HESSD 2019-613 R1

Mehdi Bahrami, 01 Feb 2021

The manuscript is well written and covers all aspects of the study area.

[Authors response]: Thank you.

HESSD 2019-613 R2 Comment

The spatial-temporal variation of streamflow is a relatively common topic in hydrology. This paper has investigated the change of streamflow for the Mackenzie and Nelson River Basins in Canada. While peer studies tend to be based on a small number of stations with a common period of years, this paper makes use of a large number of stations with different series of years by using dynamic time-wrapping. The results show twelve streamflow regime types and also reveal hydrological factors that contribute to the different types. In general, the paper is in a good shape.

[Authors response] We would like to thank the reviewer for taking the time to provide these comments.

There are a few comments for further improvements of the paper:

[1] Firstly, a flowchart that showcases all the datasets and methods is in demand. The paper involves datasets of streamflow and landscape attributes. The different datasets are related to one another through a number of statistical methods. A flowchart would make the methods and results more accessible.

[Authors response 1] The paper is already quite long and a flow chart would not help with that issue.

[2] Secondly, hydrological changes are related to landscape attributes in the analysis. For a given gauge station, its streamflow is affected by landscape attributes of the upstream catchment area. From this perspective, the upstream catchment area ought to be determined for all the gauge stations. If so, please explicitly illustrate this step.

[Authors response 2] This is an interesting suggestion but feel it would be tangential to the main presentation.

The issue of nested basin is noted at line 156. -160. The downstream propagation of the signal is noted at line 487, and the resulting clustering at line 741. We suspect the reviewer is suggesting that some form of differencing between two stations could be adopted to isolate signals to downstream portions of the watersheds.

[3] Thirdly, the dynamic time-wrapping plays a critical role in the analysis. It links a large number of stations with different series of years. It is an existing method. Please illustrate whether the analysis represents a novel application of this method.

[Authors response 3] Dynamic time warping is an important element and while our application is novel, it is not the first use of dtw in hydrology. We have added three references citing earlier work on this topic.

The use of dtw in hydrology to address these issues has been suggested previously (Ehret & Zehe 2011; Ouyand et al. 2010; Mansor et al. 2018). Overall, the use of dynamic time warping overcomes the timing differences due to latitude and elevation.

Ehret, U., and Zehe, E.: Series distance - an intuitive metric to quantify hydrograph similarity in terms of occurence, amplitude and timing of hydrological events, Hydrology and Earth System Sciences, 15, 877-896, 2011.

Mansor, N. S., Ahmad, N., and Heryansyah, A.: Performance of Time-based and Non-time-based Clustering in the Identification of River Discharge Patterns, in: Improving Flood Management, Prediction and Monitoring: Case Studies in Asia, Emerald Publishing Limited, 133-140, 2018.

Ouyang, R., Ren, L., Cheng, W., and Zhou, C.: Similarity search and pattern discovery in hydrological time series data mining, Hydrological Processes, 24, 1198-1210, 2010.

HESSD 2019-613 R3

The authors of this studied the data from several stream gauges in western Canada, analyzing several aspects, like regime types, trends and their relation with trends in vegetation, water and snow availability derived from satellite.

The paper is of clear scientific relevance, the content is interesting and the methods are generally well explained (with maybe a few exceptions, see below).

The main issue that I found is the extent of the manuscript, which is too long (~15000 words) and detailed, in a way that a reader can get lost in many small details and miss the interesting aspects. Starting from the abstract, which is almost 500 words long, and IMO fails to summarize the message of the paper and its scientific relevance. My suggestion is a major reshuffle of the text, making it more compact and streamlined, and concentrating a bit more on the important aspects of the research, possibly moving the details to dedicated paragraphs in the supplementary information.

[Authors response] We appreciate the reviewer having taken the time to provide comments our manuscript.

More detailed comments below-

The abstract is very long, with too many methodological details. A much shorter account of the used methods should be given, while the output and its scientific relevance should be highlighted more.

- l.38: "The overlap between hydrological ..." not clear what you mean here.

[Authors response] "The overlap between hydrological and satellite index trends were not consistent across the study area."

Replace with:

"Hydrological trends and satellite index trends were not consistent across the study area, or sometimes within watersheds."

- l.180: for clarity I would state explicitely that p=0.05 and 5% can be considered linked, as the p-value is a guess on the probability of the null hypothesis.

[Authors response] Replace: “A threshold of 0.05 was used in tests of significance, and 5% was also used as an indicator that the number of trends exceeds the number expected by chance alone. “

With:

“A threshold of 0.05 was used in tests of significance, and accordingly, 5% was also used as an indicator that the number of trends exceeds the number expected by chance alone.”

- l.185: table 1 is rather ugly and not of real scientific content, IMO does not show well in the main manuscript, I would move it to the SI.

[Authors response] “Table 2 provides the starting dates of the five-day period corresponding to each 5-day period number from April through November.”

A previous HESSD reviewer asked for such a table to be added so the reader could connect periods and actual dates. Similar tables have been provided in other papers (e.g. Rickenbach et al. 2020). It needs to be in the main paper as a reader should not have to go to Supplementary Material for this table.

Rickenbach, T. M., Ferreira, R. N., and Wells, H.: Springtime onset of isolated convection precipitation across the southeastern United States: Framework and regional evolution, Monthly Weather Review, 148, 891-906, 2020.

- 190: I find a bit difficult to follow this numbering of the 5-days period, also because sometimes the numbering is shifted (e.g. period 23 becomes period 1). Would it be meaningful to lable the 5-day periods with the starting date? E.g. the period that starts with the 1st of March would be 03-01. This would make the discussion and the figures clearer.

