Comments to Author

The authors have addressed many of my comments on the original manuscript, but did not present adequate response to a number of comments. I have annotated a PDF file (attached) of their response with further comments and suggestions. In addition, in Line 862 and 1510 in the marked manuscript, Hayashi and Rosenberry (2002) is cited as a reference regarding ecoregions, but this paper did not discuss ecoregions at all. Please remove the reference from these sentences.

Response to reviewers: “Watershed classification for the Canadian Prairies”

Please note that we have changed the manuscript title to: “A WATERSHED CLASSIFICATION APPROACH THAT LOOKS BEYOND HYDROLOGY: APPLICATION TO A SEMI-ARID, AGRICULTURAL REGION IN CANADA”.

Approximate page and line number references for the changes are in (page#, line#) format.

Response to Referee #1

Response to GENERAL COMMENTS

We thank the reviewer for their comments, and we appreciate the time taken to provide them. Yes, these traits of the Canadian Prairie may have been known by select individuals qualitatively for some time, but it is necessary to conduct this analysis quantitatively so as to begin to address some of the most pressing water management issues on the Canadian Prairie. This manuscript alone is a sizeable body of work, requiring careful and lengthy description. Extension to an application of the classification would render a single manuscript unwieldy. Applied use of the classification results will be pursued in subsequent papers. We agree that one of the scientific contributions of this work is in improving quantitative understanding of classifications in this region, which is why we expanded discussion of comparisons to previous classifications in this new version.

Response to SPECIFIC COMMENTS

Line 102, 108. How is “watershed” defined? Is it straight forward to define watersheds in an unambiguous manner? Please clarify that here, or in the methods.

> *We thank the reviewer for their comments. We have added clarification on operative definition of watershed used here in the methods, as well as additional detail on derivation of watershed boundaries. (23, 959)*

Line 117. How is the Canadian Prairie defined? Please present a brief definition, and the source of the ecozone boundary shown in Figure 1.

> *We have added a brief description on the ecozone, including vegetation, to section 2.1. The source for the ecozone boundary has been added to Figure 1. (23, 971)*

Line 119. The upper bound of precipitation (650 mm) seems to be too high….

> *We have changed the value in the sentence and those of mean annual air temperature and provide clear references to the source of these statistics. (23, 965)*

Line 128. Related to my comments on Line 102 and 108, how are these watershed outlet selected? Please explain.

> *We define the use of “outlet” for the purpose of this study on section 2.3.2., whereby it is the lowest elevation along the watershed boundary. (25, 1042)*

Line 136-138. As it is written, the sentence indicates that the watershed of the Saskatchewan River is excluded from the analysis, which is clearly not the case.

We thank the reviewer for this suggestion, and agree that the sentence was misleading. We have removed the sentence and adjusted text for clarity. (22)

Line 140. Please indicate roughly how many kilometers are equivalent to 15 arcsecond in the Canadian Prairie.

We thank the reviewer for this comment, which was shared by Referee #2. We provided the metre equivalents at Saskatoon, Saskatchewan, which is located within the Prairies ecozone. The paragraph now reads: "Delineations of candidate study watersheds were obtained from the HydroSHEDS global dataset (Lehner and Grill 2013). Watershed boundaries within this dataset were based on Shuttle Radar Topographic Mission (SRTM) digital elevation model (DEM) calculated at a 15 arc-second resolution. The resolution is equivalent to for example approximately 285 m east-west and 464 m north-south at Saskatoon, SK." (23, 966)

Line 141. The authors describe watersheds by referring the reader to Figure 1. However, Figure 1 does not show watersheds. Please refer the reader to Figure 5 instead, or add watershed boundaries to Figure 1.

We have removed the reference to the figure at line 141 as it was decided to be unnecessary. (23, 960)

Line 145. What is the total area of 4175 watersheds? How does that compare to the total area of the Canadian Prairie?

The area for the Prairie ecozone (4.7×10^5 km^2) and the watersheds included in the study (4.2×10^5 km^2) are now provided. (23, 960)

Line 156. Please see my comments above on CANGRID.

CANGRID is the only gridded product data that uses the Adjusted Homogenized Canadian Climate Dataset, and we felt it the most appropriate to use in this region where precipitation undercatch in gauges is very pronounced. We have added clarification in the text. (24, 1022)

Line 161. Temperature-index methods such as Thornthwaite do not give reliable estimates of "potential evapotranspiration" … please explicitly acknowledge its limitation.

This acknowledgment was addressed by including the following sentences: "To maintain consistency among climate data, and use the same temperature data as described above, options were limited with which to calculate PET. PET was calculated from the Thornthwaite equation (Thornthwaite 1948) using the SPEI package (Vicente-Serrano et al., 2010). A disadvantage of the Thornthwaite approach is it assumes a correlation between temperature and radiative forcing and adjusts for any lag in this relationship using corrections for latitude and month." (24, 1037)

Line 162. The balance between precipitation and evapotranspiration is reflected in ecoregions of the Prairie, as plants are good indicator of long-term water balance. … Please provide an explanation.

