Response to Reviewer #1

Summary

This study aims to demonstrate the utility of a hybrid classification and virtual modelling approach for assessing the sensitivity of a portion of Canadian Prairie catchments to climate. They first developed a class-based virtual basin model for a portion of the Canadian Prairie and then used the virtual basin to explore the sensitivity of hydrological response to climate.

Assessment

I enjoyed reading this manuscript. Class-based virtual basin modeling approach is an innovative tool to roughly identify regional-scale landscape vulnerability. I find the manuscript provides some interesting material for readers of HESS. This seems an appropriate topic to further generalize and extrapolate basins hydrologic functions. However, I think this paper needs major revisions and clarification on the rational of the study, and manuscript presentation and methodology. After these issues are resolved, I believe this paper could be a nice addition to HESS.

Thanks for the comments and encouragement to continue with the review process.

Major Points:

This paper uses only one class of basin developed in the classification approach published by the authors before. This class includes a portion of Canadian PPR. Almost up to the middle of the paper this point is not very clear.

As an example, in line 109-111 the objective of the paper was explained as: "This paper aims to demonstrate the utility of a basin classification–based virtual modelling approach for assessing the sensitivity of Canadian Prairie catchments to climate". But they only explored one class out of 7 classes of catchments within Canadian PPR. I suggest that they explain only one class throughout the paper and focus the text to this class only.

We are explicit that we focus on the High Elevation Grasslands class early in the paper (Lines 143 – 155) We can address this comment by better weaving in "the High Elevation Grassland virtual basin" into the text more, rather than only say "the virtual basin".

The point that the classification approach they developed in the previous published work is insensitive to inclusion of climatic features could show the inability of classification approach in categorizing basins based in their functional similarities. The timing, rate, frequency and type of precipitation could be the most dominant factors controlling how a landscape functions irrespective of their physical characteristics.

For evidence that the classification approach was not insensitive (but not for inclusion in the paper) below we list the areas of each catchment class, and their differences. Some classes were not much different (e.g., Sloped incised and High Elevation Grasslands), but some changed by 20% or more (e.g., Interior Grasslands and Southern Manitoba). This provides some comfort that the classification approach was able to capture nuances that differentiate functionality. Climate can be a dominant factor controlling hydrology, agreed. However, we chose not to include climate in this classification as we would be varying and manipulating the climate to explore the variability of current hydrological response and sensitivity to future climates in the region. We will remove the sentence on Line 139 because it is misleading, as yes, the two classifications were similar, but different in several ways. Furthermore, we will rewrite this content at the end of this paragraph to be "This was done because including climate could bias the results of the climate scenarios by "hard wiring" some functionality into the model parameters."

	This papers delineation (km^2)	Wolfe et al. 2019 (km^2)	Fractional difference
High Elevation Grasslands	79667.52	75176.03	0.94
Interior Grasslands	47211.54	57812.5	1.22
Major River Valleys	34533.41	39892.18	1.15
Pothole Glaciolacustrine	77844.3	70399.25	0.90
Pothole Till	120881.2	106541.9	0.88
Sloped Incised	35388.6	34432.65	0.97
Southern Manitoba	21149.17	32421.21	1.53

The major assumption of the paper is that the authors previously published classification approach can classify landscapes based on (dis)similarity in their hydrologic functions. However, Oudin et al. [2010] clearly demonstrated that physically similar catchments, with similar physical features, are not truly functionally similar within the context of basin classification. As they used physical features to classify catchments, it is not clear if their classification is functional.

Further to our points above, we can cite a new paper that used hydrological response to classify watersheds on the Prairie (Whitfield et al., 2021) and how it qualitatively aligns well with our classification. This paper's classification did include streamflow in its calculations and had the purpose of designing the virtual basin model structure and for its parameter selection. We will also reference Wolfe et al. (2019) which showed that the approach to classification has the capability to be functional.

If one of the objective of the paper is to demonstrate the utility of a new regionalization approach (e.g., Class-based virtual basin modeling approach) below points should be clarified and discussed in the introduction, method, and discussion sections:

The rational for needing such approach: In which way the approach works better than other regionalization approaches in the literature (see Blöschl et al. [2014] for details of different regionalization approach).

We can explain this further in the first section in methodology as we describe the framework. In particular, we will include content such as "The low density of gauged non-regulated streams in the region make extrapolation of observed streamflow data highly uncertain."

How computationally efficient is the method? Is it faster than, for example, the approach suggested in Knoben et al. [2018] that only used three functional indices to globally regionalize basins streamflow signatures. If not, why we have to use the proposed method? Does it respect functional behaviour of basins stronger than other methods? How accurate is their method compared to the other regionalization schemes? The paper shows some graphs related to the accuracy of the method. But there is no quantitative and/or qualitative comparison with other methods.