[Authors response] While we partially agree with this comment, the shift that is made (where period 23 of the year become point 1 in a time sequence) is due to software package and the manner in which it frames the input data. The dates associated with periods are shown in all plots so referring to them in the text in the manner suggested seems to make things more, not less, complicated.

- Figure 2 and others (also in the SI): the authors should label the panels with ids, e.g. a, b, c, and in the text reference the panels instead of referring to the "bottom left" panel.

[Authors response] Figure 2, and others that are similar, are single figures with coupled panels, not separate panels. While we understand the convention we have chosen to leave the description as originally stated.

- l.239: ".. unable to group hydrographs when they are not aligned in time" depends on the metric you use in k-means?

[Authors response] "Statistical methods, such as k-means (Likas et al. 2003; Steinley 2006) or self-organized maps (Kohonen & Somervuo 1998; Hewitson & Crane 2002; Kalteh et al. 2008; Céréghino & Park 2009; van Hulle 2012), are unable to group hydrographs when they are not aligned in time (Halverson & Fleming 2015)."

The reviewer is incorrect, the k-means metric is not the issue as k-means performs Euclidean distance from point to point and we have used dtw to cluster the median hydrograph, not k-means.

- l.261: It is not entirely clear what metric was used in the k-means (and subsequently, the mathematical reason of why you do need a DTW). Furthermore, the results of k-means depend on the initial guess of the centroids, the seeds. How did you select them?

[Authors response] We did not use k-means in this section and have stated that in lines 244 to 248:

"The clustering of annual median streamflow time series was done using dynamic time warping, DTW, (Berndt & Clifford 1994; Wang & Gasser 1997; Keogh & Ratanamahatana 2005) which measures similarity between time series that may vary in magnitude and timing by aligning the two standardized (zero mean, unit variance) curves in time, essentially matching the shape of inflections to create clusters (Sarda-Espinosa 2017; Whitfield et al. 2020)."

- l.270: I would rephrasing stating more clearly that a trend was estimated for each of the 5-days time preiods.

[Authors response] *"Trends in the five-day periods for the annual common time window were determined for the period of record of each time series, using Mann-Kendall tests as described above, following the approach of Déry et al. (2009) for examining trend magnitude for a fixed endpoint in time."*

Changed to:

"Trends in each of the five-day periods for the annual common time window were determined for the period of record of each time series, using Mann-Kendall tests as described above, following the approach of Déry et al. (2009) for examining trend magnitude for a fixed endpoint in time."

- l.283: Also here the metric used in the k-means and the seed method should be explained.

[Authors response] *"The individual annual trend scores for the annual common time window for the 395 stations were clustered using the method of k-means, which partitions observations into clusters having similar means and which is well suited to clustering of features such as patterns of significant differences (Likas et al. 2003; Steinley 2006; Agarwal et al. 2016). The number of clusters chosen (six) was based upon the elbow method (Ketchen & Shook 1996; Kodinariya & Makwana 2013); using more than six clusters did not improve the modelling (not shown). "*

- l.289: these figures are in the SI, why pinning them here?

[Authors response] *Sorry if this was confusing. This was simply tracking in the manuscript where each figure and supplementary figure was used. These will be removed in the final version.*

- l.306: capable -> useful

[Authors response] *"Google Earth Engine (GEE) allows for cloud-based planetary scale analysis while it serves as a database for petabytes of open access satellite imagery such as the Landsat archive (Google Earth Engine Team 2017; Gorelick et al. 2017), and is particularly capable for this study. "*

Change to:

Google Earth Engine (GEE) allows for cloud-based planetary scale analysis while it serves as a database for petabytes of open access satellite imagery such as the Landsat archive (Google Earth Engine Team 2017; Gorelick et al. 2017), and is particularly useful for this study.

- l.318 on: landsat composites for NDVI, NDWI, NDSI should already be already available in gee. Did the authors recompute them?

[Authors response] *"For each basin, three time series of spectral index averages were derived from the 16-day mosaics of Landsat 5 TM data."*

We are unsure what the reviewer's point is as this text describes that we simply created time series of basin averages of the indices that are available for each pixel using mosaics we created from Landsat imagery.

- section 3: this section is very long, I believe it would be possible to greately reduce and summarize it, putting the details (e.g. the detailed description of each streamflow regime and its spatial distribution) in the SI.

[Authors response] *We agree that this paper is long, but we disagree with this suggestion. We believe such a move would make the presentation even more difficult to follow, since the reader would have to switch back and forth between the main paper and the supplementary material and that, in our opinion, would be detrimental.*

- l.357: the dashed lines are plotted below the other lines and you cannot see them for many clusters.

[Authors response] *Figure 6 "Each of the twelve plots contains a line for each gauged basin in that Type and the heavy dashed line is the centroid of all members; the colour of the lines is based upon stationID. "*

This is an issue with the originating software that uses ggplot where the centroid is plotted first and over-plotted with the individual lines. We have changed the description to be more precise:

"Each of the twelve plots contains a line for each gauged basin in that Type and the heavy dashed line, where visible is the centroid of all members; the colour of the lines is based upon stationID."

- Figure 8: isn't this figue a repetition of figure 6 (if figure 6 was adjusted to show the centroids, now the dashed line is often hidden).

[Authors response] *Yes, Figure 8 shows only the centroid that is not always visible in Figure 6. This is stated in lines 416:*

"The standardized streamflows plotted in Figure 6 make it difficult to compare the Streamflow Regime Types; plotting the z-score centroids of each (Figure 8) makes the comparisons simpler."