Please see above for a more detailed explanation on ecoregions. Briefly, we acknowledge vegetation as an indicators of the water balance. However, in the Prairies, much of the local "natural" vegetation in not reflected due to human land modification (e.g., agriculture). We use the

landcover types from AAFC to consider portions of the natural vegetation, such as woodlands and grasslands.

Line 167. How were these non-effective areas determined? Please briefly explain the method and cite a reference. This is well known to Canadian Prairie hydrologists, but HESS is an international journal.

These were defined by (Mowchenko and Meid, 1983). We will include this citation and provide a brief description. We also provide more detail in Section 2.3.2 as to the impact of non-effective areas to prairie hydrology, and we included the following description: "The location of these regions are shown in Figure 1. This definition stems from work by Agriculture and Agri-Food Canada where prairie drainage areas were divided into gross and effective drainage areas, whereby the former describes the divide that is expected to contribute under highly wet condition, and the latter is the area that contribute runoff during a mean annual runoff event (Mowchenko and Meid, 1983). Thus, at its simplest, the non-effective area is the difference between the gross and effective drainage area; however, the exact area contributing runoff is dynamic and the controls complex, which include antecedent storage capacity and climatic conditions (Shaw et al., 2012: Shook and Pomeroy, 2015)." (24, 1016)

Line 177. Please change the wording to "seasonally flooded prairie potholes". Potholes are permanent landscape features, whereas flooded areas can be seasonal.

Thank you for the clarification, and we have considered this comment in our revision. Given suggestions made by Referee 2, we have adjusted the sentence to indicate what is meant be "prairie potholes" as follows: "As such, "wetland" in this context can include some seasonal ponds (i.e., prairie potholes) as well as larger or more permanent shallow water bodies". (25, 1034)

Line 180. Is (wetland density) needed here?

We thank the reviewer for the suggestion. We removed this fragment and adjusted the sentence for clarity. (25, 1037)

Line 191. Please briefly explain the meaning of mu and beta, and indicate the dimension or unit. These must have a unit of area to maintain the dimensional homogeneity.

We thank the reviewer for the suggested and the paragraph was modified to describe the meaning of the Pareto distribution parameters and the units. The paragraph now provides explanation of the meaning of the parameters within our context and the units. (26, 1053)

Line 195. Is it true that all pixels in the Canadian Prairie have "monthly" satellite images? I do not think that is the case. Please clarify that in the texts.

We thank the reviewer for their comments. The maximum water extents were computed from Landsat images over the 32-year period, which have 8-day or 16 day revisit times. In this context, the Canadian Prairies has monthly satellite images. We have removed the sentence of concern and added the following for clarity: "Note that because the sizes of the water bodies were taken from infrequent remote-sensing measurements (i.e., the Landsat data have a minimum revisit time of 8 or 16 days), they also are biased against short-lived water bodies." (26, 1062)

Line 197. What do you mean by "the median area of the largest wetland"? Please

re-phrase so the reader can understand what you mean.

We have clarified this in the text by adding more detail in the description of the term, as well as in the Line of concern. It is the median of the distribution of the "area of the largest wetland" (W_L) for the watersheds within each class. We provide the following description in the text: "The median area of the distribution of largest wetlands for each watershed class provided an indication of the maximum sizes of the water bodies exhibited those watersheds, and thus provided a maximum value to simulate fitted values". (26, 1058)

Line 205. Surficial geology is mapped by geologists in each province using different terminologies. I am not sure if the "comparison across provincial boundaries" is straight forward. Please add a brief explanation on how the difference in terminology and mapping methods was reconciled.

Amelioration among surficial geology definitions was performed by grouping more defined classification into broader categories describing depositional features. Grouping was performed by comparing definition of each feature type using the provincial government metadata and informed by advice from a colleague in geology. We acknowledge that these are broad groupings and ideally we similar framework used across the provinces would be ideal. However, for our current purposes, these broad descriptions were useful in capturing a variation in at least broad geological settings.

Line 208. In the Canadian System of Soil Classification, colour indicates more than just an appearance of soil. For example, Black Chernozem and Dark Brown Chernozem are distinct soil types developed under distinctively different climatic conditions. The distribution of these soil types often coincides with ecoregions (e.g. Black Chernozem is associated with Aspen Parkland). Please consult with local soil scientist to give a better context to soil classes. Also, somewhere in the paper, perhaps near the beginning of the method section, it will be useful to present a process-based framework to understand the eco-hydrological functions of the Canadian Prairie landscape (see my comment on Line 162).

We thank the reviewer for this insight and have edited the text accordingly. We recognize that the "colour" is only a descriptor and the function of the soils are different among soils types, and that they develop under specific climatic conditions, geology, and vegetation. These were implicit in the data that we used. We also included soil texture class data to provide additional description of soil characteristics. (28, 1110)

Line 223. Please indicate the unit of DSF. It must be the inverse of length.

We thank the reviewer for the comment. We adjusted the description to indicate that DSF is in units of km^{-1}. We also added units for perimeter (km) and area (km^2). (27, 1100)

Line 255. Please indicate these prairie stations in Figure 5. I assume these are the "study watersheds" described in Line 472. Please point that out here.

