**Author response to reviewer #3 comments for HESS manuscript "Modelling the effects of climate and landcover change on the hydrologic regime of a snowmelt-dominated montane catchment" [Paper #: hess-2024-361]**

Dear Reviewer,

We would like to thank you for taking the time to complete a thorough review of our manuscript submission. You raised many important issues that will certainly result in a stronger manuscript. We disagree with some of your comments, but appreciate your thoughtfulness and value the constructive scientific discussion. Please find below a list of responses to your comments. We hope our responses satisfy the spirit and intent of your remarks.

Sincerely,

Russell Smith

**Reviewer #3 comments**

**Summary**

The authors describe a study to explore the effects of climate change and landcover change on the hydrology (SWE and various discharge variables) of a forested nival watershed in the interior of British Columbia. Results are derived via simulation using hydrologic modelling in combination with various forest disturbance scenarios and mid- and end-century climate projections. The research deals with an important and timely issue, particularly considering the dramatic increase in wildfire risk experienced in western North America. However, the work suffers from the strictly qualitative, predominantly graphical, approach to both analyse and interpret the results. This results in a work that generally lacks any real scientific impact. This is particularly puzzling considering that the experimental design employed (full factorial) and the available data lend themselves well to the application of common statistical techniques (ANOVA and regression) that could give quantitative results and analysis of individual and combined effects of climate and forest cover. This is a missed opportunity. I feel that with a more quantitative assessment of results, this work would make a worthy scientific contribution; hence, my recommendation is to accept but with major revisions.

- Thank you for acknowledging that the research deals with an important and timely issue. We agree!
- We respectfully disagree with your point that the work suffers from a strictly qualitative approach to analyzing and interpreting the results. We acknowledge that we followed a conventional/traditional approach in hydrology to presenting and interpreting the results; however, graphical plots are a means of presenting quantitative results, not qualitative results. Moreover, we discuss the patterns and effects quantitatively throughout the manuscript. If our work lacks any real scientific impact because of the approach to analyzing and presenting the results, then it should also be said that much of the existing hydrology literature also lacks any real scientific impact.
- Further to the points above, our work incorporated a substantive analysis in the frequency domain of peak flows, low flows, and mean annual discharge (i.e., annual yield). We're not aware of other hydrology literature that has examined all three of these hydrologic variables in this way. This approach led to valuable insights regarding contrasting effects on hydrology and associated watershed risks. We related these results to changes in climate, snowpack dynamics, and vegetation, which led to identifying mitigative influences. Moreover, three other peer reviewers have not had a concern with our overall analytical approach. We're confident that our work is scientifically meaningful.
- We acknowledge that the approach you've proposed for analyzing the model outputs is interesting and novel; however, not following your suggested approach certainly does not mean that our work lacks complexity or impact. Further, we have some uncertainties about the validity of your proposal:
    - It seems you are suggesting ANOVA and regression as a means of hypothesis testing. We question whether it would be statistically valid to do so, as outputs from a pre-calibrated deterministic model cannot be considered random samples from a population.
    - We can see some value in using ANOVA or regression in a descriptive sense to help understand the model; however, we question the utility of that approach when we have the luxury of simply modifying the climate forcings and/or landcover, and examining the predicted effects on hydrology directly. This is a luxury that does not exist with field studies, making ANOVA and/or regression necessary/appropriate in those cases; however, with field studies, it is also important to ensure a randomized sampling design.

**Major Issues**

**Experimental Design**

The experimental design, which is central to the study, is described in an ad hoc fashion. The authors are encouraged to re-organize the document so that the experimental design is clearer. And in fact, it should be noted that the authors employ what is formally known as a factorial experimental design, where multiple factors (climate state/change, disturbance level, elevation, etc.) are tested for their influence on the outcome of a response variable (peak flow, max SWE, melt-out date, etc.). The experiments could be described as follows:

- Phase 1:
  - a 13 x 2 design (26 categories) with 13 x climate states and 2 x disturbance levels
- Phase 2:
  - SWE: a 2 x 2 x 3 x 3 design (36 categories) with 2 x slope aspects (north and south), 2 x disturbance level (forest and clearing), 3 x elevations (low, middle and high) and 3 x climate states (current, 2050s and 2080s
  - Discharge: a 5 x 3 design (15 categories) with 5 x disturbance levels and 3 x climate states

Although the factorial design could have been exploited to formally explore main effects and interactions between the various factors, using such methods as ANOVA and regression, the authors opted instead for a qualitative and graphical approach and, I feel, missed an opportunity for a more impactful study.

- Thank you for your ideas on how to describe the experimental design. We will revise the text to describe the experimental design more clearly near the beginning of the manuscript.

