REVIEWER 2

Comments by the reviewer are in black text

Our responses are in blue text

Where applicable, proposed changes to the revised manuscript are in blue and italics

Major comments

The manuscript describe a study using a numerical model to investigate the factors that influence mobilization and transport of carbon in soils in permafrost regions. Warming and permafrost thaw are likely to cause the transport of carbon from soils through aquatic systems, so a better understanding of physical mechanisms is crucial to the development of more accurate models and climate simulations.

We thank the reviewer for their thoughtful review. In response, we propose several modifications to address their comments and improve the manuscript. Our primary focus will be on providing a thorough justification of the meshing strategy, which was previously overlooked. Although it is not feasible to simulate multiple sites with this model, we will better clarify how our results can be generalized within a broader context. Additionally, we have added several clarifications throughout the text to enhance readability. We are confident that these changes will significantly improve the manuscript.

1. The study is designed as a highly idealized case for a small area in Svalbard. The grid setup is not standard. The time periods noted in the text are variable and confusing. The short time period of the abrupt that complicates interpretation. This application of a single model in this fashion makes extension of the study findings to broad regions of the Arctic questionable. The results would be more meaningful by applying the model for at least one or two other configurations based on observations of soil texture and organic carbon content, weather data, elevation gradients, which are available at select sites in Europe, Siberia, Canada, and northern Alaska. Doing so would allow the authors to have more confidence that what they are seeing represents a robust response and not an artifact of a unique model setup. The authors should also state the choice of application to the Endalen valley on Svalbard where observations are sparse. Is there a rich history of field study there?

We realize the description of the grid setup was not adequately explained and we will revise the text to better provide the justification. The key point is that the use of a flow-aligned, variable-width representative hillslope is a well-established modeling approach in catchment and hillslope hydrology. Please see our response to comment #9 by Reviewer 1. And while we agree that a study of multiple sites would be of interest in a future project, that's clearly out of scope for this project, which addresses our study site and takes advantage of available field observations. Importantly, we are confident that the understanding gained in this study is transferable to other similar locations and sites, although of course not generally to any permafrost region. We will clarify and elaborate on the potential applicability to different locations in the revised manuscript. Further, we will address the issue of only modeling this specific site by better highlighting the available field observations, which include groundwater flow velocity observations with the help of tracer experiments. This is in line with comment #27 by Reviewer 1.

2. The 0.25 km2 catchment is relatively small. Why not use a traditional uniform horizontal grid and with vertical soil layers? How many grid cells would be needed at resolution of 1 m? Would the computational expense be prohibitive? How does the setup result in “a natural equilibrium without artificially imposed boundary conditions”, and what does that phrase mean? Boundary conditions and zero-flux boundaries in a typical 3-D model setup could be the same as those in this study. In other words, I don’t see the advantages of this current grid mesh, nor the implications for the interpretation of results. While the authors refer to computational expense, more justification should be presented.

A variable resolution mesh with high refinement around a region of interest is very standard practice in groundwater and surface/subsurface integrated hydrologic modeling. This is especially true for combined flow and transport models because finer meshes are typically required to resolve transport. In our case, we refined our mesh in the region of our numerical transport experiments and kept a coarse mesh away from that region to set the appropriate flow conditions for transport. We stand by this choice of mesh as it is best practice for these types of simulations. The advantage of the domain setup used in this study is it allows for a feasible computational runtime and simplifies analysis of results in a way which is suitable for the aims of the study. For reference, the climate change scenarios alone required over 3 months runtime to compute on a high end workstation. Although a full 3D domain mesh could in principle be considered, it is not warranted for the aims of this study.

We agree the phrasing “a natural equilibrium…” is unclear and we will revise the text. What we meant was that we use hydrometerological forcing as driving conditions for the model, as opposed to cumbersome and often simplified boundary conditions. This enables the thermal-hydrological dynamics of the model to be controlled by weather inputs in a way which more closely resembles nature by precipitation, air temperature, and factors influencing evaporation, etc, as opposed to boundary conditions representing hydraulic pressures or hydraulic gradients, and thermal inputs, etc. Importantly, this is only possible for integrated surface/subsurface models like ATS. We will clarify the phrasing referred to as “a natural equilibrium…” in the revised text.

3. The wide variation in time steps mentioned is awkward and makes interpretation of the influence on results difficult. There is a rate for tracer injection of per second and a time step of monthly for the warming experiments. If it were implicitly daily, could mineralization in the future simulations like the present day simulation? The time step of per second is confusing in light of a monthly step in the future simulations.

Internally, computations are performed with an adaptive time stepping scheme but output frequency is user-controlled. For the multi-decade simulations representing climate change scenarios it is of course not practically possible to output results at a high frequency such as every second. Therefore, output frequency is assigned to monthly snapshots in those cases. However, for present-day conditions and when focusing details such as seasonal variability, it is relevant to consider higher frequency outputs. Although this leads to a range of different time representations in outputs, it is necessary and reflects the complexity and novelty of our study. We will add a sentence explaining that ATS is using adaptive time stepping with a maximum time step of 1 day and a minimum timestep of 1e-10 seconds and that output frequency depends on the scope of the simulation (i.e. seasonal vs. climate change scenario).