We note the "study watersheds" in Line 473 is misleading. Here we are referring collectively to the 4100+ watersheds used in the clustering analysis. We have revised the section for clarity. (28, 1130)

Line 265. Please explain how V1 and V2, and W1 and W2 are defined. Please note that most readers of HESS are not familiar with CCA. You do not have to present

detailed explanation of CCA, but you need to give a brief outline so that the reader can understand the basic concept.

We thank the reviewer for the insight. We have made necessary adjustments to describe the methods in more clarity. This concern was shared with the other reviewers. We have re-ordered some of the sentences in the paragraph so that it now reads:

"Briefly, CCA involves correlating streamflow to physio-climatic characteristics of gauged watersheds to create canonical variables. These canonical variables (i.e., V1, V2, W1 and W2) are constructed from linear combinations of the original variables such that the correlation (λ) of the canonical variables is maximized. Positive canonical correlation coefficients imply positive relationships and negative canonical correlation coefficients imply negative relationships. There are two canonical variable sets; one for physio-climatic variables (i.e., V1 and V2) and another for hydrological variables (i.e., W1 and W2). Canonical variables plotting similarly on X-Y plots (W1-W2 and V1-V2), indicate good correlation (Spence and Saso, 2005). If canonical correlation values are above 0.75 (Cavadias et al., 2001), that set of variables was deemed useful for estimating hydrological variables from physio-climatic ones. Those physio-climatic variables passing this threshold were included as variables in a multiple regression to develop a predictive equation for Q2. Analyses were performed using vegan package (Oksanen et al. 2018). (29, 1152)

Line 266. What are "the original variables"? Please explain, using a table if appropriate.

We have adjusted the sentence for clarity by referring to the Table summarizing the original variables. (29)

Line 290. ". . . attributes and is the basis . . ." for matching the tense.

We thank the reviewer for the comment and have edited.

Line 301. Please define alpha.

We thank the reviewer for the comment and have edited. (31, 1204)

Line 310. What does this mean? Based on Line 269, does it mean that the result was very useful for V1-W1, and barely useful for V2-W2? Please explain.

We have adjusted the sentence for clarity by referring to the Table summarizing the original variables. (30, 1152)

Line 311. What correlation value would indicate "strong"? Does it have a statistical level of significance, like in the standard correlation analysis? Does a negative value indicate negative correlation? Please explain.

Thank you for the suggestions. Yes, positive correlation coefficients imply positive relationships and negative correlation coefficients imply negative relationships. We have included these descriptions to the methods description of the CCA, as included in the new paragraph above. There is a sentence included that says "if correlation values are above 0.75 (Cavadias et al., 2001), those were deemed useful for estimating hydrological variables from physio-climatic ones." (29, 1159)

Line 311-312. It is true that the correlation value is strong between Q100 (1:100 flow)

and W2, but it is weak for Q2 (mean annual flow) and W2. On the other hand Q2 and W1 has a strong correlation. Also the lambda value is much greater for V1-W1 combination than for V2-W2 combination. Given that, why was W2 chosen? Is it because the classification is designed for 1:100 flood prediction? Please provide an explanation.

The second set of canonical variables (V2 and W2) were chosen because the individual canonical correlation coefficients were higher than V1 and W1. We rephrase the paragraph to discuss bias and reason for choosing the variables: "This sentence has been included into the text: "The canonical coefficients from the CCA were λ1 0.97 and λ2 0.77, respectively. Mean canonical correlation values between the hydrological variables and W2 were greater than those with W1 (Table 1), and because both values of □ were acceptably large (Cavadias et al., 2001) the physio-climatic variables strongly associated to V2 were used in the multiple regressions0 … Plots of observed and predicted runoff Q2 (R2=0.45) and Q100 (R2=0.48) show moderate agreement at lower flow values (Fig. 2). There is a negative bias estimated between 26 and 29%,…..". (33, 1262)

Line 322. How is rock fraction area calculated? I cannot imagine there are many areas of exposed bedrock in the Canadian Prairie. Please explain.

There are regions of exposed bedrock, particularly in Southern Saskatchewan. We invite the reviewer to the following map of surficial geology at http://publications.gov.sk.ca/documents/310/93756-Surficial%20Geology%20Map%20of%20Saskatchewan.pdf . Rock is shown in pink, and is labeled "R". This landscape was mainly associated with dissected valleys and riverine systems.

Line 326. Please list the classes of surficial geology used in the analysis.

We have included a table of the surficial geology classes, as well as over components of the compositional datasets, in the supplementary data (Table S3).

Line 347. What are the "PCs from compositional datasets"? Are these different from PC1-PC6 in the header of Table 3? Please explain.

These are not the same Principal Components (PC). The "PCs from compositional datasets" were used to capture the main gradients in the physiogeographical dataset (e.g., surficial geology) that are then used in the PCA for the cluster analysis. This was comment was also echoed by the second reviewer. We provide a figure that shows our workflow.

Line 358. "Weaker", not "less strong".

We have revised accordingly.

