Response to Reviewer #1

Manuscript Title: Cold Climates, Complex Hydrology: Can A Land Surface Model Accurately Simulate Deep Percolation?

Authors: Alireza Amani, Marie-Amélie Boucher, Alexandre R. Cabral, Vincent Vionnet , and Étienne Gaborit

Reviewer Comments and Author Responses

- **The reviewer's comments in orange**
    - (page # line #) refers to the page and the number where the annotated comment is placed.
- **The author's response in blue**
- **Excerpts from the manuscript in *green***

We would like to thank the reviewer for their time and constructive feedback. The comments are mentioned and addressed in the same order that appeared in the annotated PDF. When multiple comments point to the same issue, a response is provided under the first comment in this document.

Comment 1 (page 1 line 22)**:** Trenberth's seminal article makes no mention of groundwater or groundwater recharge. Please cite a more relevant article.
Response: We replaced Trenberth with the following references (Page 2 Line 49):
Green, Timothy R., et al. "Beneath the surface of global change: Impacts of climate change on groundwater." *Journal of Hydrology* 405.3-4 (2011): 532-560.
&
Taylor, Richard G., et al. "Ground water and climate change." *Nature climate change* 3.4 (2013): 322-329.

Comment 2 (page 2 line 32)**:** What is 'them' in this context?
Response: The introduction is largely revised/rewritten and the referenced ambiguity is resolved, where 'them' referred to indirect methods of estimating deep percolation.

Comment 3 (page 3 line 67)**:** According to Figure 2, L1 and L2 appear to have different heights. Is this correct and, if so, it deserves a brief mention? What other features distinguish the two lysimeters?
Response: The only difference between the L1 and L2 lysimeters is their heights. The comparison between L1 and L2 was the subject of a study by our team (*Field-based assessment of the design of lysimeters for landfill final cover seepage control*; Kahale et al. 2022). It was shown that the L2 lysimeter collected almost as much percolation as the L1 (less than 5 % difference). We mentioned this in the revised manuscript Section 2.1 (Page 5 Line 111):

The lysimeters differ only in height: L1 is 2.10 m high, while L2 is 1.50 m. The comparison between L1 and L2 is the subject of Kahale et al. (2022), where it was demonstrated that the L2 lysimeter collected nearly the same amount of percolation as L1 (less than 5% difference).

Comment 4 (page 4 line 74)**:** Please specify the soil water retention model used and cite the source. See later comments about some inconsistencies between various SWR models (and their parameters) that seems to have been carried over from previous published works.
Response: We introduced the Clapp and Hornberger soil water retention model with proper citation (Page 8 Line 153).

Comment 5 (page 5 Table 1)**:** Again, I would suggest denoting the SWR model as a footnote.
Response: We mentioned the Clapp and Hornberger soil water retention model in the table caption.

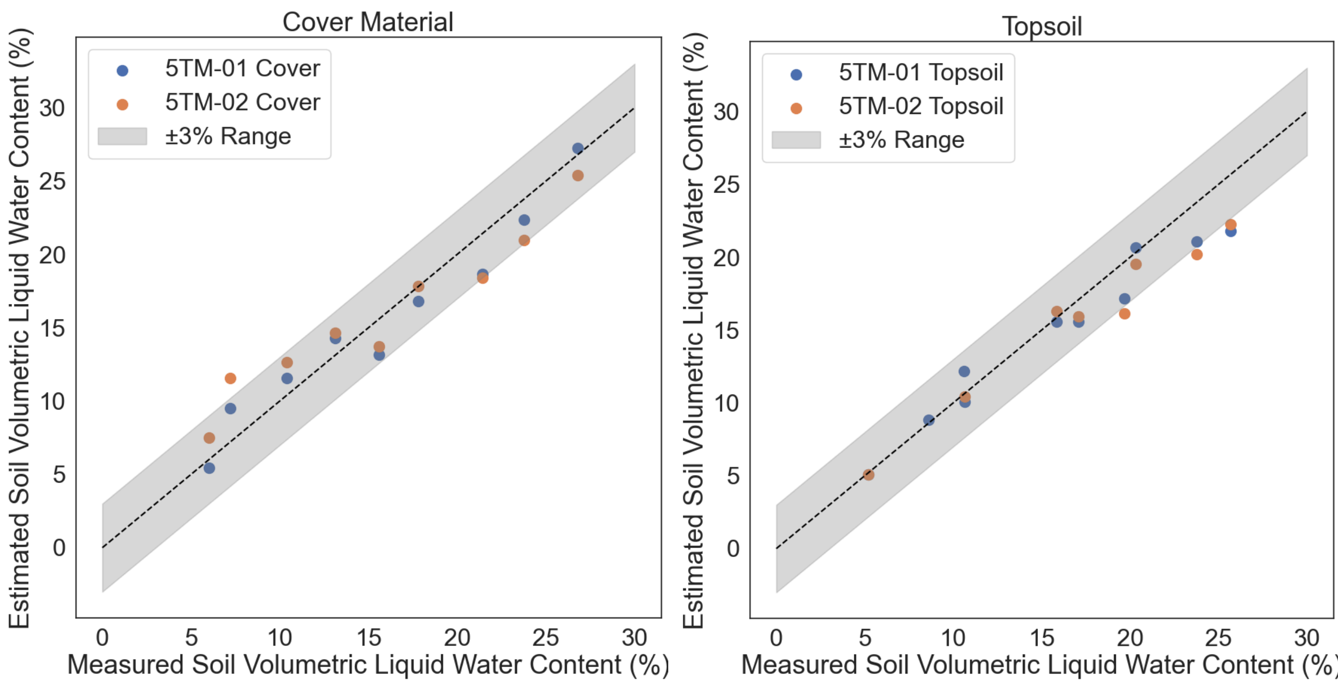
Comment 6 (page 6 line 103)**:** From what I can tell by looking at the literature cited here (I have not used SVS), SVS uses the Clapp and Hornberger (CH) model rather than the Brooks and Corey (BC) model. The two models have a similar foundation but there are key differences. CH uses a different definition for soil water saturation (BC defines effective saturation using a residual water content whereas CH does not). Also, the 'b coefficient' represents a pore size distribution index (BC's original terminology), but note that CH effectively defines b = 1/lambda, where lambda is BC original pore size distribution index.
Response: We acknowledge the difference between BC and CH models and we introduced CH as the correct model for SVS in the revised manuscript (Page 8 Line 153).

