Author response to Anonymous Referee #3: Our replies to referee comments (*black italics*) are provided below in blue.

Anonymous Referee #3 comments:

Summary of the paper:

In this study, multiple tracers were used to identify dominant runoff generation mechanisms over two hydrologic years in Star-east and Star-west watersheds. Principal component analysis was used to reduce the complexity that may arise by analyzing every tracer combination. The study concluded that streamflow during early melt was dominated by hillslope groundwater. As snowmelt peaked, the entire landscape became connected and all the water sources contributed to streamflow (the proportion from different sources is not computed). During the Fall season, hillslope and bedrock groundwater became the major sources of streamflow in Star West watershed (proportions not computed), however the sources were unresolved in the Star East watershed. The au
thors then went on to conclude the subsurface flow pathways in this region are complex and this complexity along with slow release of groundwater from glacial till ensures hydrologic resilience in this region.
This study tries to resolve the seasonal sources of streamflow which is a very interesting research topic and definitely fit for this journal. However, quantitative estimates of source proportions are missing from this study which is possible to compute given the number of tracer variables that were monitored.

Reply: We thank the referee for their review of this draft manuscript.

Major comments:

1. The abstract and introduction talks in detail about the concept of hydrologic resilience, however I do not find any attempt to quantify this statistic in the remainder of this article (except a very brief discussion on recession rates at the end). I will recommend either quantifying resilience or removing it (at least from the abstract).

Reply: All sections relating to hydrologic resilience will be removed.

2. The source apportionment which is the key focus of this study was done qualitatively because TVR was below 2. A TVR value below 2 signifies that sources are not completely differentiable. In such cases, the uncertainty in the contribution of different sources is higher, which does not mean that an EMMA is useless. I will encourage the authors to undertake a simple EMMA and report the results for the same. An easy way to do this will be using one anion and one cation (reason in #3 below) and some variant of an EMMA. On the point of violation of assumptions, instead of a conventional EMMA, a Bayesian mixing model can be used where the error distribution can be parameterized and later verified.

Reply: The mixing model portion of EMMA was not run because of the violation of key assumptions, the large variability in source water, and Star East stream water not being bound by its sources. The seasonal variation in stream water and large overall variation in source water added uncertainty to mixing results; median or mean values of source water would not have physical meaning during a given season or month. Additionally, while not mentioned in the draft manuscript, small numbers of samples per source can also add uncertainty to mixing proportions. Small et al. (2002) suggested that greater than 20 samples per source are required to reduce this uncertainty; however, in many cases, we have far fewer than 20 samples. Due to the combination of the factors above, the error associated with the un-mixing model would be very large and results would not be particularly meaningful. We decided that a qualitative description of these data displayed in a PCA would still provide insight into the hydrological processes in our study region because the principal components (PC1/PC2) were created from the variability across multiple tracers.
It is important to note that it has been shown that less accurate predicted mixing proportions can arise from reducing the number of tracers used in the un-mixing model (Barthold et al., 2011). While others have used two tracers historically, close scrutiny of predicted portions using known mixtures have shown that larger number of tracers generate more accurate results (Collins et al., 2017; Sherriff et al., 2015). Thus, undertaking a simple EMMA or a Bayesian mixing model with 2 tracers as suggested would have the same problems regarding source water variation and large uncertainty in predicted proportions compounded with overall less accurate mixing proportions. The importance of testing mixing model

predictions using mixtures, rather than goodness-of-fit tests for the prediction of measured tracer values in mixed waters, as was conventional for many years, has been critical in revealing the dangers of using overly reductionist signatures.