Line 389. The Canadian Prairie has now been divided into seven classes, which seem to be consistent with our current understanding of eco-hydrology. For example, C1 roughly coincides with the ecoregion "Lake Manitoba Plain (162)" in the Ecozones and Ecoregions of Canada (Ecological Stratification Working Group, 1995). Then, what new knowledge and insights can we learn from this exercise? It will be nice to see a clear demonstration of the contribution of this study to new advances in "Hydrology and Earth System Sciences". Please try to present that in the discussion section.

We thank the reviewer for their insights into the use of eco-hydrology and comparing our findings to these classifications. We included references to ecoregions and discussed the similarities and difference in these two approaches in the Discussion. Briefly, we see some relationships with boundaries, however, we can identify areas that are not considered in the more general ecoregion description, and provide a discussion on new insights gleaned beyond ecoregions. (41)

Line 412. Glacial till and hummocky landforms. Does this refer to one thing, or two separate things (till and hummocky landforms)? Hummocky landform is a sub-class of glacial till terrain. Please clarify.

We thank the reviewer for this observation. It is true that hummocky landforms are associated with glacial till deposits. However, the landforms dataset describes forms that include aspects of surficial geology, relief, among others. Therefore the two datasets are related. We feel that both datasets offer information on local geography. The hummocky landform designation is particularly useful for characterizing landscape drivers depressional storage and overland flow. (41)

Line 453. Brown Chernozem is associated with the "Mixed Grass (159)" ecoregion, which covers much of the driest part of the Canadian Prairies, commonly referred to as the "Palliser Triangle". Accordingly the outer boundary of C5 roughly coincides with the outer boundary of Mixed Grass. However, Figure 5 shows a patch of C6 in the core of the Mixed Grass, which is the driest part of Alberta having distinctly different eco-hydrological characteristics compared to the band of C6 parallel to the western boundary of the Prairie. Is the new method picking up new information, or is it erroneously classifying watersheds? Are there too many classes in the system? These are worth discussing in this section.

Thank you for your observation. The classification indeed classifies watersheds outside of what would be defined as a traditionally eco-hydrologically-based region. We expand on this idea in the Discussion of our revised version. Briefly, we have confidence that the majority of watersheds are being classified similarly resulting from our resampling analysis. Although some watersheds might be seemingly spatially disparate, they exhibit characteristics that warrant membership to a specific class. In the case of C5 and C6, they coincide well with the Mixed Grass ecoregion; however they differ fundamentally in physical controls on hydrology (e.g., slope, non-effective area), and thus provide additional information beyond ecoregion description. (41)

Line 472. Are there 11 study watersheds, as indicated in Line 255? If so, is that a high enough number to examine all seven classes? Please explain.

We address the concern with the miscommunication of the "study watersheds". However, we acknowledge the concern of extrapolating data from 11 watersheds. However this is an approximation of a hydrological runoff variable.

Line 490-493. It is true that few studies have classified "watersheds" in the prairies, but there have been numerous studies examining the spatial distribution of ecohydrological functions of the Prairie landscape. For example, ecoregions are an integral measure of hydro-climatology. Please acknowledge previous efforts and highlight the newness of this work.

We discuss this above. We added acknowledgement of the contribution of ecoregions in the Discussion (41). We thank the reviewer for the insight.

Line 502. This is an example demonstrating the strong effect of ecoregions on hydrology.

We discuss this above and thank the reviewer for the insight. We added acknowledgement of the contribution of ecoregions in the discussion under section 5.1.2 (41).

Line 633. Yes, but the delineation has been available for many decades in the form of ecoregions. Please acknowledge it.

Given the comments related to ecoregions, we have added a section within the discussion to discuss the similarities and differences it the approaches, and insights gleaned (41).

Line 637. Geography may not be an appropriate term here, because geography encompasses many things, not just landforms. I would say topography or landform is more appropriate.

We agree with this edits and the sentence has been revised to consider the comment. "Geography" was switched to "topography". (42, 1559)

Line 661. Figure 8 just shows wetland density and area delineated in satellite images, which is dependent of climatic factor (wetness) in addition to depressional storage capacity. Overall, I believe that the data from the 11 study watersheds can be utilized more to demonstrate the validity and usefulness of the new classification method. For example, are there distinct differences in the hydrological characteristics of seven classes of watersheds?

As mentioned above, the 11 watersheds were only used for the CCA. The issue with using these to compare the classes is that these watersheds do not compare to the same scale as the watersheds derived from HydroSHEDs. Moreover, they tend to represent large, river-dominated systems, and mostly coincide with C4, C6, and C7. We use the wetland simulated data to compare how the classes represent observed data. We thank the reviewer for their comments, and we have elaborated on this in the text. (45, 1638)

Response to Referee #2

Response to GENERAL COMMENTS

We appreciate the helpful suggestions and advice provided by Referee #2. Overall, the suggestions constructively added to the content of the manuscript. Specifically, we have added additional references and re-ordered the structure of the Introduction to emphasize applicability to an international audience. We also divided the Methods section into Data Collection (2) and Data Analysis (3) as per the suggestions of Referee #2. We felt this suggestion added to the readability of the manuscript. Finally, we have added more detail on the CCA method, which was a concern shared by other reviewers.