- l.397: this hydrograph, what hydrograph?

[Authors response] "Streamflow Regime Type 5 basins were also common in the Prairie ecozone (n=49), in the Boreal Plains (n=81) as well as along the Mackenzie River to below Great Slave Lake; this hydrograph shows an earlier and briefer peak than Type 1 with a rapid recession (Figure 6)."

Changed to:

"Streamflow Regime Type 5 basins were also common in the Prairie ecozone (n=49), in the Boreal Plains (n=81) as well as along the Mackenzie River to below Great Slave Lake; the hydrograph for this Type shows an earlier and briefer peak than Type 1 with a rapid recession (Figure 6)."

- l.414: the closing parenthesis at the end of line has no opening match.

[Authors response] "While these descriptions are explicit, there was overlapping of types in space (particularly Streamflow Regime Types 2 & 5), and cases where individual basins of a Type occur quite separately from each other (Types 9 & 12) as is evident in Figure 7)."

Changed to:

While these descriptions are explicit, there was overlapping of types in space (particularly Streamflow Regime Types 2 & 5), and cases where individual basins of a Type occur quite separately from each other (Types 9 & 12) as is evident in Figure 7.

- l.446: bsin -> basin

[Authors response] Changed.

- l.477: remind here briefly that the expectation "by chance alone" raises from a choice of p=0.05

[Authors response] "For a period of 10 years, 5% of the cases show significant trends, as would be expected by chance alone."

That choice was stated earlier in the paper, but could be repeated here.

For a period of 10 years, 5% of the cases show significant trends, as would be expected by chance alone (based on a p value of ≤ 0.05).

- paragraph 3.3: the trends in sats indices look weak, furthermore, you did not prove that the trends may be enhanced or hindered by multi decadal oscillation (while for the streamflow trend this has been proved). I believe the discussion comparing with the expected probablistic 5% threshold is acceptable, but this limitation should be mentioned.

[Authors response] "The spatial patterns of trends in the mean values of the three normalized difference indices are presented in Figure 13. The spatial patterns of the trends in the maximum, mean, and minimum of NDVI, NDWI, and NDSI are provided in Figures S22-S24 and are also summarized in comparison with Streamflow Regime Type (Table 7), Trend Pattern (Table 8), and Ecozone (Table 9). The tables show the fractions of stations grouped by trends that were significant at $p \leq 0.05$. In the figures significant trends ($p \leq 0.05$) are shown as red (decreasing) or blue (increasing) triangles, trends whose significance was ≤ 0.10 are shown as red or blue dots, and those with no trend are plotted in black. There was a stronger association of the trends in the three indices with spatial location and with ecozones than with Streamflow Regime type or trend pattern. Frequently, the trends in vegetation, water, and snow satellite indices occur in a spatial domain that follows the margin between two or more ecozones (Figures 13 & S22-S24)."

This is Results section 3.3 and the reviewer's comment is really related to the discussion where a section on PDO and AO in included at lines 969-979.

- section 4: also the discussion is a bit overextended. Furthermore, some concepts that were already mentioned in the results section are repeated here.

[Authors response] Gosh, if we need to remind readers that the 5% is related to a $p \leq 0.05$, then we need to remind the reader of these details.

- l.713: remove one of the 2 "that"

[Authors response] Removed

- conclusions: I would suggest to shorten also the conclusions, providing a more summarized account of what has been done and the originality of the methodology, of the most interesting results, of their relevance for the scientific community and possible future development.

[Authors response] We disagree. The conclusion is a summary of a study of a large spatial domain with many watersheds and multiple facets being considered and several novel techniques. Each of these is summarized as briefly as is reasonable.

Reply to HESS-613 R4

Whitfield et al detail a study that examined trends in streamflow and satellite indices for a large area in central Canada. Many government operated hydrometric stations in this region are only active during warm or open-water season. Because of this, many previous studies looking at long-term flow trends have not used these stations and instead focused on stations operated year round. The authors overcome this issue by using an approach called dynamic time warping to account for differences in timing of streamflow across latitude and elevation. They apply these methods to 395 stations and classified 12 streamflow regime types. They also looked at seasonal trends in streamflow using k-means clustering. They complemented the flow analysis with an analysis of trends in three satellite indices (NDVI, NDWI and NDSI). They suggest that trends in streamflow are organized spatially and influenced by location and ecozone, but not streamflow regime type. They document regional trends and speculate on the underlying hydrologic changes that may be driving those regional trends.

The study is interesting and the methods applied seem useful. My biggest issue with the manuscript is that it is long and feels like it lacks focus at times (I would echo many of the comments of RC3). During my readings of the manuscript, I often felt that content was repeated numerous times (e.g., the conclusions simply repeat some of the discussion), I wasn't sure what the key points of the study were (e.g., was the application of relatively novel methods or was it the hydrologic insight - in the case of the hydrologic insights, I wasn't always sure what I was supposed to take away from the results), and there were often statements made that seemed to have major implications for understanding the study, but were not expanding upon.

[Authors response] We acknowledge that the paper is long, but not that it lacks focus. We sought to be thorough in our descriptions, to ensure the work is reproducible, and to provide the reader with the necessary information to elucidate our results.

It is challenging when a reviewer says that the paper is too long yet asks for additional details and information as below. We have tried to balance these requests in our responses.

To build on this last point, the following are some statements that I wished were developed more:

[Authors response] The reviewer has highlighted a number of statements that we have revisited and revised accordingly.