I am also not convinced that the graphical results can be used to isolate each individual effect (disturbance and climate) as stated. For example, the 'disturbance effect' in Figure 10 shows the effect of each disturbance conditional upon various climate states. I.e. the disturbance effect is not independent of climate state.

- We agree that the disturbance effect is conditional upon various climate states. That was a key point of the analysis and interpretation, and the reason we presented the results the way we did – that is, to examine how disturbance effects vary with climate, and vice versa for climate effects.

As said prior, formal statistical approaches could be used to quantify the effect of each individual factor as well as the various combined effects. The authors could also take their existing results one step further. As each synthetic climate series is stationary, each 100-year series could be divided into non-overlapping decadal periods that would provide ten replicates per category, producing a full factorial design with replicates. Any experimental design text will explain how to more formally exploit this approach.

- You've proposed using formal statistical approaches to quantify various effects; however, it is important to acknowledge that the dependency of the disturbance effect on climate would need to be examined through incorporating an interaction effect (e.g., in the case of a regression model). To any extent the model residuals might violate the various requirements of regression (randomness, homoscedasticity, normality, independence, lack of correlation with the independent variable), the results and inferences of the regression modelling would be weakened, invalid, and/or biased/skewed. In contrast, directly examining the modelling outputs from various model perturbations leads to a more direct and unbiased interpretation of the various effects as represented by the catchment model.

**Phases:**

I am not clear on the requirement of the Phase 1 portion of the methodology. It seems that the main purpose was to select the climate experiment to be used for Phase 2. If so, the authors should be aware that they, perhaps inadvertently, chose the 'driest' scenario (see Table 1); only CSIRO85 projects slightly decreased precipitation in the 2050s and has the smallest precipitation increase in the 2080s. I would highly recommend that the authors redo the Phase 2 analysis with at least one additional climate experiment (perhaps the wettest one, e.g. MIROC85 or MPI45).

- This manuscript is a resubmission of a paper that was originally submitted to HESS in the fall of 2023 and later rejected (hess-2023-248). The original manuscript was submitted with the Phase 2 results only (i.e., only CSIRO85 had been modelled because of limitations in available funding). The original reviewers requested modelling of additional climate projections. We obtained additional funding and completed the modelling of four additional climate projections, and incorporated those results with CSIRO85 into Phase 1. The paper was already very long with many results to synthesize and package for the reader; thus, we did not think it would work well to complete a detailed (i.e., Phase 2) analysis of the results from all five climate models. We also did not have funding available for that amount of additional analysis. As it turned out, however, CSIRO85 generated climate effects on the spring freshet hydrograph that were generally intermediary between the other climate projections in relation to peak flow timing and discharge, and hydrograph elongation; thus, we saw it as appropriate to present the CSIRO85 results for the detailed examination of disturbance and climate effects (Phase 2).

- We believe CSIRO85 was the most suitable selection for Phase 2, for the reasons discussed above. However, we acknowledge that CSIRO85 is relatively “dry” and, thus, might result in more severe climate effects on low flows compared to other climate projections, for instance. We are prepared to complete event frequency analyses for the outputs from the other four climate projections. We are hesitant to include a full presentation of these additional results in the main body of the manuscript because of the substantial length of the existing manuscript; however, the additional event frequency analyses could be provided in the supplement and discussed briefly in the uncertainty section (6.4).

**Minor Issues**

Line 24: Not sure ‘three dimensional’ is a suitable term for what you are describing. Perhaps ‘multi-facetted’?

- Thank you for the suggestion. We will consider other language options.

Line 146: More specifically, Raven estimates cloud cover, etc. from T and P. The term ‘accounts for’ is a tad vague.

- Agreed. We will revise the wording to reflect that these components are estimated.

Section 2.2.2.1: If available, why wasn’t data from the Penticton Airport station used directly as a model forcing?

- The model was originally set up to use both P1 and Penticton Airport data for model forcing; however, the Penticton Airport data generated nonstationarity in the model parameters over the ~45 year optimization record that could not be resolved. The nonstationarity was related to the temperature data. We inferred that the nonstationarity might have been related to urbanization, or possibly an instrumentation issue at the station. In either case, the nonstationarity did not occur when only P1 data were utilized.

Line 173: You are using incorrect nomenclature. What you are describing as “emissions pathways” are what should be called climate projections (i.e. a projection is produced from a combination of an emission scenario and a global climate model). Your design only includes two, not five, emissions pathways/scenarios, RCP4.5. and RCP8.5.

- Thank you for pointing this out. We will adjust our language accordingly.

Line 174: That bound “90% of projections” for what variable(s) over what region?

- They bound 90% of the air temperature and precipitation range of CMIP5 projections for the southern Okanagan of British Columbia (Supplementary material, Spittlehouse & Dymond, 2022). We will add this information and citation to the text.

Line 194: Confirm the adjustment is made seasonally, and not monthly.

