Author response to reviewer #1 comments for HESS manuscript "Controls on groundwater response and runoff source area dynamics in a snowmelt-dominated montane catchment" [Paper #: hessd-10-2549-2013]

Dear Sir or Madam,

We would like to thank you for your taking the time to complete such a thorough and thoughtful review of the original manuscript submission. Addressing your comments will undoubtedly result in a stronger submission. 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 #1 comments

General Comments

I find the overall presentation of the material cluttered and hard to follow. I think many of the sections could be reorganized and shortened to improve clarity for the reader. I also have some concerns about whether or not the results are sufficient to support the papers conclusions regarding lateral vs. vertical flow and hydrologic connectivity. In particular, I am concerned that given the deep soils in the catchment, that sites that were deemed unresponsive or transient may not have been installed deep enough to measure the water table elevation. I also am not clear on how the authors are defining hydrologic connectivity in such a large catchment, which is a major theme of the discussion.

- The concerns identified above are repeated in greater detail below, along with our responses.

Specific Comments

1. Throughout the manuscript, the word 'persistence' seems to be used interchangeably with 'duration' (e.g. P2550, L13). I recommend just using 'duration' since it is presented as one of the response variables in section 2.3.3.

- We agree that using "persistent" in place of "duration" could be confusing in some instances. We will abandon its use as a synonym for "duration", but retain it as an adverb (e.g. "persistent responses") wherever it results

in clearer wording.

2. In the abstract, the authors state that “Runoff source areas expand and contract throughout these periods according to an interplay between catchment wetness and the spatial patterns of topographic convergence.” I find “interplay” too vague a term and suggest the authors be more specific about the interaction between catchment wetness and topographic convergence.

- We suggest the following wording: “Runoff source areas expand and contract throughout these periods coincident with catchment wetting and drying, and follow the general spatial patterns of topographic convergence.”

3. On P2550, L16 of the abstract, the word “differential” is used. I think the authors should be more specific about what type of variability they are referring to (spatial, temporal).

- We suggest the following wording: “However, spatial controls on the timing, intensity, and quantity of snowmelt and controls on vertical versus lateral flux partitioning in the soil overwhelm the influence of topographic convergence on runoff source area dynamics during early spring freshet periods.”

4. When discussing the ‘fill and spill’ concept, the authors should reference Spence and Woo (2003) where the term was first used.

- Good point. The text will be revised accordingly.

5. Are there other studies that can be referenced along with Redding and Devito (2008, 2010) on P2552, L3. While these two papers do discuss the mechanisms involved in lateral and vertical flow in deep glacial soils, they do so by way of an irrigation experiment, which creates rare conditions (in the Boreal Plain) of soil saturation and high intensity precipitation events. Haught and van Meerveld (2011) may be appropriate here or elsewhere in the Introduction, although their soils are relatively thin compared to this study. But, this paper doesn’t really address flow paths deeper than 2m.

- Good point. Thank you for the reference. We will try to find additional suitable references to cite.

6. On P2552, L4 and elsewhere, the authors use the term ‘water inputs’. I think it

would be helpful to be more specific here that you are talking mainly about spatial variability in snowmelt inputs.

- We use the term “water inputs” on purpose to avoid differentiating snowmelt versus rainfall inputs. The catchment runoff regime is snowmelt-dominated; however, rainfall inputs are common during the spring, summer, and fall, and contribute to the space-time dynamics of groundwater response. Since our analysis was focused on the annual regime rather than individual runoff events, we believe it is appropriate to treat both inputs collectively.

7. In the last paragraph of the Introduction, the authors lay out their six hypotheses. I think that this paragraph should also include a clear statement of the overall goal and specific objectives of the study. While this can be inferred from the preceding discussion of knowledge gaps in the area of groundwater dynamics in large snowmelt dominated catchments, I think it is best to be explicit about what is in the current study.

- Agreed. The paragraph will be re-framed as a statement of objectives and the hypotheses will be incorporated as conclusions, wherever appropriate.

8. Be specific about what the ‘other factors’ are in your first hypothesis statement.

- See response to comment 7.