- "Clearly historical budgetary and management decisions in the Canadian hydrometric program affect any current analysis of hydrological trends and change." (L217-18) How does it affect analysis? How does it affect this study in particular? What are the implications?

[Authors response] The data that were available for this study were not specifically collected to support any specific analysis. What is presented in this paper in an approach

that looks at spatial patterns in streamflow regimes, streamflow trend patterns, and trends in satellite indices based on available data. The paragraph in question starts with: "The stations in this study were operated for differing periods of time and with differing operating schedules."

"Clearly historical budgetary and management decisions in the Canadian hydrometric program affect any current analysis of hydrological trends and change."

Changed to:

The complexity of the dataset results from historical budgetary and management decisions in the Canadian hydrometric program. Assessing hydrological regimes and trends in this dataset requires approaches that are different from "standard" methods.

- "The dataset contains nested basins." (L156) Okay - what should the reader make of this? My first thought is that these stations will exhibit correlation and may be considered a form of pseudo-replication... how does this impact the analysis?

[Authors response] It is important to inform the reader that some stations are nested within this study. The reviewer is correct that this would result in correlation between stations which would be considered a lack of independence. That would be an issue for many statistical tests, but has no negative effect on the analysis which looks at the spatial pattern of groups with similar hydrologic regimes or trend patterns. The nesting results in a signal that propagates downstream along some rivers as we acknowledge at lines 491 and 745. That nested sites are clustered together is logical support of their similarity.

Text has been added to state this explicitly at line 156. "The dataset contains nested basins and may have some correlation; the analysis performed does not require sites to be statistically independent."

- "Determining the magnitudes of trends in annual runoff or other annual attributes using this methodology is not appropriate." (L714-715). Why and what is the implication here?

[Authors response] This originated as a response to a previous reviewer. This is a simple statement that when one compares trends in records from different time periods or with different coverage in a common period of years one should not compare magnitudes.

No change was necessary.

- "While these GRACE trend regions are consistent with the results presented here, the mechanisms that Rodell et al. (2018) suggest are not." (L823-824). This seems to come out of nowhere. What are the mechanisms presented by Rodell et al? How do they compare with the speculations made in this study? I think this is partially attempted in the following paragraphs, but the comparison between this study and Rodell is not clear.

[Authors response] Thank you for this comment. We have rewritten the lines:
"… These four regions correspond to areas with changes in water storage (Rodell et al. 2018); precipitation increases in northern Canada, a progression from a dry to a wet period in the eastern Prairies/Great Plains, a region of surface water drying in the eastern Boreal, and of no change along the Rocky Mountains. While these GRACE trend regions are consistent with the results presented here, the mechanisms that Rodell et al. (2018) suggest are not.

And replaced them with:
"… These four regions correspond to areas with changes in water storage reported in Rodell et al. (2018). These four GRACE trend regions are consistent with the results presented here. Rodell et al. (2018) suggested the regions result from the following mechanisms: [1] precipitation increases in northern Canada; [2] a progression from a dry to a wet period in the eastern Prairies/Great Plains; [3] a region of surface water drying in the eastern Boreal; and, [4] a region of no change along the Rocky Mountains. Based on the results presented here, mechanisms different from those suggested by Rodell et al. (2018) should be considered.

- "It is important to note that climate signals, particularly for the Pacific Decadal Oscillation [PDO] and Arctic Oscillation [AO] were not considered here. Oscillation in the climate system can be manifest as tests on short time periods." (L970) This seems like a big deal and was an immediate concern that popped into my mind as I was reading the manuscript the first time. Given that the stations sample different years, how does one consider the potential influence of periodic climate signals that may be in different modes for available station records?

[Authors response] It is indeed an important caution. This study reported changes and trends in a large spatial domain where warming is taking place [lines 53-84] but the focus was on detecting change and not attribution. In the paragraph at line 970 we discuss why some of the observed changes could be driven by PSO and AO and suggest a method which could resample the available data to be able to examine relationships to climate indices.

- "Interpreting the results from any fixed period may not be representative of historical variability (Hannaford et al. 2013)." (L1026-1027) Can the implications and importance of this statement for this study be expanded upon?

[Authors response] This originated as a response to a previous reviewer who was of the opinion that any trend study needed to have a common period of years. Our study did not have a fixed period but used all available data.

We removed this from the conclusion and moved it to the discussion. ~line 705.

I have a few specific comments listed below, but another part of the analysis that seems to warrant some consideration is how does size of the basin impact the trends in the satellite indices? I would guess that larger basins may aggregate different temporal patterns across the space thereby smoothing out potential trend signals. Are identified (or lack of) trends related to basin area?

[Authors response] This is an interesting question and one that should be considered in future research but is really beyond the scope of the present work. As the text indicates, we chose a simple straightforward approach that meets the needs of this study with nearly 400 basins. For each basin three time series (mean, maximum, minimum) from composite imagery for NCVI, NDWI, and NDSI and then determined trends using Mann-Kendall. This was necessary to treat these in a comparable basis in this context. As is mentioned in the text, more complex analysis such as bfast would be suitable for assessing change, but that analysis results in a large collection of statistical models that make comparing and contrasting this large number of basin impractical.

Specific comments

Abstract: Agree with another reviewer that this is too long. The first paragraph could be substantially reduced and incorporated into the second paragraph. There's perhaps too much detail provided here. Distill this down to the key points of the study.

[Authors response] We agree that the abstract is long in an attempt to summarize the key aspects of a complex study of a large and variable hydrological dataset.