Comment 7 (page 6 line 108)**:** Soil water potential or soil matric potential would be preferable.
Response: We replaced *suction* with *soil matric potential* in the text (Page 9 Line 158)

Comment 8 (page 7 line 131)**:** K_sat previously.
Response: We ensured that the revised text is consistent regarding symbols and capitalizations.

Comment 9 (page 8 line 174)**:** Our group's experience indicates that manufacturer reported soil moisture sensor accuracies are highly optimistic, especially if no site-specific soil moisture calibration is undertaken. When using 'factory' equations that are not soil-specific, the 5TM used in this study (and other similar low cost sensors) often have much more significant absoluter errors, although they are generally reasonable for assessing temporal trends. There are numerous studies in the literature that have identified these issues, such as:
Mittelbach, H., Lehner, I., & Seneviratne, S. I. (2012). Comparison of four soil moisture sensor types under field conditions in Switzerland. Journal of Hydrology, 430, 39-49, https://doi.org/10.1016/j.jhydrol.2012.01.041

Response: We acknowledge the limitations of the probes used in our study. The comparison between measured and simulated time series in our study was focused on identifying general trends.
Additionally, our group's limited calibration exercise in the laboratory suggests that a ±3% accuracy range is reasonable for these probes and the calibration equation suggested by the manufacturer, as illustrated in this figure:



Scatter plot showing the relationship between measured and calibrated soil moisture sensor readings for both cover material and Topsoil. The shaded area represents the expected accuracy.

We mentioned this in the revised manuscript (Page 11 Line 229):

Applying manufacturer calibration functions to low-cost sensors like the 5TM may lead to larger deviations from the true soil moisture values (Mittelbach et al., 2012), our limited laboratory calibration exercise indicates that a ± 0.03 m3·m−3 range is reasonable for these probes and the manufacturer's calibration equation in our soils.

Comment 10 (page 9 line 212)**:** Why is there no data shown for the 2021-2022 winter period? Figure 3 shows snow data and SVS results for the period, so it seems curious that they are not included here (and in subsequent figures). Were they omitted due to the (relatively) poorer SVS model performance for snow depth in late winter/spring 2022, which would presumably then affect soil temp and moisture estimation? Or perhaps the soil moisture sensors were not collecting data in this period and I simply missed it?

Response: We did not collect soil moisture/temperature data after October 2021 (the data loggers were used in another project).

We augmented the figures to contain the simulated values for October 2021 to October 2022. We also mentioned this in Section 2.1. Of the revised manuscript (Page 5 Line 114):

Soil water content and temperature were monitored half-hourly by dielectric sensors (METER Group Inc., 5TM), represented by blue dots in Figure 2-c, until October 2021. After this date, the data loggers were used in another project, and soil data collection stopped. However, percolation measurements continued until October 2022.

Comment 11 (page 10 line 218)**:** Based on the sensor used, this is presumably liquid (unfrozen) water content, so it should be specified for both observations and simulated results.
Response: We made it clear in the revised text (including figures) that our manuscript only reports soil volumetric liquid water content measured value (Page 8 Line 147):
(throughout this article, soil moisture refers to the liquid water content in the soil)

Comment 12 (page 15 line 267)**:** The literature cited here misses early fundamental work by Stahli on preferential flow in frozen soil. Stahli, M., Jannson, P.-E. and Lundin, L.-C. (1996), Preferential water flow in a frozen soil - a two-domain model approach. Hydrol. Process., 10: 1305-1316. doi:10.1002/(SICI)1099-1085(199610)10:10<1305::AID-HYP462>3.0.CO;2-F
In addition, Mohammed et al. (2018) published a conceptual model of water flow and infiltration dynamics in frozen, macroporous soils that deserves mention here since it nicely highlights the complexities involved. Mohammed also goes on in subsequent field and modeling studies to demonstrate the influence of macropores on snowmelt infiltration and deep percolation in frozen soils.
Response: We thank the reviewer for the suggestions. We incorporated these references into our revised text (Page 19 Line 340).

Comment 13 (page 15 line 278**:** Mohammed (noted above) previously observed similar non-sequential wetting in the Canadian prairies, with deep percolation and groundwater recharge occurring through a significant soil frost zone. Mohammed, A. A., I. Pavlovskii, E. E. Cey, and M. Hayashi (2019), Effects of preferential flow on snowmelt partitioning and groundwater recharge in frozen soils, Hydrol. Earth Syst. Sci., 23(12), 5017-5031. doi:10.5194/hess-23-5017-2019
Response: We included this point as well as the suggested reference in the revised text, in Section 5. (Page 19 Line 353):

These findings are consistent with the results of Mohammed et al. (2019), which studied the impacts of preferential flow on snowmelt partitioning and groundwater recharge in frozen soils at three grassland sites (Stauffer, Spyhill, and Triple G) in the Canadian Prairies. They found that preferential flow paths, likely created by macropores in the soil, allowed for rapid infiltration and bypass flow, even when the ground was frozen, significantly influencing the partitioning of snowmelt into infiltration or runoff, and therefore the amount of water available for groundwater recharge.

Comment 14 (page 15 line 292)**:** I'm not sure what the point is here. Deep soil moisture readings tend to be relatively stable (in this and other studies) since there is less influence of surface temperature and evapotranspiration, so the relatively small CRPS seems to largely be a reflection of that stability. I am not certain it is truly a reflection of 'better' model performance.
Response: We agree with this assessment and we revised the text to reflect it (Page 21 Line 371):

Figure 8 reveals that the simulated ensemble has a relatively high overlap with the observations during the cold months where the movement of water in the soil is predominantly downward, reflected in a small CRPS value of 0.007 m3·m−3, which is likely influenced by the inherent stability of deep soil moisture during this period.