4. The term seasonal variability in the title is odd and unnecessary, as seasonal variability in a study like this is essential. Having the term ‘sensitivity study’ in the title would be more meaningful.

Seasonal variability stands in contrast to climate change and highlights the fact that we do both in this study. However, we do not conduct a full sensitivity study regarding model parameters or forcing. Hence, we prefer to keep the title as it is in the current version of the manuscript.

Minor comments

5. Line 6: grammar, “Here, we analyze of solute transport…”

Noted

6. Line 43: Unclear. What is “they” in the statement “With permafrost acting as a largely impermeable layer between the two, they are…”

Will change “they” to “the shallow and deep aquifers are mostly disconnected systems”

7. Figure 3 caption: It appears from the graphic that width (y-direction) decreases with distance in the x-direction. As x increases from 0 to 1000, the width gets smaller. Please clarify.

We will add a sentence to the Figure caption to explain this nature of a convergent hillslope setting.

8. Line 159 states that the model is run for an average yearly cycle. Then at line 172 there is reference to year-to-year differences. How can year-to-year differences in spinup be determined? Further clarify the spinup time period and protocol.

For each day of every year of the spinup, a certain state of the ground and the ground surface is calculated by the model. In the beginning, when the model is not in a cyclic steady state yet, there will be a difference between those days between one year and the next. Once these differences decrease to a negligible amount, the system is determined to be in “cyclic steady state”. We will include a plot in the Appendix showing differences in the surface temperature at a specific location for each day of the 10 years, showing that the difference between each e.g., June 1 becomes negligible. This is common practice in the modeling community (see comment #22 by Reviewer 1). We will add the following sentence to the manuscript:

“*We confirmed that the spinup process places the system in a cyclic steady state (see Appendix) where conditions in each day of the year are nearly identical to those of the same day of the previous year*”.

9. Line 171: typo “a the”

Noted

10. Line 173: What ten years? Is weather data not for an average yearly cycle (climatology)? Be specific about the ten year period.

We will change it to “ten years of average weather conditions”

11. Line 186: Clarify how numerically does a tracer become injected at a rate with a time step of a second. What is the model time step? If it is daily, how can something occur every second? Does the mass enter as a total each day? Intrinsic time step of the model is not clear.

ATS uses an adaptive time stepping scheme with user-assigned limits. This is advantageous as it adjusts the time step automatically using an iterative procedure to ensure numerical convergence depending on the dynamics of the system. For example, in spring and summer, with meltwater infiltration and variable groundwater flow, the time stepping scheme iteratively reduces the computational time step typically to a small value (less than a second), whereas during less dynamic periods, such as winter, the time step is typically increased to a maximum of one day, which helps reduce runtime. Of course the specified quantity is an injection *mass rate*; the actual mass injected during a time step depends on the length of that time step.

12. Line 235: typo, for for

Noted

13. Figure S3: component mass overlay is difficult to see in the graphic. Also add label for elevation for the y axis to make clearer what is being shown.

These will be reworked for clarity in the revised manuscript. Please also see our response to comment #49 by Reviewer 1.

14. Line 18 in supplement: units of %C ¿ 10 are unclear.

Noted

15. Line 337: typo, “have lead to”

Will be changed to “have led to”

16. Line 357: Do the magnitudes (+25 and +60 cm) of the abrupt active layer deepening have any grounding in reality? That is, do any studies report abrupt thaw that is some meaningful fraction of the amount modeled here? Some perspective on the magnitudes would aid in interpretation of the results.

We will add a reference to active layer variability in the Adventdalen valley. Strand et al., 2021 have shown that an abrupt increase of 30 cm is possible between a cold and a warm year. Hence, the modeled abrupt thaw is within the currently observed variability and additionally exaggerates it by a factor of 2. In a rapidly warming Arctic such as in the RCP8.5 scenario, heat extremes will be more common and may lead to abrupt thawing of the active layer, which is why we choose to consider scenarios which may seem extreme today, but may well be feasible within the time frame of the warming scenarios / coming century.

17. Line 368: “...is released from both the injection column…” grammar and unclear meaning.

We will remove “both”.

18. Line 391: At the mention of ancient carbon, it would be helpful to remind the reader that this is frozen carbon. For readers that skim the abstract and discussion, the terms ancient and buried will have unclear meanings. Making clear that ancient means ancient currently frozen carbon would help.

We will add this to the text.

References:

Strand, S.M., Christiansen, H.H., Johansson, M., Åkerman, J. and Humlum, O., 2021. Active layer thickening and controls on interannual variability in the Nordic Arctic compared to the circum-Arctic. *Permafrost and Periglacial Processes*, *32*(1), pp.47-58.