We have made some edits to shorten the abstract by some 30%.

L35-38: Not clear what this sentence is trying to say.

[Authors response]
"Streamflow regime types, the trend patterns, and satellite indices trends each showed spatially coherent patterns reflecting the influence of sources in the Canadian Rockies and other range in the west and poorly defined drainage basins due to post-glacial topography in the east and north. "

As part of the rewritten abstract, now:
"*Regime types, trend patterns, and satellite indices trends each showed spatially coherent patterns separating the Canadian Rockies and other mountain ranges in the west from the poorly defined drainage basins in the east and north. "*

L70-71: 'hydrological' is repeated twice.

[Authors response] Second occurrence deleted.

L74-77: Run-on sentence.

[Authors response] The whole region is subject to strong seasonality, continental climate, near absence of winter rainfall, seasonally frozen soils and from the mid-boreal forest northwards and at high elevations the sub-surface below an annually thawed active layer are permanently frozen (permafrost).
Changed to:

The whole region is subject to strong seasonality, continental climate, near absence of winter rainfall, and seasonally frozen soils. From the mid-boreal forest northwards, and at high elevations, the surface is an annually thawed active layer below which materials are permanently frozen (permafrost).

L78: Change one of the 'dominate'

[Authors response] Changed to:
Tens of millions of lakes and wetlands cover the northern and eastern part of this region where the Canadian Shield dominates topography and hydrography.

L87-89: Seems like these two sentences repeat the same point.

[Authors response] The hydrographs of all rivers in this domain reflect contributions from snowmelt, the magnitude of which differ in both space and time. There are important regional differences in how snowmelt contributes to the seasonal pattern of hydrographs.

We agree, changed to:

The hydrographs of all rivers in this domain reflect contributions from snowmelt, the magnitude of which differ in both space and time.

L92: The rationale for ecozones comes on a little abruptly. So 'ecozones were chosen as an appropriate level for comparisons', but the reader doesn't really know what we will be comparing yet.

[Authors response] Ecozones were chosen as an appropriate level for comparisons (Marshall et al. 1999; Eamer et al. 2014; Ireson et al. 2015) rather than physical attributes such as climate, permafrost, or geology, since ecozones represent regions where the ecology and physical environment operate as a system.

Was changed to:

Ecozones (Marshall et al. 1999; Eamer et al. 2014; Ireson et al. 2015) were chosen as an appropriate level for comparisons of hydrologic patterns and trends rather than physical attributes such as climate, permafrost, or geology, since ecozones represent regions where the ecology and physical environment operate as a system.

L85-112: Some of this feels more like a description of the methods and data used. Consider revising this to focus the introduction on setting the context of the study and what research gaps are addressed, and use the methods section to focus on details about the data used.

[Authors response] This important background information for the reader. An alternative would be to move it to a separate section but we did not feel that would improve the paper.

Hydrological processes differ widely in this domain, which spans 11 of Canada's 15 terrestrial ecozones, and includes many small basins where streamflow is only temporary (Buttle et al. 2012). The hydrographs of all rivers in this domain reflect contributions from snowmelt, the magnitude of which differ in both space and time. Other flow contributions, from glaciers and rainfall, all vary spatially across the domain with glacier contributions focussed in high mountain headwaters and rainfall contributions increasing at lower elevations and latitudes. Ecozones (Marshall et al. 1999; Eamer et al. 2014; Ireson et al. 2015) were chosen as an appropriate level for comparisons rather than physical attributes such as climate, permafrost, or geology, since ecozones represent regions where the ecology and physical environment operate as a system. It is important to note that many river basins in this study originate in one ecozone and cross through other ecozones whilst maintaining the characteristics of their source.

Streamflow data in this domain is taken from stations that were operated either year-round or seasonally (MacCulloch & Whitfield 2012); seasonal stations generally provide records from April through the end of October, because there is either no streamflow in the winter, or because the channels become completely frozen. This approach contrasts with many studies that use only stations having continuous records and a common period of years (e.g. Whitfield & Cannon 2000); one novel aspect of this study is that it demonstrates a method which incorporates records from both continuous and seasonal stations. Trend assessment is conducted on an annual common time window for both continuous and temporary streams.

Landscape changes may cause or result from hydrological changes. Satellite imagery and derived spectral indices were used to assess the changes in the landscapes of basins in relation to their hydrological response. Normalized Difference Indices of vegetation, water, and snow (NDVI, NDWI and NDSI) were constructed using optical imagery from the Thematic Mapper (TM) sensor (USGS 1984) on board the Landsat 5 satellite for individual basins (e.g. Hall et al. 1995; Su 2000; Hansen et al. 2013; Pekel et al. 2016). Although the temporal coverage of the indices differs from that of the hydrometric data, the trends in these indices over many basins provide another perspective of change over the study domain.

L110-112: The difference in temporal coverage of the streamflow and vegetation/water/snow indices feels like a big deal that is somewhat glossed over here. In addition, I'm not sure what 'trends in these indices over many basins provide another perspective of change over the study domain' means. Could this be clarified?

[Authors response] Although the temporal coverage of the indices differs from that of the hydrometric data, the trends in these indices over many basins provide another perspective of change over the study domain.

We were not intending to 'gloss over' but to emphasize that those satellite data have coverage for another time period. Changed to:

The temporal coverage of the indices differs from that of the hydrometric data used in this study. Trends in these indices over many basins from the satellite imagery that is available provides a complementary perspective on hydrological change over the study domain.

L115: Do you need both 'separately' and 'individually' here?
L115: What is 'hydrological structure'?