3. On visual inspection of Figure 4, it seems that Cl^- is markedly different from the other tracers. Most of the cations are positively correlated and offer complementary information. Is this the case? If yes, why not simply use one anion and one cation instead of doing a principal component analysis using all the tracers. The problem with PCA is that readers do not know which tracers influence PC1/PC2 and to what extent, losing physical significance. This will also ensure that an EMMA model can be setup in a very simple way (using one cation and one anion as the tracers)
Reply: Yes, most ions are positively correlated but Cl^- is also positively correlated except for a few samples that had higher concentrations. SO_4^{2-} better separates the source and stream water samples along a biplot axis and would likely be the better choice if conducting a 2-tracer mixing space/model. However, as explained above, a 2-tracer mixing model would still have large uncertainties associated with the estimated proportions since overly reductionist signatures generate less accurate proportions versus known mixtures (either virtual or actual). The methods used in this study were intended to maximize the statistical information provided by the tracer suite without overstating the results or conclusions that we could make from this dataset.
To address the problem that readers do not know which tracers influence PC1/PC2, a simple option to help clarify the physical significance of the PCA plot is to include a table with the tracers that influence PC1 and PC2. These types of tables are provided often along with PCA plots and can be added here for clarification.

4. Sections 5.2 and 5.3 can be combined into one section, that will make it easier to read the sections and also help avoid repetitions.
Reply: Referee #1 suggested we combine sub-sections within Sections 5.2 and 5.3 to avoid repetitions and we agree that this will help streamline the discussion. We will consider combining the source water contributions and stream water contributions discussion sections as suggested by Referee #3. However, we do not want to lose the ability to stress some key discussion points in source water dynamics. Regardless, we will ensure that the Discussion is less repetitive and more streamlined.

Minor comments:
1. The number of sources are different in different parts of this article (eg: P1L15, P3L67, P5L112, P9L232, etc.). I will recommend using the same number of sources at different instances in the article.
Reply: The number of sources in the identified lists will be revised for consistency.

2. How many of the 11 snowmelt samples came from North York Creek? (P5L124)
Reply: Two of the snowmelt samples came from North York Creek. This will be clarified in the text.

3. Were EC measurements also taken? These can also be used to verify if the seep water is coming from a groundwater pool. (P11L249)
Reply: EC was taken from seeps and the stream but not the till and bedrock wells, hillslope/riparian wells, or suction lysimeters. We agree that EC could be used to verify the source of the seeps and we will explore these data.

4. Water temperature has been discussed at different places in the article, however there are no figures of water temperature in the article. I will recommend to include at least one figure for water temperature.
Reply: We will consider adding a figure for water temperature. Referee #1 also had a similar suggestion about groundwater temperatures. Water temperature of seeps were discrete measurements taken during three flow conditions over two seasons. It is likely that a box and whisker plot will be the best way to display these data because only the till and bedrock wells and stream have continuous temperature data. Water temperature was not measured in hillslope/riparian wells.

5. The reported hillslope groundwater includes riparian water and soil water. How is a riparian zone part of hillslope? (P11L260)

Reply: Hillslope groundwater included riparian water when the chemical signature of riparian water was not statistically different from the chemical signature of soil water. While the processes that occur in the riparian area certainly differ from those on the hillslope, there was not a significant difference between these sources at Star West Lower and Star East Upper so these sources were grouped together.

6. Section 5.3 indicates some kind of a hysteresis pattern in the PC plots of streamflow (anticlockwise direction in Star west (Figures 5, 6) and clockwise direction in Star east (Figures 7, 8)). I will encourage more discussion about the reason behind this.

Reply: The difference in direction between Star West and Star East is likely a product of the PCA process, not due to a difference in hydrological processes, as the tracer's eigen vectors have opposing signs (positive vs negative) in Star West and Star East. More discussion about the difference in direction of the hysteresis loops will be added for clarification.

7. There is no work done on the water age, how have old or new water been defined? (P1L18, P18L418, L420, etc.)

Reply: No work has been conducted on water age. Referee #2 has also suggested we clarify the definition of old water to mean "reacted waters". References to old water will be changed to make it clear that we are talking about water that was already in the watershed prior to snowmelt in contrast to new water such as rain and snow.