Comment 15 (page 15 line 296)**:** Readers can also see from Figure 10a that SVS is severely underestimating the dynamic soil moisture fluctuations at shallower depths, whereas it is largely overestimating the soil moisture fluctuations at greater depths. It highlights that SVS is struggling to capture soil moisture drainage and evapotranspiration in both shallow and deeper portions of the soil profile. Thus, it suggests to me that the capillary barrier effects are not the sole reason for the apparent discrepancies between model and observations. For example, the influence of preferential flow undoubtedly has an important influence on summer soil water flow and storage conditions as well, which is also likely contributing to the difficulty in SVS simulating field observations.
Response: We agree that factors beyond the capillary barrier effect are likely contributing to the observed discrepancies. We pointed out the need for further investigation and raised other possibilities, such as the potential influence of lysimeter characteristics and inaccuracies in ET estimation as contributing factors. This is mentioned in the revised manuscript (Page 21 Line 387):

Figure 11-a reveals a notable divergence between the simulated and observed near-surface soil moisture (at 75 mm depth), where the SVS model consistently underestimates soil moisture throughout the entire period from May to September 2020, and this underestimation becomes particularly pronounced during the summer months (July and August). This raises questions about potential contributing factors. Did the no-flux boundary walls of the lysimeter or the capillary barrier alone contribute to this discrepancy? Could the limitations of the one-dimensional single-continuum representation of the model in capturing the complex dynamics of water movement and storage in the vadose zone contribute to this discrepancy? While errors in evapotranspiration estimation could play a role, the absence of direct measurements limits our ability to fully quantify their impact. A thorough evaluation of the effects on deep percolation simulations necessitates more integrated data, including deep percolation and soil moisture measurements, particularly during summer.

Comment 16 (page 18 line 327)**:** I would argue preferential flow is also likely contributing to the issues you see in summer soil moisture/flux estimates from SVS also. 1D single-continuum representations of soil water movement and storage rarely capture the complex realities that we observe in the vadose zone.
Response: This possibility is included in the revised section (mentioned in the above response to comment 15).

Comment 17 (page 18 line 329)**:** G.Y. Niu has long advocated for improved inclusion of cold regions processes, specifically frozen soils, in land surface models and more recently has highlighted the need for inclusion of macropores/preferential flow in the vadose zone to properly capture the hydrology of cold regions. I think your study nicely emphasizes the need for this same inclusion as it relates to deeper percolation and groundwater recharge, since surface and subsurface systems are connected across scales.

Niu, G. Y., & Yang, Z. L. (2006). Effects of frozen soil on snowmelt runoff and soil water storage at a continental scale. Journal of Hydrometeorology, 7(5), 937-952.

Agnihotri, J., Behrangi, A., Tavakoly, A., Geheran, M., Farmani, M. A., & Niu, G. Y. (2023). Higher frozen soil permeability represented in a hydrological model improves spring streamflow prediction from river basin to continental scales. Water Resources Research, 59(4), e2022WR033075.

Response: Thank you for suggesting those references, which we included (see Page 23 Line 420). They convey a similar message to ours by emphasizing the representation of preferential flow within our models.

Comment 18 (page 21 line 430)**:** What is wi in this case? Soil frozen water (ice) content? Earlier (Eq 2 and 3), theta was used to denote volumetric soil water content, so I would advise sticking with a consistent variable set and nomenclature (e.g., soil water content vs soil moisture).

Response: wi refers to soil ice content. We ensured the suggested consistency and clarity in the revised text.

Comment 19 (general)**:** ... the study doesn't conclusively answer the title's question …

Response: Our modified conclusion explicitly answers the question posed in the title (Page 23 Line 425): *However, until these limitations are addressed, SVS cannot be considered a fully reliable tool for simulating deep percolation in cold environments.*

Additional notes: Any other in-line comments concerning grammar, dictation, and references not mentioned in this document are addressed during the revision.

Response to Reviewer #2

Manuscript Title: Cold Climates, Complex Hydrology: Can A Land Surface Model Accurately Simulate Deep Percolation?

Authors: Alireza Amani, Marie-Amélie Boucher, Alexandre R. Cabral, Vincent Vionnet, and Étienne Gaborit

Reviewer Comments and Author Responses

- **The reviewer's comments in orange**
- **The author's response in blue**
- **Excerpts from the manuscript in *green***

We would like to thank the reviewer for their time and constructive feedback. The responses are provided in separate sections, similar to the document submitted by the reviewer.

Comment: The title is not appropriate and does not correspond to the conclusions of the study
Response: Our modified conclusion explicitly answers the question posed in the title (Page 23 Line 425):
However, until these limitations are addressed, SVS cannot be considered a fully reliable tool for simulating deep percolation in cold environments.

The introduction needs to be more improved

Comment:. I find the introduction disorganized and incomplete. Preferential flows need to be mentioned, and also add more details on LSMs and lysimeters. To facilitate the reader's understanding of the impact of your study, I suggest an organization similar to the below: i) context : groundwater, complications and challenges of cold climates, characteristic and properties of landfill soils (this last part is really missing); ii) models can be used, including Lsm (give more details: their history, why they were developed, aren't they mainly used on a regional or global scale? What's the point of looking at an LSM on the scale of a lysimeter?); iii) Talking about lysimeters: this part is too limited and needs more details.
Response: We appreciate the feedback and the suggested organization concerning the introduction. The introduction is restructured and largely rewritten. The modified introduction follows the suggested organization and includes content on LSM and lysimeters.

Page 2 Lines 38
Physics-based numerical (hydrological) models, such as Land Surface Models (LSMs), are valuable tools for understanding and modeling deep percolation. Atmospheric scientists originally developed LSMs to address the limitations in early General Circulation Models (GCMs) which used fixed boundary conditions for land surfaces and failed to capture the complex interactions between the atmosphere and the land surface (Fisher and Koven, 2020). Manabe (1969) developed one of the first-generation LSMs to simulate these interactions using a simple "bucket" model for representing soil moisture. This model simplified land surface processes by focusing on the top meter of soil, where precipitation instantly infiltrated, evapotranspiration was the only way water left the soil, and runoff occurred when the soil became saturated (Yang, 2004). Subsequent generations of LSMs, such as Noah (Niu et al., 2011), and ISBA (Decharme and Douville, 2006), recognized the need for more realistic representations of vegetation and soil which have a crucial role in the exchange of water/energy at the land surface.