[Authors response] "The objective of this study was to examine the hydrological structure and changes in seasonal streamflow patterns by combining data from perennial and temporary streams, which had previously only been treated separately and individually, to diagnose hydrological structure and change in western Canada's cold interior."

Changed to:

"The objective of this study was to examine the hydrological structure and changes in seasonal streamflow patterns by combining data from perennial and temporary streams to diagnose hydrological process differences and change across western Canada's cold interior."

L120: What 'mechanisms' were considered in this study? It seems like some basic indices were related to streamflow patterns. Why should this be considered as 'mechanisms'?

[Authors response] "[2] how are climate related trends and mechanisms distributed?, and "

Changed to:
"[2] how are climate related trends distributed?, and "

L142: What is meant by 'annual time window' here?

[Authors response]
The alternative approach used here includes data from a large number of seasonal and continuously observed sites, which are considered using only an annual time window described below.

Not sure why this is an issue as the common time window is described in the next paragraph.

Changed to

The alternative approach used here includes data from a large number of seasonal and continuously observed sites, which are compared using only the data available from April to the end of October.

L150-151: Why is the link to CCRN process-based studies important?

[Authors response] "and including these stations provides a link to CCRN process-based studies (DeBeer et al. 2016)."

This text is to inform the reader that our study of many stations, provide a spatial and temporal study of change which complements these process intensive study locations.

"and including these stations provides a link between the spatial patterns reported in this study and intensive process-based CCRN studies (DeBeer et al. 2015; in press)."

L164-166: Check grammar.

[Authors response]

"Satellite imagery and derived spectral indices are valuable for assessing effects of environmental changes the hydrological responses of the gauged basins as these methods allow determination of changes in vegetation, water bodies and snowcover for large areas (e.g. Hall et al. 1995; Hansen et al. 2013; Pekel et al. 2016; Su 2000)."

Changed to:

"Satellite imagery and derived spectral indices are valuable for assessing effects of environmental changes <u>and</u> the hydrological responses of the gauged basins; these methods allow determination of changes in vegetation, water bodies and snowcover for large areas (e.g. Hall et al. 1995; Hansen et al. 2013; Pekel et al. 2016; Su 2000)."

L180: What 'related packages'?

[Authors response] "Maps were plotted using `maps` and related packages. "

Deleted.

L310-314: How much of an issue was this? I can imagine certain regions across the study region are more prone to cloud cover than others. Did this bias the analysis in any way?

[Authors response]

"Theoretically this would allow for spatially-complete mosaics given the 16-day revisit time of the satellite. "

This is described in the sentences that follow in the text. We actually determined the actual coverage fraction of images a typical example of which is shown in Figure S14. We used the annual maximum, minimum, and mean of these composited over basin areas. Visual inspection suggests that these time series of indices demonstrate a common seasonal patterns despite the missing observations.

Section 3.4: There are a number of speculations made about hydrologic processes underlying the detected trends in streamflow and satellite indices (e.g., increased greenness results in increased ET, increased NDWI is due to increases in PPT). There are a number of limitations to these indices and making some of these links to actual hydrologic fluxes and storages may be confounded by other influences. It seems like some of these speculations could be checked by looking at available meteorological data, such as precipitation trends for some of these NDWI trends, for example.

[Authors response] We felt it important to make the suggestions and others might want to explore those in specific areas within our study. But, this suggestion would be outside the scope of our study.

L730-733: This feels like a major issue - that the key hydrologic runoff event for some of these catchments is not always accounted for using the time window considered in this study.

[Authors response]

"and c) the snow melt period often occurs before the beginning of the annual common time window."

This really is a simple statement of fact. Since we are not looking at annual runoff but changes in runoff across five day periods the data are being compared on a consistent and logical manner.

L745-748: I'm left wondering what the deal is with these flow regime types that contain only one or two stations. Why did these stations stand out? What makes them unique? It seems a lot could be learned about the hydrology and/or the methods used in this paper by looking at these specific classes more closely.

[Authors response]

Three clusters (Streamflow Regime Types 6, 7, & 8) each contained a single member and one (Type 10) had only two members. These Streamflow Regime Types are very different from those containing large numbers of members and any clustering of hydrographs in this way must balance common characteristics against uniqueness.

This is a classic issue with hierarchical clustering as it often separates individual of small groups of cases preferentially when dealing with time series. Our decision was to acknowledge these more 'unique' cases, but to focus on the groups with larger memberships.

L758-759: Could you elaborate on why these results are different from those reported by these other two studies?

[Authors response]

"Basins in mountainous areas generally lack consistent patterns of trends, but some exceptions do occur. This is counter to many studies that have shown late summer streamflow declines (Jost et al. 2012; Fleming & Dahlke 2014)."

Changed to:

Basins in mountainous areas generally lack consistent patterns of trends (Trend Pattern 5), but some exceptions do occur. Many studies in British Columbia have shown late summer streamflow declines (Leith & Whitfield 1998; Jost et al. 2012; Fleming & Dahlke 2014). This may be the result of hydrologic resilience (Harder et al. 2015) in mountain basins east of the continental divide.

L990: In what way are the available data 'messy'?

[Authors response]

The available data are messy, and there was an uneven distribution of stations by ecozones.

Changed to:

The structure of the available data is complicated, the dataset contains streamflow records from entire years and/or only the warm season, and for a variety of periods of years, and the sites are not evenly distributed across ecozones.

L991-994: Run-on sentence.