Response to SPECIFIC COMMENTS

1. International readers might not be able to place the Canadian Prairie on a map (line 55). A brief statement about the geographical extent of the Prairies would help.

> *Increased detail regarding the Prairie region, and what distinguishes it, was also suggested by reviewer #1. As discussed in our response to reviewer #1, we provide greater detail of the Prairies ecozone in Canada in the methods and introduction, including the spatial extent of the region in the introduction. (pages 19-21)*

2. "Hydrological characteristics" (line 71) is unclear. Do the authors mean catchment attributes (e.g. topography, soils), climatic conditions, statistical properties of the streamflow regime or something else?

> *Yes, here we mean statistical properties of streamflow regime. This clarification has been added in the text. (21, 897)*

3. It would be helpful for the reader to briefly summarize how well earlier classification attempts have worked (line 74-78) and where the authors see current challenges.

> *In this regard, we are not concerned with whether these approaches have not "worked" but rather that although there have be attempts to classify watersheds/regions, they either do not extrapolate across provinces or are too coarse to represent heterogeneity within the Prairie. This is now better described in the Introduction. As reviewer #1 pointed out, ecoregions have been used to represent hydrological response by landscape characteristics in eco-hydrology. Our response to this latter comment can be found in our response to Referee #1. We appreciate the suggestion from reviewer #2 and provide detail to address some of this concern. (21, 897-915)*

4. The HydroSHEDS webpage (https://www.hydrosheds.org/page/development) lists a few regions where the data set is prone to errors, including areas with low or not well-defined relief. Is this of concern in the Canadian Prairies?

> *The error associated from datasets derived from SRTM can be of concern for the Prairies. Given this, the dataset does provide us with delineations at the scale of interest (~100km^2), and is the only dataset of that sort available. As a result, we deem it sufficient for our purposes given the current state of data availability for the region. We acknowledge the uncertainty in the dataset in the text with the following revision: "As with other SRTM products, the HydroSHEDs dataset may be prone to errors in regions with low relief due elevation precision of 1 m. However, the dataset*

provided an objective delineation over the region of interest and was sufficient for purpose of the current study." (23, 968)

5. Approximately how many meters are 15 arc-seconds (line 140) in this area?

This comment was shared with Referee #1 and we provide the distance measure in meters: "The resolution is equivalent to for example approximately 285 m east-west and 464 m north-south at Saskatoon, SK." (23, 966)

6. What motivated the choice for these specific area (line 142) and urbanization (line 143, Table S1) thresholds?

The choice in threshold areas was to remove very small "watersheds" or those that were very large, which tended to relate to lake basins (e.g., Lake Winnipeg). The urbanization threshold was informed by visual inspection of watersheds surround known large urban centers. A threshold of 40% removed most of those that had a large portion covered in urban development. We wanted to focus on those watersheds that were more "rural" and reduce the immediate impact of cities or development, which are known to produce unique impacts on local hydrology. We could not remove urbanized areas completely due to the number of rural communities and roads that exist across the Prairie region. We acknowledge the legitimate impact of cities and urbanization on water quantity and quality necessitates consideration, but these questions are not in the scope of the current manuscript. We added: "Because HydoSHEDs includes the basins of larger water bodies, including lakes, watersheds consisting of majority water were removed as the study concerns the uplands of these systems. Finally, highly urbanized areas (i.e., watersheds with cover being >40% urban) were removed." (23, 974)

7. The spatial resolution of climate data (line 157) seems large compared to the resolution of the watershed boundaries. Can climate data on this resolution still be considered representative for the smaller catchments?

Please see related comment on the CANGRD in response to Referee #1.
The text now states that the original data has been interpolated by kriging to a higher spatial resolution raster. (24, 998)

8. What is the rationale for choosing the Thornthwaite method (line 161)?

This comment was shared by Referee #1. The text now includes an acknowledgement of the reason for choosing this method and a limitation: "To maintain consistency among climate data, and use the same temperature data as described above, options were limited with which to calculate PET. PET was calculated from the Thornthwaite equation (Thornthwaite 1948) using the SPEI package (Vicente-Serrano et al., 2010). A disadvantage of the Thornthwaite approach is it assumes a correlation between temperature and radiative forcing and adjusts for any lag in this relationship using corrections for latitude and month." (24, 1006)

9. Snow formation and melt can strongly influence the seasonal water distribution and accounting for the fraction precipitation that occurs as snowfall has recently proved valuable in hydrologic similarity research (Knoben et al, WRR, 2018; https://doi.org/10.1029/2018WR022913). Is there any particular reason why the authors use only mean P and ET in their clustering?

We thank the reviewer for the suggestion, and we agree that inclusion of this parameter is and likely valuable for the Prairies. We focused solely on precipitation and ET because these variables were available at the temporal length and spatial extent for the study. Given the

limitations of the dataset we used, calculating parameters at a seasonal scale might introduce additional uncertainty, and thus was not included here. However, fraction of snowfall should be considered in future iterations provide the data resolution is available.