Page 3 Lines 65

First, direct measurements of deep percolation by lysimeters are difficult, costly, and scarce (Li and Shao, 2014). Lysimeters are containers or vessels, like a soil column, filled with soil or other material. They come in various types such as weighing (Payero and Irmak, 2008) or non-lysimeters pan lysimeters. Pan or zero-tension lysimeters are large-scale and are typically used for measuring percolation over a relatively large area. Pan lysimeters that are carefully designed can account for spatial variability of the soil, the presence of preferential flow pathways, and resolve percolation rates as low as 0.1 mm/yr (Malusis and Benson, 2006).

Regarding the soils: cover material and topsoil used in filling the lysimeters and building the experimental plot described in our study are locally available materials and do not belong to any special category such as ‘landfill soils'. That being said, our field investigation included other experimental plots filled with contaminated soils. Those plots were not part of this study.

Regarding the comment on the point-scale evaluation:
The main advantage of performing a point-scale evaluation is the reduction of biases in the meteorological input data. This makes it more likely to identify potential shortcomings related only to the model's structure/parameterizations. There are similar studies concerning land surface models, such as the references below:

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., ... & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, *116*(D12).

Jiménez-Rodríguez, C. D., Sulis, M., & Schymanski, S. (2022). Exploring the role of bedrock representation on plant transpiration response during dry periods at four forested sites in Europe. *Biogeosciences*, *19*(14), 3395-3423.

Comment: I find the literature search interesting: however, it is also necessary to give some values in the text (and not only in the Figure). Also, if the values correspond to frequencies, then why does the sum not equal 100? And I also don't understand why there are 57 cases in fig 1a while there are 28 in fig 1b; Be coherent.
Response: We provided statistics within the revised text as suggested:

Page 3 Line 81
Our survey shows that about 60 % of the studies evaluated models' deep percolation at coarser-than-daily resolutions.

Page 3 Line 89
The case studies are concentrated in temperate (58 %) and arid (24%) regions with a smaller representation in continental zones (15%), indicating a focus on areas where snow is less prevalent.

We also revised the figures (Figure 1a, b), using percentages instead of frequencies and changed the bar label from "Total" to "Annual / cumulative", to avoid the misinterpretation (the " Total" bar may have been interpreted as representing the cumulative count of cases).

Comment (page 1 line 15)**:** Why are they important?
Response: Currently, the importance of numerical (land surface) models is explained using examples (Page 2 Line 47):

These complex models can explore hypothetical scenarios, enhance our understanding of physical processes, and provide explanatory, process-based predictions (Willcox et al., 2021). For example, they help quantify how changing precipitation patterns impact groundwater recharge (Taylor et al., 2013; Green et al., 2011; Wu et al., 2020). This is particularly important in regions where fluctuations in groundwater resources heavily influence ecosystem water availability (Huang et al., 2019; Orellana et al., 2012). Beyond natural systems, LSMs can *guide the design of landfill final covers. By simulating the interaction of prospective covers with local hydrometeorological conditions (Ho et al., 2004) designers can optimize these covers to minimize leachate production, thereby saving time and resources.*

This part is preceded by a brief overview of LSM development which emphasizes their importance for climate modeling.

Comment (page 2 line 34)**:** There are many other limitations (Vereecken et al., 2019).
Response: We acknowledge that there are other limitations associated with numerical models that are not mentioned within the text or addressed by our model evaluation effort. Our focus was limited to identifying potential process representation within the SVS model that needs to be improved to better simulate deep percolation in cold regions.
In the conclusion of the revised manuscript, we added (Page 23 Line 428):

This approach, along with considering other limitations of the LSMs (Vereecken et al., 2019), can further improve deep percolation simulation by land surface models and enhance our understanding of cold region hydrology and its response to a changing climate.

Comment (page 2 line 53)**:** "This study aims to improve deep percolation simulation". I cannot see the improvement, as there is no comparison with the initial version...
Response: We revised the sentence to (Page 3 Line 93):

This study comprehensively evaluates the Soil, Vegetation, and Snow (SVS, Alavi et al. (2016); Husain et al. (2016)) LSM in a cold environment setting.

The improvement we originally wrote about pointed out the benefits of a robust model evaluation, in which we can identify the model's weaknesses and ratify them. The addition of the soil freezing module, enables the model to simulate soil freezing/thawing and its impact on soil hydrology. It also enables the model to simulate soil temperature for each layer within the model. This is a physical necessity and was not implemented simply to improve the performance of models as quantified by any performance metric.

Comment (page 2 line 58)**:** Not soil matric potential?
Response: The experimental plot (inside and outside the lysimeters) were also equipped with soil matric potential sensors at various depths. However, these measurements were not included in any of our analyses concerning this manuscript, therefore we decided not to include them in the figures and not mention them in the text.
In the revised manuscript, we acknowledged the presence of soil matric potential sensors in the revised text (Page 5 Line 118):

Soil matric potential sensors were also installed at various depths in the plot, but are not shown in Figure 2-c because they were not used in the analyses.

Comment (page 3 line 60)**:** Line 60 : Maybe for SVS, but not for LSMs (Boone et al., 2001 ; Decharme et al., 2016)
Response: We revised the sentence to make it clear that the soil freezing scheme is evaluated for SVS (Page 5 Line 100):
Additionally, we implemented a simple soil-freezing scheme, which is assessed for the first time in the scientific literature for SVS.

Description of lysimeters, soils, datas and parameters are incomplete

Comment: The description of the soils and lysimeters is really incomplete, which makes it difficult to interpret the results. What are the characteristics of these "landfill" soils? What makes them special? And is the soil covered with vegetation? Please comment texture values when you describe soil. 0% of Clay?
Response: We augmented Table 1 to include the USDA soil classification for the soils. We have also included soil water retention curves of the soil samples with the accompanying curve-fitting, Figure 3 in the revised text.

- As addressed above, no special type of soil is used in filling the L1 and L2 lysimeters.

- The soil is covered with grass and this is mentioned in the revised text (Page 5 Line 105):
  The plot was fully covered with grass one month after construction.