We interpret this comment as not so much about the modelling as much as it is the classification, which is not the point of this paper. In our defense, Knoben et al.'s paper used three indices, but applied them across the globe. It is arguable that much of the Prairie could not be differentiated using those three indices (aridity, the seasonal range of water-versus-energy availability, and the fraction of snowfall), as they are similar enough across the Canadian Prairie, relative to the globe, and this is reflected in the figures within Knoben et al.. There are important differences in small basins across the Canadian Prairies, differences due to topography, soils, depressional storage, runoff connectivity within the basin and groundwater connectivity to surface hydrology. This classification was intended to elucidate and define these differences for the purpose of hydrological model design and parameterisation.

We can include a sentences such as "A more thorough detailed approach is needed within this region than methods such as Knoben et al. (2018) which are too coarse to differentiate functional differences. This classification was intended to elucidate and define these differences for the purpose of hydrological model design and parameterisation."

There is a dearth of literature on the impacts of climate change and wetland drainage on streamflows of PPR watersheds in both USA and Canada. The authors mostly referred to their previous papers in this regard. In both introduction and discussion, other groups' works must be acknowledged. It should be clarified, how their findings are (in)consistent with other works.

Will we expand on this more in the Introduction (framing the problem) and Discussion (context of other studies). We thought we were keeping the self-citations down, but we will expand on current content in Lines 568 – 592 that cites other groups' work papers by including papers that Golden and Johnson wrote in the region in nearby North Dakota.

Minor Points:

Line 44: How long does it take to run one model. How long would it have taken to run a model for every catchment in the database or every catchment in the cluster? No talk of efficacy of virtual experiments in discussion, yet it seems like one of the central purposes of your paper in the introduction.

For each 46 year run (1960 to 2006), the model takes around 7 minutes on a PC with CPU of Intel Core i7-8650U. One reason we could not run the model for each of the 796 catchments in the HEG class is that we could not find a decent gridded climate dataset (we tried three). So, we are currently running the model based on the availability of station observations. This resulted in the seven AHCCD stations. With 35 climate scenarios run for seven stations, it took 7 hours to complete all the simulations. We can add these details to the methodology and add a paragraph about efficiency to the discussion.

Line 139: Why would there be bias?

We can add content to address this question such as: "The influence of recent climate, historical climate change and how this interacts with Prairie biogeophysical features and drainage to influence basin hydrology behaviour has been recently explored by Whitfield et al. (2021). They showed the coherence between climate effects and biogeophysical and drainage effects on basin classification. Therefore, bias could be introduced from a classification that includes climate when applying future climates to force the virtual basin model. For instance, if a current climate classification resulted in a virtual basin class that precluded semi-arid basins, this would become non-functional with the projected increase in precipitation and temperature for the region which could change the area to semi-arid. By excluding the direct impact of climate in the classification and restricting this to topography, biogeophysical features and drainage systems then the new climate of a set of basins can be more fully explored."

Whitfield et al., 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, Hydrology and Earth System Sciences, 25, (2021), 2513:2541, doi: 10.5194/hess-25-2513-2021.

Lines 209-210: Why is the maximum threshold less than minimum?

We missed mentioning the minimum is for 100% rainfall. We can fix this by editing the sentence to "… for 100% snow and a minimum temperature threshold of +2°C for 100% rainfall and distributes …..

Line 193: (Shook et al., 2013) lots of other places where comma is missing as well

We will fix these.

Line 283: Why don't you look at cluster center (compare median catchment with virtual model)

We can add a sentence such as; "The cluster median was not used for comparison, as the range as represented by snow on the ground (Figure 4) and streamflow (Figure 5) better demonstrates that model simulations are within the envelope of plausible hydrological functioning rather than comparable to an individual basin within the class."

Line 434: To me advance in max snow depth date implies further in time, i.e., closer to summer. Rephrase.

We believe that in many studies looking at the day of year of maximum snow water equivalent, or snowmelt or snow free dates, advance implies it is arriving earlier. We can rephrase to "The date of annual SWE peak (with no change in precipitation) arrives 40 days earlier."

Line 517: should it be "temperatures , but these patterns"?

Respectfully, no. The patterns that vary across the region reflect how precipitation may cancel the direction and slow the magnitude of streamflow changes caused by increased temperatures.

Blöschl, G., M. Sivapalan, M. Wagener, A. Viglione, and H. Savenije (2014), Runoff Prediction in Ungauged Basins, Cambridge University Press.

Knoben, W. J. M., R. A. Woods, and J. E. Freer (2018), A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data, Water Resources Research, 54(7), 5088-5109.

Oudin, L., A. Kay, V. Andréassian, and C. Perrin (2010), Are seemingly physically similar catchments truly hydrologically similar?, Water Resources Research, 46(11).