- Seasonally (Spittlehouse & Dymond, 2022)

Line 199: What parameters are used, and then presumably adjusted, to describe the distribution of tasmin, tasmax, and precipitation.

- Daily values. Only Tmax and Tmin need adjustment.

Lines 249-254: The process being described seems to be better explained as a process if intersecting various layers as opposed to imprinting. In other words, discretizing HRUs is the process of intersecting five individual layers: sub-basins, BEC variants, disturbance history, vegetation type, and 2K x 2K grid (to limit HRU size to <= 4-km2).

- Agreed. We will revise the text accordingly. Thank you for the suggested language.

Lines 279-285: The use of the word “constrained” implies that these parameter values were adjusted during calibration. Do you really mean the values were estimated a- priori using observations?

- We’re meaning that the parameters were calibrated on historical data and, thus, their values were constrained by the need to fit the simulated outputs to the historical observations. However, our language has generated some confusion among multiple reviewers. We will revise the text to be more specific with respect to assigning parameter values versus setting calibration ranges and calibrating on observed data.

Lines 301-306: This section implies the LAI and crown closure are closely related, however, in my mind they describe different vegetation characteristics. LAI is a vegetation property (i.e. leaf area density for individual plants) whereas crown closure is a stand property (the density of individual trees/plants). The authors seem to have conflated the two properties. Is LAI, then, a combined value of vegetation and tree density?

- Good question. Thank you for pointing this out. LAI and canopy closure do not specifically represent density, but are correlated to each other and to tree density. LAI can be both a property for individual plants or a stand property when averaged across space, just as tree height can express the height of an individual tree or the mean height of a stand (or a particular layer in the stand). For catchment modelling with Raven, the specific LAI related parameter is the maximum LAI of the vegetation type (i.e., stand level). We will revise the text to clarify that the specific parameter is maximum LAI, and that it relates to the vegetation type as a whole (i.e., stand level).

Line 329: Which years were used for calibration and validation?

- The years are shown in the supplement on the respective calibration plots (Figures S2.1 through S2.5) and validation plots (Figures S3.1 through S3.5). We will insert a reference to the supplement.

Line 332: How was the composite function constructed from the various indicators (I’m assuming it was the arithmetic mean). Were the individual metrics weighted? Show the mathematical description of the composite function.

- The composite function was calculated as the arithmetic mean of all components, and all components were assigned an equal weighting. These points will be added to footnote #1 for Table S3, and added to the text in the main body.

Line 350: Replace “emissions pathways” with “climate projections”.

- We will make that change. Thank you.

Line 353-354: Would recommend instead saying “For each 100-year simulation the landcover state was static and the meteorology derives from a stationary climate state”

- Thank you for the suggested language. We will revise the text accordingly.

Line 356: May use “dynamic equilibrium” instead of “wet”.

- We think that dynamic equilibrium would make sense to people with experience in hydrologic modelling, but might generate confusion for others. We propose revising the text to the following: “The first year of simulation was used as a warm-up (i.e., spin-up) to ensure soil wetness had reached a dynamic equilibrium (i.e., “wet”) leading into the subsequent simulation years.”

Section 2.5: There are two experiments being conducted to assess climate and landcover change on SWE, a point-scale experiment and a catchment-scale experiment. It’s not clear, however, how the two are used, or which is experiment is being referred to in the results sections (5.1.2 and 5.1.3). For the point-scale experiment, where are the sample sites located? Which experiments supply the results for Table 5 and Figure 5? Are both experiments necessary?

- With respect to Section 5.1.3, both experiments (point scale and catchment scale) are a sensitivity analysis, as described in Section 2.5. As a sensitivity analysis, the point scale experiment does not represent specific locations in the watershed; however, it represents site conditions that are realistic for the watershed. Figure 5 is based on the point scale experiment. We will revise the text to clarify where point scale versus catchment scale data are being presented/discussed.
- Section 5.1.2 and Table 5 (net precipitation) are based on the same high elevation sites as those in the point-scale snowpack results (Figure 5). We will revise the text to clarify this point.
- We believe the point scale and catchment scale outputs are both valuable. The catchment scale plots show very clear changes for different climates, and help to convey a sense of broad, landscape scale changes. The point scale results facilitate a focus on specific contrasts, and lead to identifying specific mitigating influences.

Line 377: Day numbers are hard to interpret. It would be helpful to add the corresponding calendar dates, 172 = June 21 and 264 = Sep 21. Or just give the dates instead.

- The date associated with a specific day of year varies from year to year. We propose revising the text to the following: “lowest 30-day mean discharge between day of year 172 through 264 (approximately June 21 to Sept 21).

Line 385: Incorrect reference.