10. What is meant with a wet cycle (line 176-177)?

We removed reference to a "wet cycle" and the sentence now reads: "The 30-year period was chosen to capture natural climate variability". We thank the reviewer for their comment, and we think this edit better reflects our intentions. (24, 997)

11. Please include a (short) definition of potholes (line 177).

Thank you for the comment. Given suggestions made by Referee 1, we have adjusted the sentence to indicate what is meant be "prairie potholes" as follows: "As such, "wetland" in this context can include some seasonal ponds (i.e., prairie potholes) as well as larger or more permanent shallow water bodies". (25, 1034)

12. Why is the Lw/Lo metric (line 184) relevant? What does this metric tell us about watershed behaviour?

The metric identifies how close (or far away from) the largest wetland depression is to the watershed's outlet. It is meant to be an indicator of hydrological gate-keeping and thus controlling the likelihood for the watershed contributing flow to the downstream watershed. We explain this concept in the Introduction and beginning of the Methods. We considered placing more context in this regard, and we added the following clarification: "Both WL and LW/LO can be used to evaluate the relative importance of hydrological gate-keeping; for example, larger wetland depressions located closer to the outlet control the likelihood of the watershed contributing flow downstream and attenuating peakflow (Shook and Pomeroy, 2011; Ameli and Creed, 2019)." (25, 1043)

13. The climate data (line 156), land cover data (line 230 and further) and hydrological data (line 252 and further) cover different periods in time (1970-2000 for climate, 2011/2016 for agriculture land use, 1990-2014 for hydrologic data). For a general classification of similar regions, overlapping time periods for the data sources would be more appropriate. What is the rationale for not doing this?

We think the reviewer offers a valid concern and we thank them for the insight. Land cover because we wanted the most recent measurement to show current cover. The older climate data was used because of the reduction in reliable precipitation data from Canadian climate stations since 2000. Additional explanation of this now provided in the text. (24, 996)

14. Estimation of mean flow Q2 and flood Q100 (line 252) for 4175 watersheds using only 11 stations (line 255) seems ambitious to me. Spence and Saso (2005) show a significant uncertainty in their predictions. Can the authors provide a statement about their confidence in the Q2 and Q100 estimates?

Spence and Saso (2005) evaluated uncertainty in predicting streamflow using canonical correlation analysis and suggest that Q2 and Q100 estimates could exhibit errors of approaching 50% but exhibited bias of only 13%. We have elaborate on this topic in the text. (29, 1142)

15. What is the reasoning behind the 80% threshold for PCA components (line 279)? Perhaps the authors can include a plot or table that shows the importance of each PC

to support this choice.

The Scree plot in Figure 3 shows the importance of each PC in the analysis. The 80% threshold is commonly used as a cut-off value for PCAs, which informed our decision how to limit PCs considered for these dataset.

16. Were variables standardized to a fixed interval (e.g. [0,1]) in addition to the logtransform (line 282)?

Fractional variables were standardized to a fixed interval because of the nature of the data. However, other variables were not fixed (e.g., elevation).

17. Line 286-287 needs clarification. Which variables are the "complete suite of variables"? The previous section gives the impression that all variables were converted to PCs, of which only those above 80% would be used. A table with a summary of all variables used, their data source(s) and their hydrologic relevance could help clarify what is going on.

We recognize the vagueness of "complete suite". We have included the reference to Table 3 to indicate the variable that were included in the analysis. The sentence now reads: "Clustering analysis was performed on the complete suite of physio-geographic variables, which included PC variables derived from pre-processing (Table 3)." (30, 1179)

18. Retaining PCs above 50% (line 291) seems to contradict retaining PCs above 80% (line 279).

The agglomerative clustering approach requires selecting the number of PCs included in the analysis. This cut-off was chosen based on inspection of the contribution of PCs to the clustering approach and described multiple co-related variables, rather than individual variables, which tends to be the case for increasing PC number. This reasoning is why these two thresholds differ. We have included the following with the intention of being clearer: "Retaining these first PCs at a threshold of 50% allowed for clearer focus on main trends in the data and reduced the impact of noise on subsequent analyses, which might occur if subsequent, less influential, PCs were retained." (30, 1190)

19. A short description of Ward's criterion (line 295) would be helpful.

Thank you for the suggestion. We added additional description as follows: "Ward's criterion decomposes the total inertia of clusters into between and within-group variance, and this method dictates merging for clusters (or watersheds) such that the growth in within-group inertia is minimal (Husson et al. 2010). Within-group inertia represented the homogeneity, or similarity, of watershed within a cluster." (30, 1194)

20. I suggest replacing "and thus did not explicitly affect the clustering analysis" (line 303) with "and are not included in the clustering procedure" (assuming that I correctly interpreted this sentence).