9. I would like to see some foundational discussion in the Introduction to support the third hypothesis statement regarding intra-annual variability of groundwater response. It sort of pops up out of nowhere.

- See response to comment 7.

10. The fourth hypothesis statement is too wordy and hard to follow.

- See response to comment 7.

11. The term ‘groundwater response’ should be defined earlier in the Introduction (probably in the second last paragraph on P2553). Does this mean that the water table is measured in a well?

- As defined on lines 13 and 14 of page 2553, “groundwater” was used to refer to phreatic water regardless of the depth below the soil surface. For

observed data, it meant that a water table was measured within the well. In response to your suggestion, the text will be revised to clarify this point in the introduction.

12. I think there are too many hypotheses being tested here. Can some of them be combined? Hypothesis 3 seems redundant.

- See response to comment 7.

13. In section 2.1, where the authors give an estimated range of ET, I would like more information on how these values were modelled. If they don't want to take up space describing it here, the information could be moved to the Supplementary Material.

- We revised the estimation of ET using a simple catchment water balance calculation (i.e. precipitation minus runoff). Annual ET is approximately 400 mm based on three years of data. The text will be revised to incorporate this information.

14. On P2555, L21 the authors state that spring snowmelt dominates the hydrologic regime. It would be useful to quantify this somehow, e.g. what percentage of annual flow is from snowmelt?

- Good point. Determining the portion of annual flow from snowmelt requires detailed catchment modelling, which is currently in progress. However, approximately 65% of the annual precipitation in the catchment falls as snow. The portion of annual flow from snowmelt would likely be at least as high due to the relatively high runoff efficiency during snowmelt periods compared to snow-free periods. We will revise the text to incorporate this information.

15. P2555, L23-25: It would be helpful to know how many years this range covers.

- The 150-600 mm values represent the range of peak snowpack conditions in the catchment during a typical year, which was determined based on three years of data. This point will be clarified in the text.

16. P2556, L13-15: I do not understand what is meant by ": : : based on the USDA soil classification system (Smith, 2011)." No classification information is given here, just data on soil texture and structure. If some sort of classification is

intended, why isn't the Canadian System of Soil Classification used?

- The soil make-up was quantified through particle size analysis using sieve and sedigraph methods. The particle size classes were defined based on the USDA soil classification system, which is in common use throughout North America for particle size analysis. The text will be revised to clarify these points.

17. P2556, L15-19: Where were observations (or lack thereof) of soil macropores or cracks made? Soil pits? Road cuts? This should be stated. Also, is it really possible to visually assess the abundance of burrowing animals and insects over such a large catchment?

- The primary author visually observed soil characteristics while hand-digging over 400 soil pits with depths ranging from 0.41 m to 1.64 m, and along several kilometers of new forestry road (road cut depths ranged from 2 m to 8 m) that were constructed in the catchment at the end of the study. He was present in the catchment for approximately 315 days, and visited the 50 hillslope monitoring sites and seven streamflow sites (which were all well distributed throughout the catchment) regularly. Moreover, data from the extensive field campaign to measure hydraulic conductivity profiles at all 50 hillslope monitoring sites provided numerous observations of the porosity and macropore structure of soils within the catchment. We consider this an exhaustive effort to characterize the soils. The text will be revised to include this information.

18. P2557, L22: Was more than one well installed at each site?

- No, only one well. This point will be clarified in the text.

19. P2557, L25: Does 'groundwater initiation' mean the water table moves up into the well and/or an existing water table elevation changes? Please clarify.

- "Groundwater initiation" means that a water table developed within the well either by rising from below or from development of a perched water table. This point will be clarified in the text.

20. I think that the authors need to provide a bit more information on why wells were installed in soil pits that were then backfilled. Depending on the depth of the pits there could be significant disturbance to the soil structure surrounding the well, no?