- This is the guidance provided in Section 1.4 of the Raven user’s manual:
    - To cite Raven technical details for technical reports, this manual may be cited as:
        - Craig, J.R., and the Raven Development Team, Raven user’s and developer’s manual (Version 3.5), URL: https://raven.uwaterloo.ca/ (Accessed xxx, 2022).
- We will add a date of access to our reference.

Line 478: Both crown closure and LAI are used interchangeably to describe stand density. Terminology needs to be clarified and unified across the text.

- Good point. Both can be used to describe vegetation density at the stand scale, and the two variables are correlated. Generally, LAI is more impactful in the model than canopy closure, which is why LAI was referenced in line 478 and elsewhere. However, canopy closure was used to represent stand density in classifying mature stands for discretizing HRUs (lines 294-299), as LAI is not represented in the catchment-wide VRI forest cover mapping. This is the only section where canopy closure is referenced as a variable for stand density. We will revise the text to clarify these points, and to state that LAI is otherwise used to describe stand density in the analysis.

Line 503: This conclusion is not obvious from the results.

- In Figures 5b, 6e, and 6f, the disturbance effect is positive for almost all sites, but small or negligible for some sites at the low elevation. We will revise the text to reference these figures, and to clarify the exception for the low elevation.

Lines 515-516: This conclusion if also not obvious from the available figures.

- We agree. More specifically, the disturbance effect is variable at the middle elevation in the catchment scale analysis, and comparing to the associated canopy closure (Figure 2d) is not straightforward. Notwithstanding, with respect to mitigating influences, we believe it is helpful to communicate that the greatest disturbance effects in the middle elevations were associated with high density stands located in the MS zone, based on a detailed review of the output files.
- We propose revising the text to the following: “At the middle elevation, this disturbance effect was minimal for the vegetation types represented in the site scale analysis (i.e., moderate density P/F; Fig. 5e), and variable in the catchment scale analysis (Fig. 7d). The greatest disturbance effects in the middle elevations were associated with high density stands located in the MS zone (based on a detailed comparison of the data for Fig. 7d and Fig. 2d).”

Line 533: I don’t think the figures support this conclusion that clearly.

- We agree. Comparing to the associated canopy closure (Figure 2d) is not straightforward. We propose revising the text to the following: “The climate effect was greatest for lower elevations, and for higher stand densities at middle and higher elevations (based on a detailed comparison of the data for Fig. 7b-c and Fig. 2d)”.

**Tables**

Table 2: Do the LAI values in this table reflect the spatial variation in crown closure? Are LAI and crown closure correlated (see earlier comment)?

- There is a positive correlation between LAI and canopy closure. As described in the manuscript and discussed above, canopy closure was used for discretizing the HRUs. The relationship between LAI and canopy closure from the vegetation surveys was used for setting calibration ranges for LAI, then LAI was calibrated on SWE and discharge. The LAI values in Table 2 generally reflect the spatial variation in canopy closure, but not based on a specific equation/relation between the two. As discussed above, LAI is more impactful in the model than canopy closure, which is why LAI is presented in Table 2.

Table 4: Indicate in the table header weather the variable is a mean (Winter, Summer) or a maximum (Spring).

- Good point. It's the long-term mean of the total for winter and summer, and the long-term mean of the maximum for the specified duration for spring. We will revise the caption to clarify, and add info to the table header.

Table 5: Which experiment are these results from?

- Point scale, as discussed above.

**Figures**

Figure 5: Which experiment (point or catchment) do these results derive from? The categories (y-axis) on the panels take a while to interpret as they change between columns and are not explicitly identified. Also note that the 'Disturbance effect" and "Climate effect" are not strictly independent (disturbance effect is conditional on climate state and climate effect is conditional on disturbance state). Do the bars show the mean or the median (assuming catchment scale results) or single sites (point scale results).

- Please see response above regarding data independence.
- The existing caption states that these results are from the site scale snowpack sensitivity analysis, and that the data are the mean of annual maximum SWE (mm) and median timing of snowpack melt-out (day of year).
- Multiple reviewers have expressed concern regarding the interpretability of this figure and Figure 10. We will revise and add to the y-axis labelling, and make some rearrangements to the panel layout for clarity (e.g., swapping the disturbance effect and climate effect columns). We will also experiment with different bar fills and thicknesses to hopefully improve clarity.

Figure 10: The categories (y-axis) on the panels take a while to interpret as they change between columns and are not explicitly identified. Also note that the 'Disturbance effect" and "Climate effect" are not strictly independent (disturbance effect is conditional on climate state and climate effect is conditional on disturbance state).

- Please see response above for Figure 5.

Figure 12: Decimals missing on secondary y-axis labels. Recommend showing the summer period on the graph (i.e. as background shading).

- We assume you're referring to the missing decimals on the discharge y-axis. Thank you for noticing that error. We will correct it.
- Good idea about adding the shading. Thank you.