Variables included in the analysis as "supplementary" had their relative location in PCA-space calculated (i.e., eigenvalues were calculated for the variable for each PC). However, they did not impact the PCA directly, which is in contrast to "active" variables. The suggested revision is not completely accurate; we have adjusted our original explanation to mitigate confusion. We have include the following sentence, which is now in the previous paragraph to denote that this step

occurred before the HCPC: "The majority of physiogeographic variables were included as active variables in the PCA and thus influenced the arrangements of the PCs. In contrast, watershed area, DSF, latitude, and longitude were used only as supplementary variables, and thus did not explicitly affect the clustering analysis. These variables did, however, aid in watershed class characterization and interpretation." (30, 1184)

21. Not all readers will be equally familiar with canonical regression analysis. I find it difficult to interpret the results in section 3.1. A (very) brief description of CCA might help. Some questions I'm stuck with: are those lambda values high or low? What would either tell us? What does it mean that hydrologic variables are associated with W2?

We provided more detail in regards to the CCA method and include references where necessary. This concern was shared by other reviewers. (29)

22. I would say these regressions are not particularly convincing (line 314). It looks as if the one high value could be inflating the correlation value. Did the authors use Pearson or Spearman correlations? Predicting streamflow characteristics in ungauged basins (i.e. regionalization) is an active field of study but achieving robust results has proven very difficult. How does this impact the extrapolation of this information to the 4100+ watersheds and what are the consequences for the subsequent analysis?

The bias in this relationship is 29 – 26 %. Perhaps this is to be expected give the small sample size. It is higher than that documented by Spence and Saso (2005) in their study. Content to this point has been added to the manuscript. (29)

23. Section 3.2 (PCA results) lacks a logical conclusion (or perhaps an introduction). How did the authors choose how many PCAs to discuss and which PCAs are selected to be used in subsequent steps?

We intend for this section to provide an account of the main variables associated with the PCs of the compositional dataset. We see these as intermediate steps within our procedure and is intended to provide a brief overview of this preliminary step. We thank the Referee for the suggestion. We have provided elaboration on the clustering PCA as per comment #25 to increase clarity. (32)

24. The difference between active and supplementary variables needs to be defined (line 348).

Thank you for the suggestion. We have clarified the difference between active and supplementary variables in the Methods section as per comment #20. (30, 1187)

25. Section 3.3 lacks a logical conclusion. Which PCAs are carried over to the clustering analysis?

The intention of this section was to describe the PCs and the variables associated with them. We considered it an intermediate step within our procedure, and the 6 PCs were used in the following clustering analysis. We appreciate the reviewers comment, and added sufficient detail to strengthen the relationship between this step and the cluster analysis. This includes a paragraph outlining trends and important characteristics briefly, followed by a more detailed account on the

relationships of individual parameters to each principal component. We have also provided a figure in the supplementary material displaying our workflow to improve clarity (Fig. S1). (31)

26. What do the authors mean with “definition of clusters” (line 370)?

Here, “definition” refers to the distinction of each class. We adjusted the sentence to read: “Further increasing k improved definition refined the separation and definition of clusters up to seven (k=7).” (34, 1315)

27. Section 3.4 is very brief. One of the main aspects of clustering analysis is assessment of how good the resulting clusters are. Currently the authors extensively list the differences between the clusters (section 3.5) by summarising which inputs were most influential in determining the clusters. However, this only tells us something about the patterns in the data and not much about the usefulness of these clusters. The authors suggest in the discussion that these clusters can be helpful to inform management decisions, by showing which regions are expected to behave similarly and which regions are not. This statement should be backed up by proof with independent data that these cluster indeed show that. The GSIM archive (Do et al, HESSD, 2018; https://doi.org/10.5194/essd-10-765-2018) is a recent contribution of global streamflow indices which might provide the authors with independent hydrologic information that they can use to quantify how well their clusters group hydrologically similar regions. See e.g. Knoben et al, WRR, 2018 (linked above) for possible ideas.

We thank the reviewer for this insight. Comparison with independent data was also suggested by Referee #1. We elaborate on this comment at the beginning of our response. We have also included another analysis that compares the robustness of the clustering approach. In addition, we evaluate the applicability of some independent data sources, (e.g., HYDAT, wetland remote sensed data) to compare our classes and the appropriateness of their use, in our responses above and in our Introduction. We also further incorporate the comparison with simulated and observed wetland size distributions. Our intention here is to compare how the classes represent the observed data of the watersheds within each sub region. Streamflow data (from Do et al. 2018) is likely not appropriate for most of the watersheds classes and are not available at the spatial and temporal resolution necessary; although we appreciate the reference to this work. We use the wetland dataset for this purpose. Despite the limitation within these remotely sensed data, we feel it provides a useful application to the prairie regions as well as those regions that are semi-arid or do not possess a well-developed drainage area where streamflow comparisons are not representative.

