated given the sampling design. The authors develop a compelling argument that landscape position (and associated geologic setting) drive lake hydrology. However, their analysis of hydrogeomorphic features at each site is very limited due to data availability. Important site-specific variables like network order (ie how many lakes drain into the lake in question), contributing watershed area, and soil characteristics like specific yield can all be used to explain observed variability in hydrologic data. While it is understandable/acceptable that authors did not collect these variables, I think they are over reaching by saying landscape position is more important than local hydrogeomorphic characteristics. I would encourage the authors to add a caveat to this statement, highlighting measures that could be used to better explore local geomorphic drivers.

As suggested by the reviewer (and the other reviewer as well), we developed this section further and we listed other factors that may explain the observed variability in the hydrological data. We also added some correlation results in the result section to better link the result and discussion section as suggested earlier in the reviewer's general comments.

11) The results presented in figure 8 are quite confusing. The discussion should restructured in such a way that clearly delineates annual and inter-annual variability in ET and E/I ratios. As presented, the conclusions (L708-712) that recharge wetlands are more sensitive to climate extremes that discharge are unconvincing. I would encourage authors to present a conceptual model of annual variability for both recharge and discharge wetlands, highlighting how results presented in Figure 8 support the proposed model. Also, consider visualizing this information using another format.

We thank the reviewer for his/her input on this section. Our goal was not to discuss annual or inter-annual variability but observed differences between sampling campaigns, which were characterized by slightly different hydroclimatic conditions. Evaporation rates should be relatively similar among lakes as our study area consist basically of a rectangular zone of 12 km by 6 km. E/I ratios differ among lakes as groundwater inflow (I) can buffer evaporative losses (E). ET (evapotranspiration) could not be computed using the Penman formula as no solar radiation data was available for the period of interest. Our real goal was to relate observations in changes in E/I ratios and EC along a groundwater gradient between our sampling campaigns and how it relates to the existing literature. As for the format, bar plots seem to be the most appropriate format as we compare different groups of lakes while tracking changes in time.

But we do agree with the reviewer that a conceptual model combined with the existing data from Fig.8 would produce interesting synergies and further support our observations. While the literature and our general knowledge of our study area could allow us to produce a conceptual model displaying the annual and inter-annual variability, such a model would remain hypothetical as no continuous significant data set could back up such a model. As a result, we decided to adapt the conceptual model developed by Webster et *al*. (1996) in similar settings. The conceptual model of Fig 8a describes the relationship between the direction and magnitude of lake E/I and EC changes during drier conditions to lake type defined by the degree to which lakes interact with groundwater. We think that this type of conceptual model is more appropriate to support the observations we comment in this section of the manuscript.

Besides developing a conceptual model, (1) we acknowledged that some of our observations (such as changes in EC between discharge and seepage lakes) were not clear and used the conceptual model to explain trends; (2) we also cautioned the reader that it is difficult to draw conclusions on the inter campaign variability based on the limited sampling frequency; and (3) we deleted the mention of extreme conditions as we do not have such data.

12) Finally, I would encourage authors to explicitly link their results to management activities in northern Ontario. What information from this study will be useful for managers? Maybe even write this section for them. Authors begin to do this when describing mining activities. However, there is no discussion how the “cottage development” industry could use the information derived from this study.

We agree with the reviewer that the implications of our results for management purposes would benefit from further development. This point was also brought up by the other reviewer. We explained a little bit further in the paragraph how mining activities and cottage development can impact groundwater-dependent systems in the area.

Specific line-by-line comments:

L33 Abstracts should be 1 paragraph.

The abstract was restructured into one single paragraph and was revised based on comment 2) of the reviewer.

L72 Typically, anions such as Cl and Br are considered conservative tracers. While they may not exhibit conservative behavior in all settings, they are much less reactive than tracers like DOM, N, and P.

We totally agree with the reviewer that Cl and Br anions are generally viewed as conservative tracers (especially with respect to NPOC, N or P) but we defined those as non-conservative in comparison to water stable isotopes. Also, Cl and Br don't always behave conservatively as they can be affected by residence times.

L102 This is an odd place for a comma.

The coma was likely a typo and was deleted. Thanks for pointing this out.

L103 Lake typology is a new term for me, and potentially for other readers as well. Maybe consider defining it here?

A definition of lake typology was added in the introduction as suggested and we better developed the rationale for it.

L130 Please rephrase. "The esker are" is somewhat ambiguous, especially since the term eskers has not been defined.

We agree with the reviewer that the sentence was ambiguous. The sentence at line 130 was deleted and the term esker was defined as follows a little bit further in the text: "long sinuous ridges of coarse grained glaciofluvial sediments in deposits oriented in a north-south direction".

L164. How were features digitalized in ArcGIS 10.3 from google earth? I think I know what you mean, but please be more explicit here.

We agree with the reviewer that the sentence was confusing. It was reworded as follows as suggested by the other reviewer: "Lakes and other geographic features were digitized from Google Earth using the imagery dating from 7/26/2005".

L171-173 I'm not sure what this is referring too. Please rephrase to be more clear.

The sentence was rephrased as follows: "Since all lakes in the study area are kettle lakes, which are characterized by steep slopes on their shore over a small distance, buffer zone of different widths were produced. The buffer width of 100 m was chosen as this distance showed the best correlation with water tracers."

L233 How far away were these samples taken from the study area. Note, it's acceptable that precipitation samples were taken "off-site" given the level of detail/inference in this paper. However, this is still an important detail.

The distance and the orientation from the station was referenced according to the closest city (Timmins). We adapted the distance the orientation (*i.e.* 125 km NW) in the text as suggested.

L251 Please define hydrologic and isotopic steady state, and then use previous studies to confirm the validity (and limitations) of those assumptions.

We added a definition and cited a couple of studies that used the Craig-Gordon model in similar settings as suggested above by the reviewer (point 6).

L326 Please add a parenthetical designation of "E/I" ratio

This was adapted in the text as suggested.

L361 Please describe the breakpoint analysis in more detail. This detail should likely go in the methods section.

Breakpoint analysis was used to detect when there was a change in a trend in the data. In our case, we have a gradient of lakes ranked according to elevation (thus a continuous series of data just like a time series) and we observed that lakes behaved isotopically and chemically differently in two groups according to elevation. A breakpoint analysis was done and indeed revealed that the breakpoint was significant for most variables at around the same elevation. This is particularly relevant because the breakpoint line allows us to map what we believe to be the groundwater recharge and discharge zones. A brief sentence was added in the method section.

L522 Again, please describe logistic regression in the methods section.

A brief sentence was added in the method section.

Figure 1. Maybe add a line to map to indicate the potential location of conceptualized cross section presented in b

We agree with the reviewer that indicating the potential location of the conceptualized cross section could be useful to the reader. However, in order not to overcrowd the map, we specified that thus could be located at latitude 48°35'0"N in the figure caption (which is displayed on the figure on the left and right).

Figure 4 and 5. Authors should consider flipping their axis.

While we do agree with the reviewer that elevation should be displaying in the x-axis as we relate the isotopic and chemical signature of the lakes to elevation, displaying elevation in the y-axis allows the reader to see the abrupt transition that occur near the breakpoint. Elevation is typically displayed on the y-axis when looking at variables along an elevation gradient.

Figure 6. How were these grouping boundaries defined? Are you connecting points at the fringes, or are these polygons the result of an analysis?

Those boundaries consist of the connection of points at the fringes of the groupings defined by the lake typology.