- That's a good point. It was impossible to bring heavy machinery to the sites due to access issues and the depth of hand augering was restricted by the high coarse fragment content of the soils. As a result, wells were installed in soil pits that were hand dug approximately 30 cm in diameter. They were carefully backfilled with the native soil ensuring the same soil layering and avoiding compacting, which resulted in porosities that were similar to that of the original soils. As mentioned in the manuscript, up to two additional attempts were made to increase the depths of the wells at sites with limited or no groundwater responses, including installation of stainless steel drive-point wells using a sledgehammer (15 out of 50 sites, in total). Soil disturbance around the drive-point wells would have been minimal. In either case, we believe the water table in the disturbed soil equilibrated with the surrounding water table based on the volume of disturbed soil relative to the size of the hillslopes and the amount of water input.

21. P2558, L9-11: Specify what water table depth was measured relative to. I assume it is the ground surface or a benchmark, but be specific.

- It was measured relative to the soil surface. This point will be added to the text.

22. P2558, L19-22: Be specific about what depths soil saturation was measured at.

- Soil saturation was measured manually at 10 cm depth intervals to the maximum installation depth at the site. The mean installation depth among all sites was 66 cm, and varied from 40 cm to 90 cm depending on the size and abundance of coarse fragments. We will revise the text to provide this information.

23. P2558, L26,27: I think that all methodology should be outlined in this paper or its Supplementary Material. Realistically, not many people are going to want to dig up a PhD thesis to find more info.

- We provided all the necessary information that we could think of and included the reference for convenience; however, we would be happy to include other relevant information, if necessary. Moreover, we'll reference the URL for the dissertation so it can be accessed easily.

24a. I don't think Table 1 is necessary. It repeats a lot of the information already given in the Methods section and anything extra could be easily added to the text

or Supp. Mat.

- We believe it's a convenient way of presenting key facts about the installations and methods. We prefer to retain Table 1, but will review the text to remove unnecessary redundancies.

24b. The authors may want to consider using italicized subtitles to break up section 2.2 into more digestible and easy-to-find subsections.

- Good point. We will break it into subsections, as suggested.

25. P2559, L3: Reference Figure 1 after ": : : lysimeter sites: : :"

- We will make that change.

26. P2559, L4-6: See specific comment 23.

- See response to comment 23.

27. P2559, L7-14: Some mention of how these other sites compare to the study area should be made, especially for the SWE data which was obtained from a site 12km away.

- In fact, the data from the Moyie Mountain snow pillow were not actually utilized in the manuscript. We will remove reference to the site.

28. P2559, L17: What temperature were soils burned at and for how long?

- The soils were burned at 500 °C for 4 hours to remove organics. This information will be added to the manuscript.

29. P2560, L6-8: Could the Ks calculation be described briefly here or in the Supp. Mat.?

- The methods for calculating Ks are explained in detail in the Guelph Permeameter operating instructions, which are readily available on the internet. We will reference the URL for convenience.

30a. Overall, I find section 2.3.3 on Statistical Analyses very cluttered and confusing. I think this section could be broken down into subsections (e.g. 2.3.3.1

Groundwater Response Classifications, 2.3.3.2 Data Transformations, 2.3.3.3 OLR).

- That's a good point. We will revise the text accordingly.

30b. Also, a lot of the background theory on OLR could be moved to the Supp. Mat.

- OLR is not widely applied within the hydrology community. Since it is the primary analysis applied in the study, we feel it is important to have this background information present in the methods section. However, we will review the text for conciseness and reference other literature wherever appropriate.

31a. On P2561, L25-27, groundwater dynamics are classified into three classes, persistent, transient and unresponsive. Then in the following sentence, the authors use different terminology, e.g. 'temporally discontinuous' and 'detectible'. Why not be consistent with the terminology?

- Agreed. The text will be revised accordingly.

31b. I also don't quite understand where this classification comes into play. Is it just used descriptively or is it somehow related to the 'Occurrence' response variable. Please clarify this in the text.

- Good point. The information was meant to describe the range of groundwater responses. The text will be revised to make this intention clearer.

32. P2562, L2: ": : : data censoring: : :" doesn't seem like the right term here. If the authors are talking about 'occurrence' here then even no data is meaningful since it is given a dummy value of 0.

- We used "data censoring" because of zero values (i.e. groundwater undetected) in the dataset. We agree that zero values are still meaningful, which is why we chose to apply OLR instead of ordinary regression.