Comment: There is a wide diversity of lysimetric devices and methods. What are the dimensions, the installation procedures...? I later learn that there are runoff measurements... Is there an imposed slope on these lysimeters? Explain the differences between the lysimeters used in this study, etc...
Response: The installation procedures are detailed in a published article from our group.
Kahale, T., Ouédraogo, O., Duarte Neto, M., Simard, V. and Cabral, A.R., 2022. Field-based assessment of the design of lysimeters for landfill final cover seepage control. *Journal of the Air & Waste Management Association*, *72*(12), pp.1477-1488.
We also provided the information in Section 2.1

- The lysimeters are 2% sloped. This was considered in the simulations (as an SVS model's input parameter).
- Two surface water (runoff) collection systems were designed and constructed in the field. Despite these efforts, neither system yielded reliable measurements. Therefore, they were not used in our model evaluation effort.
- The only difference between the L1 and L2 lysimeters is their height. This was the focus of Kahale et al. (2022) and they have shown that the L2 lysimeter (shorter walls) collects practically the same percolation volume as L1.

Page 5 Line 106

The top and bottom of the excavation followed the regulatory requirement of a 2% slope. The soil configuration and dimensions of the pan lysimeters, namely the L1 and L2 lysimeters, are illustrated in Figure 2-c. The 4 m by 4 m lysimeters were lined with a 1.5 mm thick HDPE geomembrane mounted on a wooden frame. A 100 mm thick gravel layer was included at the bottom of the lysimeters, overlaid by a 100 mm sand layer. The gravel layer served as drainage, while the sand layer acted as a filter to prevent clogging by the cover material (both shown in Figure 2-c). The lysimeters differ only in height: L1 is 2.10 m high, while L2 is 1.50 m. The comparison between L1 and L2 is the subject of Kahale et al. (2022), where it was demonstrated that the L2 lysimeter collected nearly the same amount of percolation as L1 (less than 5% difference).

Page 5 Line 117

Despite designing and constructing two surface water (runoff) collection systems, neither yielded reliable measurements and were excluded from the model evaluation.

Comment: The scheme in fig. 2c could be improved and made bigger with more detail. For me, figs. 2a and 2b are not visible and not necessary.

Response: Figure 2 was modified, resulting in the enlargement of each subplot. The removal of one subplot contributed to increasing the readability of the remaining subplots.

Comment: Some suggestions for Table 1 :

- Add the heading "Observation" for the 1st column and "Model" for the 2nd column
- Illustrate the variability of soil (textures+parameters) for the two lysimeters
- What is N ?

Response: We included the additional headers, as suggested.

The lysimeters are filled with the same soil types. The variability associated with the physical properties of these soils is reflected in the wide range of their laboratory-estimated values. The 'N' refers to the number of tests in the laboratory. We included this definition in the caption of the revised table.

Comment: As you mentioned in your introduction, model parameters are extremely important for the quality of simulations. However, no comment is given in the text on the values obtained. Are these values comparable with these soils ? Are they the same as those generally estimated by LSMs? A value of b equal to 1 is surprising, especially in comparison with pedotransfer functions used by LSMs. Please comment, especially given its importance in the dynamics of percolation simulation.

Response: We added Table S1 in the supplementary material which compares parameter values as estimated by SVS pedo transfer functions and estimated in our laboratory.

Comment: Moreover, the parameters are estimated for the period between April to September 2019. Some studies recognised that changes to the soil and its structure can be significant in lysimeters (Weihermüller et al., 2007; Séré et al., 2012; Seneviratne et al., 2012). Have you performed this protocol at other periods? If so, are the parameters measured different? If not, this could be something to consider, especially in view of the low scores obtained after 2020, and would be a point for discussion.

Response: We have not performed inverse modeling in any other period. However, we acknowledge that the soil properties within the lysimeters might have changed over time. This could potentially explain the low model performance scores. We included a paragraph at the end of the discussion section of the revised manuscript to acknowledge the point (Page 21 Line 396):

This study's parameter estimation, based on inverse modeling (Section 3.3.1), used data from April to September 2019. While lysimeters offer controlled conditions for hydrological studies, they can change over time, influencing hydrological processes. . Séré et al. (2012) highlight how changing soil properties within lysimeters can affect water flow and solute transport. The discrepancy between the model and observations could be partly attributed to such temporal changes. Future studies with longer datasets can monitor lysimeter soil properties and use multi-year calibration to assess the impact of these changes on hydrological simulations.

Comment: If there is vegetation, please explain how you set the parameters.
Response: This information is added to Section 3.2 of the revised manuscript (Page 9 Line 174):
The experimental plot was fully covered with short grass, classified as "low vegetation" in SVS. Parameters for low vegetation, such as leaf area index (LAI), roughness length, heat capacity, albedo, stomatal resistance, and root depth, are derived from lookup tables containing values for 21 vegetation classes.

More details are provided in:
Alavi, N., Bélair, S., Fortin, V., Zhang, S., Husain, S.Z., Carrera, M.L. and Abrahamowicz, M., 2016. Warm season evaluation of soil moisture prediction in the Soil, Vegetation, and Snow (SVS) scheme. *Journal of Hydrometeorology*, *17*(8), pp.2315-2332.

Comment: Comment on the quality of your data, the gaps, and how you integrate it into your analysis.
Response: The potential influence of data gaps on the analysis is discussed in Section 4 (Results) of the revised manuscript. When calculating performance metrics, time steps with missing measurements were not considered. Specifically, we address the potential effects of gaps in different datasets:

- Near-surface soil temperature: The gap in this dataset has minimal influence on performance evaluation as it primarily falls outside the soil freezing period, during which the model accurately simulates soil temperature dynamics (Page 13 Line 278):
  There are two notable gaps in the data, from March 2020 to May 2020, and from November 2020 and January 2021. These gaps are likely to have a small effect on the performance evaluation, as they are mostly outside of the freezing period, and the model tends to simulate soil temperature well during this time.

- Deep soil moisture: The gap in this dataset likely has a minor impact on the overall analysis. The model and observations show good agreement in periods outside of the summer months, suggesting the gap's influence is limited (Page 15 Line 297):
  It is worth noting that the deep soil moisture data contains gaps from March 2020 to May 2020, and from November 2020 to January 2021, but these gaps likely have a minor impact on the overall analysis because, except for the summer months, the model and observations align closely in terms of absolute values.