28. The subsections of section 3.5 are hard work for an international audience. Perhaps figure 5 can be expanded to include a map which shows the various names used in these sections (see e.g. Addor et al, HESS, 2017; figure 1e; https://doi.org/10.5194/hess-21-5293-2017)

We thank the reviewer for their insights regarding readability for an international audience. We point to Fig. 1 for reference to the Provincial names. We also removed reference to more specific and local landmarks (such as Quill and Manitou Lakes). We keep references to the major rivers within this region. (36, 1371)

29. Line 435-437 (“Being river valleys . . . Q2 values (Table 1)) repeats line 428-429.

Thank you for the comment, we have removed the repeated line. (36, 1379)

30. I'm unsure how section 3.6 relates to the previous clustering results. I was under the impression that wetland density is one of the variables used during clustering. Should section 3.6 perhaps be moved to before the clustering results? Also, if this is part of the clustering analysis (as e.g. table 3 and 4 seem to suggest), why does this specific attribute deserve its own section? Edit: reading back, it seems to me that wetland distributions were estimated (line 186 and further). In that case, are the observations referred to in line 480 from the 11 stations? This seems a small sample of observations to compare results for 4100+ watersheds to. How confident can we be in these estimates?

The simulated wetlands by class shown in section 3.6 (Figure 8c) were calculated based on the Generalized Pareto Distribution (GPD) parameters (ξ and β) that were used in the clustering analysis. The wetland density and W_L parameters in panels (a) and (b) were discussed to provide context to the simulated data in panel (c). To clarify, the observed quantiles were based on those from each of the 4100+ wetlands, and the predicted values were from the simulated data based on the GPD parameters. Our intention was to provide an example of how the classes translate to observed data, which is consistent with reviewer suggestions that such an approach could strengthen the study. Specifically, we can predict wetland size distributions from the parameters in the classification, and that the simulated data is relativity consistent with the observed data. We elaborate on the usefulness of these data and our intentions in the discussion. We have also added section 3.4 and 4.4 to be clearer in our intention for this comparison. (page 31 and 38)

31. The authors stress the importance of accounting for human influences (Section 4.1) in classification procedures. Can they comment on the extent to which this was done in their work and do they have any recommendations for future efforts? For example, should artificial drainage density be considered as a variable?

In this regard, data availability at the appropriate geographic scale and spatial resolution is limiting, as we indicate in the text. We incorporate human dimension to a degree, with the inclusion of tillage practices and area of land cropped. Artificial drainage density would be a very useful indicator; however, a comprehensive dataset is not available for the region of interest. We plan to pursue avenues for including a proxy for this parameter in the future. We discuss the usefulness of an artificial drainage estimate in line 761. (page 42)

32. The authors mention that certain variables can dominate the clustering approach (line 579 and further). This is why it is not uncommon to standardize clustering variables to a fixed interval, because this reduces the effect of a variable's variability. Log-transforms lessen, but do not prevent this. Can the authors comment on which variables had the widest (log-transformed) range and whether this correlates with the variables that are most important during clustering?

Thank you for providing the suggestion to compare the impact of fixing variables to an interval. Scaling variables during the PCA was performed in our procedure, which might help to address this concern. In this particular case, such as the fraction of watershed below the outlet, we indicate that despite hydrological importance, a couple variables might not have been indicated as important to characterizing the classes. Our discussion attempted to elude potential overshadowing that might occur. Moreover, if one is particularly interested in such variables, one should consider strategies to weight their importance. It should be noted that the fraction below the outlet was an important variable for Class 5, just that it was not consider highly important to the other classes amongst the various other competing characters. We have adjusted our Discussion section to be clearer in this regard. (pages 43-44)

Response to Referee #3

Please see below for point-by-point comments to Referee #3's suggestions:

Ambiguity: It has been mentioned that the CCA was used for estimating hydrologic variables since only a few observing stations are available. These variables will be considered later in the classification approach to provide a watershed classification system that will be used, among other purposes, to estimate the hydrological response of a given watershed. What is confusing and contradicting here is to first estimating hydrological variables, and then using classification outputs to understand the hydrological behavior! A regionalization approach is more suited for this purpose.

> *In order to reduce the ambiguity we have rewritten this section. The second paragraph now reads:*
>
> *To address this gap mean annual runoff and 1:100 year flood magnitude had to be estimated for each of the 4175 watersheds. Canonical correlation analysis (CCA) was used for this purpose because it was felt that it provided a more independent means of regionalization than using terms directly applied within the subsequent cluster analysis. CCA was used to correlate gauged data to ……"*

I feel inconsistency in using CCA (the most appropriate classification method as recognized in regionalization studies) to estimate hydrological variables, and using another classification method, hierarchical cluster analysis, for classification.

> *As stated above, we needed a method to obtain streamflow terms for each of the 4175 watersheds that was somehow more independent. We believe we have explained why we needed to use a regionalization method to estimate Q2 or Q100, but the objective of the study was to classify the watersheds, and the hierarchical cluster analysis is a more appropriate tool.*

Equation in Line 319 is not very convincing since no precipitation or water-related variable is introduced.

> *One is not necessarily required. The canonical correlation coefficients imply Q2 can be estimated with confidence using these terms and with the values in the equation.*

Also, only 11 observations have been considered for calibration. Assessment of the uncertainty is not consistent too.

> *We felt an uncertainty assessment of the equation in Line 319 was unnecessary because of how the estimate of Q2 was used. To do so would have meant an uncertainty analysis could have been required for every other input into the cluster analysis, which was beyond the scope of the paper.*