33. P2562, L3: ": : : that did not experience groundwater responses: : :" Why not use the original terminology that you laid out, i.e. 'unresponsive'?

- Agreed. We will revise the text accordingly.

34. Building further on comments 11, 19 and 21, I am starting to wonder if any of the unresponsive wells were actually just not installed deep enough to measure the water table depth relative to the surface. I would like to know what the depths of the wells at the 13 unresponsive sites were. I fully understand the difficulties in installing wells into hard and/or cobbly soils, but since the whole point of this study was to look at groundwater dynamics in deep soils, this possibility should be addressed in some way.

- That's an important point that we will address more thoroughly. The mean depth of the unresponsive wells was 1.12 m (SD = 0.34) compared to 1.07 m (SD = 0.29 m) for the responsive wells. This information suggests that the difference in responses was related to factors other than well depth. We will incorporate this information in the text. This issue is addressed further below.

35. Are only the parameters with a symbol in Table 2 used in the OLR? If so, this needs to be specified in the caption and text.

- Good point. All parameters were tested in the models, but symbols were defined for only the parameters that were retained in the models. We will clarify this point in the figure caption.

36. It seems like the paragraph on the three types of response variables (P2563, L8 to P2564, L2) should be moved to before the description of OLR. This sets the stage for exactly what is being investigated and needs to be set out before the OLR discussion and the descriptions of the eight distinct hydrological periods.

- We will revise the section for clarity and flow, including breaking it into subsections.

37. Since all the wells were installed to different depths, shouldn't 'duration' be computed as the fractional portion of time that a water table was recorded at a specific depth below the ground surface and not just the duration of time it was in the well? And just because the water table drops below the bottom of a well, it doesn't mean that it's not there. Again, I find this measurement confusing. This comment also applies to the 'timing' response class.

- That's an important point and one that we gave extensive consideration. While this suggestion makes sense generally, that approach would have been problematic because of the large ranges in well depths and maximum water table levels measured in the wells. Both factors were

governed by the soil conditions, with the latter being influenced also by the local runoff processes. For any given reference soil depth, many sites would have been excluded from the sample population or would have had zero values due to wells being too shallow and due to the observed water tables (for the particular period of interest) being too deep. Combining these factors with the highly transient or unresponsive behaviour of many wells, the sample of observed responses would have been quite small. Given that statistical approaches utilizing ordered classes (which was necessary due to data censoring) require much larger sample sizes than approaches utilizing parametric data (e.g. ordinary regression), the statistical power for investigating the dataset would have been overly limiting. In response to this concern, we decided to incorporate well depth in the statistical models to test for its influence, as well as numerous soil parameters, including hydrologic conductivity profile data. Given that the average well depth for unresponsive sites was greater than that for responsive sites, it is apparent that other factors (e.g. soils, topography, vegetation) determined the occurrence, duration, and timing of groundwater responses, not well depth.

38a. P2563, L17-20: What do 'transient perched' and 'continuously persistent' mean? Again, can the original terms that were set out (persistent, transient, unresponsive) just be used here?

- Yes. We will revise the text accordingly to use consistent terminology throughout.

38b. Also, why were they treated as one population?

- The rationale is provided in the next sentence within the paragraph: "OLR requires that the number of cases within each response class exceed the number of predictor terms in the model, which restricted the number of classes that could be defined to two or three." We will revise the text to make this connection clearer.

39. Table 3 and the description of the different response variables and classes MUST be revised for clarity. As is, it is very confusing. Table 3 should include a new column on the far left titled 'Groundwater Response Class'. In this column, 'Occurrence', 'Duration' and 'Timing' can be listed. Leave the 'Period/Timing' column as is, but specify that occurrence and timing use the full annual dataset (I assume that is right?). I don't understand what 'Class 0', 'Class 1' and 'Class 2' listed under the 'Range of Responses' heading mean. Also, since this table is showing the range of responses for each class, this should be at the start of the caption: : : i.e. "Range of responses for occurrence, duration and timing response

classes: : :" then go on to describe units.

- Good suggestions. We'll make those changes.
- Classes 0, 1, and 2 defined three different response ranges for each response variable. The actual response values were replaced with the class values for the statistical analyses. This point will be clarified in the text and the figure caption.