- Near-surface soil moisture and deep percolation: Due to the complex dynamics governing these variables, it is difficult to definitively assess the impact of data gaps.

Comment: For me, the "Methods" section is not simply a Methods section, in particular the model description. Find a more appropriate name for this section or a better structure.
Response: We choose a more descriptive name for this section: 'Methodology and Model Description'

There is no comparison with the initial version

Comment: As I said earlier, in order to assess the contribution of the gel to this model, we need to compare it with the original version. This procedure is essential to highlight this work.
Response: The primary objective of this manuscript was to conduct a process-based model evaluation of the SVS model. Assessing the improvement associated with adding the freezing module was not part of our study.

We reran the model ensemble by deactivating the soil freezing module and compared this run with the original run. Two figures are provided in the supplementary material:

- Figure S1 compares daily averaged soil moisture at 75 mm below the surface for the two aforementioned simulation runs, limited to cold months (November to April), as during these periods, one can observe the influence of activating or deactivating the soil freezing module.
- Figure S2 shows the comparison between daily deep percolation volumes between the simulation runs, limited to cold months.

Analysis between variables (periods and criterias) are confusing

Comment: It's very complicated to compare Figs 3, 4, 5, 7 and 8, because they don't have the same evaluation periods. I can understand Fig 3, but not for the others. You have to be consistent. It is very confusing for me. In addition, evaluation metrics are not always identical from one analysis to another, especially correlation. This is a criterion to be included in each analysis.
Response: We revised the figures and implemented the changes to address the comment:
Consistent Evaluation Periods:

- All figures (Figs 5, 6, 8, and 9) now cover the same period.
- Inclusion of correlation coefficient:
  We have added the correlation coefficient to the time-series figures.

Results not presented in enough detail

Comment: You never mention two essential concepts: ETP and water balances. However, these are still key aspects of hydrology and can be estimated using lysimeters. An assessment of the model's capacity to reproduce these two elements would be advisable.
Response: During our field investigation, direct measurements of evapotranspiration (ETP) were not conducted due to the construction of pan lysimeters instead of weighing lysimeters. As previously mentioned, we tried to measure runoff from the surface of the experimental plot by designing and constructing two systems; however, neither system yielded reliable measurements. Consequently, indirect estimation of ETP is also not feasible in our case.
Evaluation of SVS's ability to calculate surface energy balance (including ETP) was the focus of another study:
Leonardini, G., Anctil, F., Abrahamowicz, M., Gaborit, É., Vionnet, V., Nadeau, D.F. and Fortin, V., 2020. Evaluation of the Soil, Vegetation, and Snow (SVS) land surface model for the simulation of surface energy fluxes and soil moisture under snow-free conditions. *Atmosphere*, *11*(3), p.278.
Leonardini et al. (2020) assessed SVS's ETP simulation capabilities across six FLUXNET sites (2004-2015) and found it generally performed well.

Comment: Then, I find there is very poor comment on the observations, yet observations are also a result in themselves.
Response: A comprehensive process-based evaluation of field observations is the topic of a separate manuscript (written by the same authors as the current manuscript) presently being reviewed for publication in Hydrological Processes.

Comment: Also, in the observations, it would be good to highlight the variability between lysimeters, as you did for Fig 8. That would make it possible to assess heterogeneity, the influence of the size of the lysimeters, changes in the soil over time, etc.
Response: This is a very good suggestion. We modified and reproduced the relevant figures accordingly.

Comment: I also think it would be interesting to have a seasonal cycle, in addition to Fig 6. to have a temporal dimension in the differences between observation and model.
Response: The caveat in displaying the comparison between the model and the observations seasonally is the duration (months of data) with absent measurements for lysimeters. This may result in a prejudiced comparison.

Comment: In the Figs 3, 4, 5, 7 and 8, it would also be good to include the statistical values over the complete period (although this is included in the text).
Response: We modified and reproduced the figures to include the evaluation metrics over the entire period.

Snow Depth

Comment: Where is the year 2022?
Response: Figure 4-c shows the comparison (SVS vs measurement) for the third winter, including 2022.

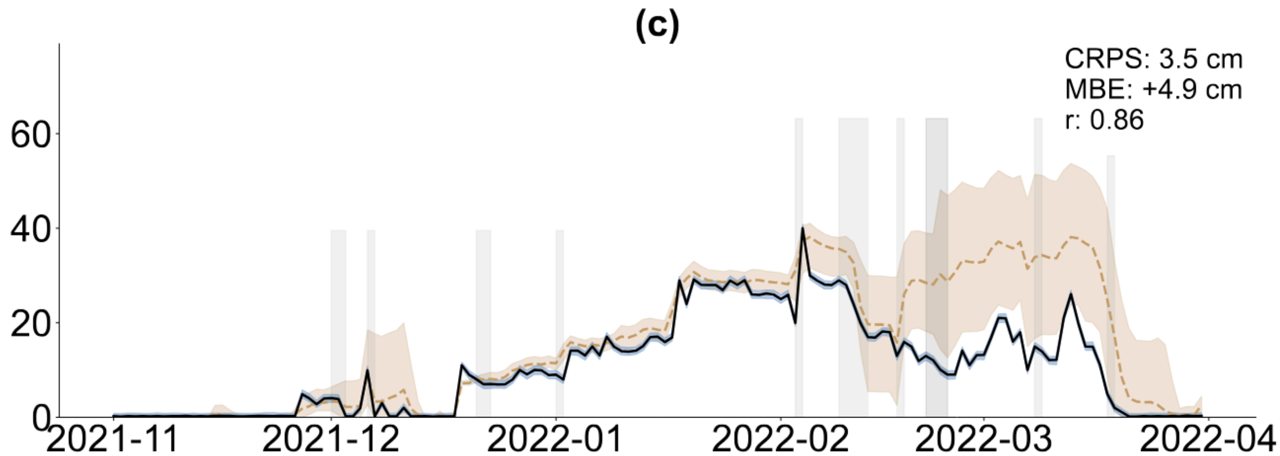


Figure 4

Comment: Comment the very similar dynamics (sensitivity) of the model at the end of winter. For example, the maximum peak is almost always at the same time of year, with a value very close to 50 cm.
Response: We investigated the results (data + model output) but did not find a specific reason for the maximum peak consistently being close to 50 cm. However, it's important to note that during the first two winters, the SVS ensemble mean closely matches the observed values at the times of peak snow depth.