[Authors response]

Determining the magnitudes of annual trends was not appropriate with the approach used because of the inconsistency of years of data between sites, but by including the data from seasonal stations and an annual common time window rather than only entire years, the results provide an interesting spatial story of trend direction.

Changed to:

Determining the magnitudes of annual trends was not appropriate with the approach used because of the inconsistency of years of data between sites. But, including the data from seasonal and continuous stations and focussing an annual common time window rather than only entire years, the results provide an interesting spatial story of trend direction.

L1002-1019: The first half of this paragraph is a little strange to read. Stating that this study is important because in connects to existing research programs doesn't seem like a strong rationale. Emphasizing the points in the second half of this paragraph would make for a stronger point.

[Authors response]

The paragraph was split into two and the first portion rewritten.

The motivation for this study came from the NSERC Changing Cold Regions Network study of 2013-2018 (DeBeer et al. 2015; in press) of Western Canada's rapidly changing cold interior. That study sought to integrate existing and new sources of data with improved predictive and observational tools to understand, diagnose, and predict interactions amongst the cryospheric, ecological, hydrological, and climatic components of this study area. The results presented here are already informing several science research agendas in Canada and internationally. The results also contribute to the Global Water Futures programme of 2016-2023 (www.globalwaterfutures.ca), which has an overarching goal to deliver risk management solutions to manage water futures in Canada in a time of unprecedented change.
The first step to managing future water changes is to understand those of the recent past and that are currently underway and this study makes a strong contribution to that process. In particular, this study represents a step forward in addressing the complexity of hydrological change; there are many studies of individual basins where, when results are considered individually, tend to be more anecdotal than systematic. It is indeed simpler and easier when there are only a few cases to consider with common variables and record length, as in many studies, but with large numbers of basins the tools for dealing with significant changes are more limited. Sensitivity studies that assess the limits of partial year analysis of hydrological structure and change are required and seem a logical next step.

Table 2: What is the point of including this table? Could be moved to the SI.

[Authors response]
This table was added in response to an earlier reviewer's suggestion as they wanted to be able to know actual dates corresponding to period numbers. Moving it to Supplementary material would make it necessary for the reader to have the SI in hand.

Figure 1: The map projection used doesn't look typical for showing this region. The northern region is stretched considerably.

[Authors response]
This is standard Mercator projection that is commonly used in North America. It indeed stretches the north.

Figure 6: The dashed lines are obscured in some of the plots (e.g., regime type 1 and 5). The axes labels could be more descriptive. Colouring the individual lines doesn't make much sense to me since we have no way to connect these back to stationID.

[Authors response]
This is an issue with the originating software that uses "ggplot" where the centroid is plotted first and over-plotted with the individual lines that are coloured based on internal code in that function. We have changed the description to be more precise:

“Each of the twelve plots contains a line for each gauged basin in that Type and the heavy dashed line, where visible, is the centroid of all members; the colour of the lines is based upon stationID.”

Many of the figures are not colour-blind friendly. I would suggest using a more colour-blind friendly colour palette.

[Authors response] Using colour requires many considerations. Where we had control, we have used contrasting and consistent colours and line types. Two important constraints we faced were [1] using the ecozone colours conforming to those commonly used in ecosystem/ecoregion mapping, and [2] using package default graphs in dtwclust that are produced using “ggplot”.

Figure	
1	*Map with red symbols and brown basin outlines on Mercator projection.*
2	*Default colours from CSHShydRology function with red down and blue up arrows*
3	*Default colours from CSHShydRology function with red down and blue up arrows*
4	*Solid lines in gray and heavy dashed lines in blue*
5	*Red dashed line and solid blue lines*
6	*Plot from dtw that relies on “ggplot”. Line colours assigned based on case order. Heavy dashed lines [centroids] plotted first.*
7	*Cluster colours based on the ggplot output in Figure 6. Base ecosystem colours conform to those used by originating agency.*
8	*Line colours defaults within dtwclust which uses ggplot.*
9	*Default colours and line weights from CSHShydRology with vertical heavy blue dashed lines added.*
10	*Red and blue pixels match colours for decreases and increases. Consistent use of six cluster colours. Base ecosystem colours conform to those used by originating agency. Added dashed lines differ in dash length and colour.*
11	*Colours conform to those used for six change patterns. These were chosen to reflect the types of clusters. Four increasing patterns are shown with distinct shades of blue [3] and darkgreen; one decreasing pattern is shown in red, and the cluster without patterns in a neutral gray. These six are used consistently throughout.*
12	*Consistent use of six cluster colours. Base ecosystem colours conform to those used by originating agency.*
13	*Red down triangles and cyan up triangles used to indicate trends. Base ecosystem colours conform to those used by originating agency.*

Reply to HESS-613 R5

Review of "The Spatial Extent of Hydrological and Landscape Changes across the Mountains and Prairies of Canada in the Mackenzie and Nelson River Basins Based on Data from a Warm Season Time Window" by Whitfield et al.

The manuscript investigated seasonal trend in streamflow observations over the Mackenzie and Nelson River basins. The authors maximized the number of observations to analyze in the study by using a warm season time window instead of common period of years. They identified 12 streamflow regime types using the dynamic time warping, six trend patterns, and three particular areas of change relating to the trends in the satellite indices. They showed that the clustering method can organize similar hydrographs that vary in magnitude and timing due to different timing of snow accumulation and melt by latitude and elevation. The streamflow trends were explained w.r.t. the changes in NDVI, NDWI, and NDSI that exhibited complex spatial variation and connection. I think that the analyses method is innovative and promising in studying seasonal trends that can shed lights on localized and shorter time scale phenomena that do not appear in annual trend analyses. The manuscript is highly relevant and worth a publication, however it is a bit difficult to follow in its current form and reorganization and modification of presentation are desirable as most of the concerns raised by referees. My specific comments are below.