40. P2564, L7: ": : : some circumstances: : :" is vague. Be more specific.

- "some circumstances" isn't necessary. We'll remove it.

41. P2565, L17-19: Does it make sense to include well depth as a predictor variable? It seems very likely that well depth would affect whether or not a well was responsive or not.

- Please see response to comment 37.

42. The OLR methods on P2564 are long. Could some of this be written more succinctly, combined with the previous OLR theory, or moved to the Supp. Mat.?

- We will revise this section for conciseness and break it into subsections for readability.

43. Figure 3 is very difficult to interpret. The grey scale used for potential radiation and snow cover extent is difficult to differentiate. Perhaps colour would work better in this regard. Also, is it necessary to keep the harvested area marked? It clutters the potential radiation map.

- We will either revise the figure for clarity or remove it entirely, whichever is most appropriate.

44. Maybe the description of snowline retreat in section 3.1 could be represented in Figure 3b instead of snow extent for each survey, which is difficult to interpret.

- We will either revise the figure for clarity or remove it entirely, whichever is most appropriate.

45a. P2566, L12-17: The authors state that the location of unresponsive wells was consistent with the model results, which is fine since those wells were not

included in the model.

- Actually, the unresponsive wells were included in the model as class 0 (Table 3).

45b. However, I don't understand what is meant by "The spatial distribution of deep soil Ks was also generally consistent with the model results based on a manual comparison." Was Ks being modelled? Please clarify.

- We'll clarify the point being made, which is that a manual comparison between the distributions of unresponsive sites and 75 cm Ks showed consistency with the model result that 75 cm Ks was important in determining the probability of groundwater response occurrence.

46. It is very hard to see the two light grey lines in Figure 4.

- We'll adjust the line shading and width.

47. Again, 'persistence' is used in lieu of 'duration' in the caption for Figure 4. Be consistent.

- Agreed.

48. The left-hand column of Table 4 should be revised in the same way as I have suggested for Table 3 in comment 39.

- Agreed.

49. The organization of the Results section is not intuitive to me. Why not organize it according to the response classes (occurrence, duration, timing)? Instead, occurrence and duration for the melt and annual periods are discussed in one section, timing in the next, and then duration for individual hydrologic periods. I find the current structure of the section detracts from the interesting results.

- Good point. We will revise the organization of these sections.

50. There is a lot of detail in section 3.4 that I think can be distilled down into the most important observations.

- We will revise the text for conciseness and possibly break it up into subsections for readability.

51. It is very hard to distinguish adjacent circle diameters in Figure 6 and I can barely see the 'no response' sites. While I like the idea of mapping the data, I'm not sure that this particular figure is the right way to do it. Maybe if annual and melt periods were displayed on separate maps the figures would seem less cluttered. This figure deserves more thought.

- We will break up the figures to improve legibility.

52a. Figure 8 is MUCH too small and some of the lines are far too faint. On P2569, L22-24, the authors state that we can see the relationships shift to higher or lower values of the predictor variables; however, this is difficult to pick out.

- Agreed. We originally had Figure 8 split into upper and lower halves, but they were combined during the publication process. We'll return to the original format.

52b. Figure 10 does clarify the observation; however, I would use the full names of each period on the x-axis instead of the numbers here. This makes it easier for the reader to interpret the changes without having to refer back to the period descriptions in the methods section.

- That's a good point. We'll either include the full names in the figure or in the figure caption, depending on how cluttered the figure is with the names.

53. P2571, L23-26: Is deep soil Ks the most important for all periods?

- To clarify, this sentence addresses the occurrence of groundwater response, not the duration of response. We'll revise the text to clarify this point.

54. P2572, L4-8: The authors compare their results to those of Redding and Devitio (2008, 2010); however, they do not compare Ks to the intensity of water inputs as Redding and Devito do. The statement that their results are consistent with the percolation-excess runoff generation mechanism seems like a jump and I think requires more evidence.

- We will elaborate on our discussion of the interplay between Ks and water input intensity to better substantiate our inference regarding percolation-excess runoff generation.