We acknowledged this in the revised manuscript at the end of Section 4.1 (Page 21 Line 379):
Interestingly, the maximum snow depth consistently reaches a value close to 50 cm for the three winters, but no specific reason for this was identified in the data. However, it is important to note that during the first two winters, the SVS ensemble mean closely matches the observed values at the times of peak snow depth.

Near-surface Soil Temperature

Comment: Do you have an estimation of the depth of frost?
Response: While there are soil temperature sensors at 600 mm depth in the E1 enclosure, their data were not available during the period relevant to the discussion in Section 5.1 (December 2019 and 2020). We addressed this at the end of Section 5.1 (Page 19 Line 330):

To better understand the potential for frost depth overestimation, future studies can gain a more comprehensive understanding by analyzing soil temperature at deeper soil depths. Although there are 600-mm depth soil temperature sensors in the E1 enclosure, their data were unavailable for December 2019 and 2020, preventing direct assessment of the frost depth during this period.

Near-surface Soil moisture

Comment: Please add the correlation in your analyse. I guess the correlation is not very good. Comment. And notably in summer when the observation is very low until model is stable. This may illustrate the fact

that the model does not represent ETP very well. Is there any vegetation that can increase ETP?
Response: The correlation (for warm/cold months) is depicted in Figure 7b. This statistic is also included in time-series figures.
We added a sentence in Section 5.3. to acknowledge the comment re. the potential influence of ET estimation on near-surface soil moisture during warm months (Page 21 Line 392):
While errors in evapotranspiration estimation could play a role, the absence of direct measurements limits our ability to fully quantify their impact.

- The plot is fully vegetated. We added this information to Section 3.2 of the revised manuscript (Page 9 Line 174):
  The experimental plot was fully covered with short grass, classified as "low vegetation" in SVS. Parameters for low vegetation, such as leaf area index (LAI), roughness length, heat capacity, albedo, stomatal resistance, and root depth, are derived from lookup tables containing values for 21 vegetation classes.

Deep percolation

Comment: Please comment the observation! What do the periods without observation lines mean (e.g. 2020-11 to 2021-03); Is this a period without drainage? or a period without observation? How is this taken into account in your statistical evaluation?
Response: To clarify Figure 9, we modified it to distinguish between times when no volume was collected and times when observations were not possible due to equipment malfunctions. These periods, shaded grey on the figure, were excluded when calculating the performance metrics.

Comment: I don't agree with your comment when you say the model correctly reproduces the observations. Perhaps the CRPS is good, but the dynamics are not good, with significant dephasing and a model that reacts too quickly and too intensely.
Response: We modified the text to better reflect the model’s struggles concerning deep percolation (Page 16 Line 302):
The CRPS values may indicate a good agreement between SVS and observations, 0.6 (mm·day−1) from September 2019 to November 2020, 0.2 (mm·day−1) from November 2020 to November 2021, and 0.8 (mm·day−1) from November 2021 to November 2022. However, Figure 9 shows that SVS struggles to match the temporal dynamics of deep percolation closely, this is most notable in October 2019, March 2020, and Winter 2022.

Comment: Is it possible that the low scores after 2021 are due to a transformation of the environment within the lysimeters over time? in structure and hydrodynamic parameters?
Response: Our analyses showed that the model's inability to capture mid-winter events in 2022 is the primary factor contributing to the low scores. This failure is due to the model's inability to simulate frozen soil infiltration and the movement of water through preferential flow pathways.

Discussion should be improved

Comment: I rather like the discussion, which is well constructed and well developed.
I think an analysis of the sensitivity of the parameters could be useful for this study. For example, if we only perturbed one parameter, or only the meteorological forcing, how will the simulations be affected?
Response: We also believe such analysis would be an interesting dimension to explore. However, it was outside the scope of the current study. This can be pursued in the future using the publicly shared dataset from our field investigation using SVS or other land surface models. We mentioned this as a suggestion in the revised Conclusion (Page 23 Line 422):

Furthermore, a detailed sensitivity analysis investigating the impact of individual parameters and meteorological forcing on model performance would be valuable for identifying the key factors driving model uncertainty and guiding future model development and calibration efforts.

Comment: Figures should be improved with the use of hovmoller diagram. These diagrams are often used on lysimeters to follow the movement of water in the soil and to help comprehension (depth vs time) (Abdou et al., 2004, Decharme et al., 2016, Sobaga et al., 2023).
Response: We agree that such diagrams would help to better visualize the water movement across the soil column. However, as only a few sensors exist, the Hovmoller diagram might not be as informative as expected. Additionally, the sensors are not placed at regular intervals, which would make the interpretation of the diagram even more challenging. Therefore, we would prefer not to include Hovmoller diagrams in our current manuscript.

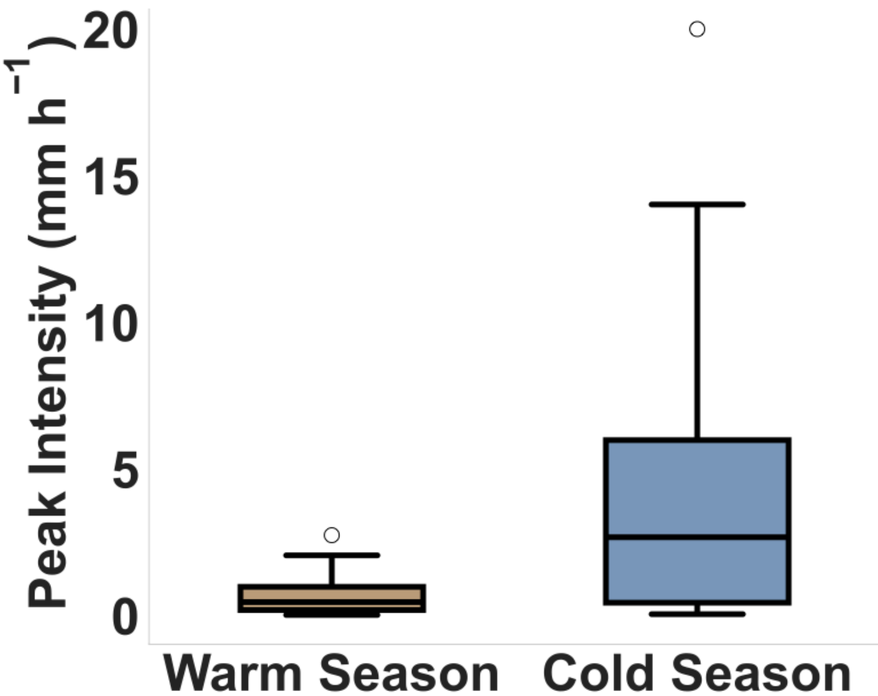
Comment: For the example you are using in this section, please use a symbol (a star?) on the previous chronicles (Fig 3 - 8).
Response: We modified Figure 5 to include zoomed-in subplots of the periods mentioned in Section 5.1., namely December 2019 and 2020. We also color-shaded winter 2021 in Figure 8, which was referenced in Section 5.2.