[Authors response] We thanks this reviewer for taking the time to review our manuscript.

Major concern

While figures are illustrative in showing different nature of data availability, hydrograph, and trends across the stations, some figures seem better suited in the Supplement section (Fig 5, 8, & 9). I concur with a referee's suggestion on focusing on the main findings (Streamflow Regime Types, Trend patterns, and three areas of changes) and separating/expanding figures related to them (Fig 7, and 12). Fig 6 is one of the main findings but claimed it's difficult to see and Fig 8 is the simplified version that is referred more in the text. While Fig 6 lines are color-coded for different stations, no corresponding map of color-coded station is provided, and the centroid is washed out. I recommend combining Fig 6 and 8 by using lighter colors for the stations (or make them less opaque) and using black thick line for centroid to make them stand out. Figure 7 and 12 are the main findings as well but it's difficult to distinguish stations because they overlap, and also the color shading for ecosystems clash with the markers. I think that it's worth separating into several maps showing a few Regime Types and Patterns per map like maps of S16-21. I also recommend using gradient shading for the ecosystems. I did not find Fig 9 to be significant.

[Authors response] We have separated the elements in this paragraph and addressed them individually below.

[1] While figures are illustrative in showing different nature of data availability, hydrograph, and trends across the stations, some figures seem better suited in the Supplement section (Fig 5, 8, & 9).

[Authors response] Previous reviewer's requested figures that provided more information about the methodology. It was for this reason that these Figures 5 & 9 were added to the paper. Figure 8 was added to make the cluster centroids comparable. Again, Figures 6 and 8 were produced using dtwclust.

[2] I concur with a referee's suggestion on focusing on the main findings (Streamflow Regime Types, Trend patterns, and three areas of changes) and separating/expanding figures related to them (Fig 7, and 12).

[Authors response] We have rewritten the abstract to improve the focus on the main findings.

It is our opinion that Figures 7 & 12 which are supported by S16-21 are sufficient to show the Trend Patterns and the Regime Types within that Pattern. Table 5 demonstrates the intersection between the Patterns and Types. To show a Figure and a description for each Pattern or each intersection between Type and Pattern would unnecessarily increase the length of the paper.

[3] Fig 6 is one of the main findings but claimed it's difficult to see and Fig 8 is the simplified version that is referred more in the text. While Fig 6 lines are color-coded for different stations, no corresponding map of color-coded station is provided, and the centroid is washed out. I recommend combining Fig 6 and 8 by using lighter colors for the stations (or make them less opaque) and using black thick line for centroid to make them stand out.

[Authors response] As we responded to R4:
This is an issue with the originating software that uses "ggplot" where the centroid is plotted first and over-plotted with the individual lines that are coloured based on internal code in that function. We have changed the description to be more precise:
"Each of the twelve plots contains a line for each gauged basin in that Type and the heavy dashed line, where visible, is the centroid of all members; the colour of the lines is based upon stationID."

[4] Figure 7 and 12 are the main findings as well but it's difficult to distinguish stations because they overlap, and also the color shading for ecosystems clash with the markers.

[Authors response] Using colour requires many considerations. Where we had control, we have used contrasting and consistent colours and line types. Two important constraints we faced

were [1] using the ecozone colours conforming to those commonly used in ecosystem/ecoregion mapping, and [2] using package default graphs in dtwclust that are produced using "ggplot" for Hydrological Types.

[5] I think that it's worth separating into several maps showing a few Regime Types and Patterns per map like maps of S16-21.

[Authors response] It is our opinion that Figures 7 & 12 which are supported by S16-21 are sufficient to show the Trend Patterns and the Regime Types within that Pattern. Table 5 demonstrates the intersection between the Patterns and Types. To show a Figure and a description for each Pattern or each intersection between Type and Pattern would unnecessarily increase the length of the paper.

If the Editor feels it necessary, we would make a csv file available that provided the details for each station including: StationID, Latitude, Longitude, Ecozone_name, Ecozone_number, H_Type, and Trend_Pattern.

[6] I also recommend using gradient shading for the ecosystems.

[Authors response] We chose the colour scheme of ecozones conforming to those commonly used in ecosystem/ecoregion mapping.

[7] I did not find Fig 9 to be significant.

[Authors response] Figure 9 provide a simple example of trends at one station and how the observed trends link into the large group of station in Figure 10. This Figure was added based on a previous reviewer comment suggesting that this step in the analysis be made explicit.

<u>Minor corrections</u>

Ln#71: hydrological repeated

[Authors response] corrected

Ln#294: "having than three years of data"=> having more than three..

[Authors response] Corrected.

Ln#342: what is bfast? R function name?

bfast is the acronym for "breaks for additive season and trends"

[Authors response] Verbesselt et al. (2010, 2012) and de Jong et al. (2012) used breaks for additive season and trends (bfast) to detect change, particularly phenological change, in satellite imagery; bfast iteratively estimates the time and number of abrupt changes within time series derived from satellite images.

We have underlined the letters to help:
Verbesselt et al. (2010, 2012) and de Jong et al. (2012) used breaks for additive season and trends (bfast) to detect change, particularly phenological change, in satellite imagery; bfast iteratively estimates the time and number of abrupt changes within time series derived from satellite images.

Ln#446: typo "bsin"

[Authors response] Corrected.