55. P2572, L18-21: I think there needs to be more discussion of the snowmelt and insolation patterns (Figure 3) and how they relate to spatial patterns of groundwater response to substantiate the statement re: controls on snowmelt intensity and hence groundwater response.

- Yes, we will elaborate on this discussion.

56. P2572, L25-28: A mechanistic explanation of the relationship between tree diameter and vertical vs lateral flux partitioning is needed here.

- We believe that maximum tree diameter was an important variable in the models (and consistently explained more variance in the models than other forest cover metrics) due to the disproportionate influence of large diameter trees (and their relatively large crowns) on snowpack shading and evapotranspiration compared to small diameter trees. We'll revise the text to clarify this point.

57. P2573, L2: How do you know that most of the catchment is "hydrologically connected"? Based on Figure 5.10 in the Supp. Mat., there are quite a few unresponsive sites between the responsive ones during the very wet snowmelt period. I think that you can only infer that the near-stream sites are connected, especially since the wells are sometimes hundreds of meters apart. Along these lines, I think you need to define what conditions infer hydrologic connectivity in this particular catchment, preferably in the Methods section. Some examination of the relationship between groundwater fluctuations and discharge, as well as the relationship between groundwater fluctuations in adjacent wells might be helpful in this regard.

- We will remove inferences regarding hydrologic connectivity since our results don't actually confirm connectivity.

58. P2573, L15 to P2574, L7: This paragraph is very wordy and hard to follow. I think it could be shortened and written more succinctly.

- We will revise for conciseness.

59. P2574, L8-11: There needs to be more of a mechanistic explanation given

here.

- We will add the points that forest cover removal generally increases snowpack accumulation and can increase melt intensity. We'll also add the point that maximum tree diameter consistently explained more variance in the models than other forest cover metrics likely due to the disproportionate influence of large diameter trees (and their relatively large crowns) on snowpack shading and evapotranspiration compared to small diameter trees.

60. In section 4.4, a lot of attention is paid to the results from other studies where I think the focus should be on the implications of the current study.

- We believe there should be a strong link to other study findings within the discussion section. Accordingly, we believe we struck a good balance between important findings from our work and other studies.

61. Also in section 4.4, I think there needs to be some discussion of how the models presented in Table 4 can be tested. Along these lines, it would have been nice to see the models verified in some way, for example, by testing on an adjacent area with similar characteristics.

- The study was an empirical analysis focused on testing hypotheses about spatial controls on groundwater response, rather than a predictive analysis. As a result, a formal test of the predictive power of the OLR models was not considered necessary. Moreover, the sample size was too small to split the dataset for verification, as logistic regression requires much larger sample sizes than ordinary regression.

62. Some of the wording in the Abstract is repeated verbatim in the Conclusions. For example, the sentence "Runoff source areas expand and contract: : : spatial patterns of topographic convergence;" is repeated verbatim.

- Good point. We incorporated some of the conclusions when writing the abstract. We'll revise both to ensure they're not simply repeating the same points verbatim.

Technical Corrections

P2550, L10: ": : :distribution of sites: : :" is vague. Perhaps change to ": : : spatial distribution of sites: : :" or ": : : number of sites: : :"

- We'll change it to "spatial distribution of sites".

P2560, L10: Change "Analysis" to "Analyses"

- Agreed.

P2561, L19: Same as last correction.

- Agreed.

P2561, L21-24: The sentence "Notwithstanding the fact: : : in the UEC catchment." Is long and awkward. It could be reworded more simply as follows "Although we do not have data for a full year, the missing data from late September and October 2008 are not of great concern since this is a relatively dry period in the UEC catchment." Or something like that.

- Agreed. We will revise for conciseness and clarity.

P2567, L23 and P2568, L6: Change "advanced" to "earlier".

- Agreed.

P2567, L25 and P2568, L7: Change "delayed" to "later".

- Agreed.

P2572, L22: Change "persistence" to "duration".

- Agreed.

P2574, L18: The word "initial" seems unnecessary here.

- Agreed.

P2576, L7: "differentiates" doesn't seem to fit here.

- We'll change to "controls the spatial distribution of"