Limitations of SVS in simulating soil moisture and deep percolation in winter

Comment: Can you examine how many events the model does not react to while the observations do react? And give a statistic for the cases where preferential flows are observed?
Response: In Section 4.4, we provided the volume and number of days that we have observations and no significant simulated value during winter 2022 (Page 16 Line 309):

This change is largely attributed to SVS missing several large deep percolation events during the winter of 2022, as also evident in Figure 7-c. Investigating the results reveals that from 2022-01-01 to 2022-03-31 there were 23 *days in which the L1 lysimeter collected more than 1 mm of deep percolation (in total 176.8 mm) whereas the ensemble mean values were smaller than 1 mm (in total 2.8 mm).*

Comment: Furthermore, don't you think that these preferential flows would not be due (or at least reinforced) by the lysimeters? Either by edge effects? or by a modification of the soil?
Response: In another manuscript–submitted before this one and currently under review after revision–, where we analyzed deep percolation observations from four lysimeters (including the L1 and L2 lysimeters), we showed that preferential flow mostly happened during cold months, reflected by distinctly high peak intensities for deep percolation events starting in the cold season. This is shown in the figure below. While we acknowledge that edge effects within the lysimeters may have reinforced preferential flow to some extent, the strong seasonal pattern of preferential flow events suggests that lysimeter artifacts are not the sole cause.



(*This figure and the relevant analysis are part of a manuscript submitted to Hydrological Processes)

That being said, we acknowledged this comment in Section 5.2 (Page 19 Line 357):
It is also important to acknowledge that the lysimeter itself, by its design, could influence the occurrence and patterns of preferential flow events, potentially amplifying or altering natural flow paths (Williams et al., 2019, 2020). Further investigation is needed to understand how these factors contribute to the discrepancies observed between the simulated and observed soil moisture and deep percolation, especially under freezing conditions.

Influence of Capillary Barrier on Simulated Subsurface Soil Moisture

Comment: Do you try to change the thickness of the last layers?
Response: Changing the thickness of the last layers from 2.5 cm to 2.0 cm (a lower value will cause numerical stability in our version of SVS) did not change the simulation results and the observed behavior.

Comment: Are these results also verified on the Lysi L2? If not, why not?
Response: We investigated the L2 observations (soil volumetric liquid water content at 1850 / 1750 mm) and it shows the same behavior as L1, mentioned in the revised text (Page 21 Line 382):

Analysis of the L2 lysimeter data reveals the same pattern of divergence between simulated and observed soil moisture

Comment: Please comment the Fig 10.a,
Response: We commented on Figure 11-a (10-a in the unrevised version) in the revised manuscript (Page 21 Line 387):

Figure 11-a reveals a notable divergence between the simulated and observed near-surface soil moisture (at 75 mm depth), where the SVS model consistently underestimates soil moisture throughout the entire period from May to September 2020, and this underestimation becomes particularly pronounced during the summer months (July and August).

Comment: A simple simulation with the same parameters and a free LBC would be welcome to confirm these observations, and to check if this is only the effect of the LBC.
Response: Unfortunately, there is no option to select a free drainage LBC in the SVS model.

Other Comments

Comment: Lsm are generally used on a regional or global scale. In your opinion, what would be the main challenges of the change of scale for these cold climates? And how would you integrate preferential flows on a larger scale?
Response: While LSMs are used at regional or global scales, their application at smaller scales is straightforward and not uncommon in research applications (see examples below). Working at smaller scales can be advantageous, as we have reduced uncertainty in parameter values and meteorological data. This allows for better model calibration and validation, ultimately leading to more reliable simulations of the hydrological processes specific to cold environments.
Some relevant references (see below for examples) support the fact that LSMs are also employed in point-scale.

- Zhang, Li, et al. "Evaluation of the Community Land Model simulated carbon and water fluxes against observations over ChinaFLUX sites." Agricultural and Forest Meteorology 226 (2016): 174-185.

- Denager, Tanja, et al. "Point-scale multi-objective calibration of the Community Land Model (version 5.0) using in situ observations of water and energy fluxes and variables." Hydrology and Earth System Sciences 27.14 (2023): 2827-2845.
- Chadburn, S., et al. "An improved representation of physical permafrost dynamics in the JULES land-surface model." Geoscientific Model Development 8.5 (2015): 1493-1508.
- Harper, Anna B., et al. "Improvement of modelling plant responses to low soil moisture in JULESvn4. 9 and evaluation against flux tower measurements." Geoscientific Model Development Discussions 2020 (2020): 1-42.

We mentioned this in the revised introduction (Page 2 Line 59):

A point-scale evaluation, in contrast to larger-scale evaluations, may allow for a detailed assessment of the model's ability to capture hydrological processes in cold environments, by reducing the uncertainty in model parameters and meteorological data. This approach has been successfully employed in similar studies (Zhang et al., 2016; Denager et al., 2023; Chadburn et al., 2015).

Comment: When you use LBC seepage for hydraulic conditions, how do you define LBC for Temperature?
Response: This is described in the A2 section of the Appendix.

Additional notes: We addressed and incorporated all the suggestions designated as 'Minor Corrections' in the revised manuscript.