We thank the anonymous reviewer for their time in reviewing our paper. We provide responses to each individual point below. For clarity, comments are given in italics, and our responses are given in plain blue text.

Permafrost and its reaction on climate change are key issues of cryosphere research and thus also in TC. As the processes that change permafrost are very complex, process understanding is still comparatively limited. Improving the modelling of permafrost is an important approach to close this research gap. The work of Zegers et al. addresses relevant scientific questions in the context of permafrost research by improving its modelling through introduction of air convection and its effects (heat advection-conduction without assuming a local thermal equilibrium between the air and matrix) into the 3D hydrology model GEOtop. In this way, new methods and results for permafrost modelling based on the existing theory of air convection and density-driven buoyancy flow of air and heat advection conduction in a porous matrix are presented.

I found the paper very interesting, generally well written, well structured and easy to follow. Only the large number of acronyms sometimes made it difficult to follow the storyline. The title of the paper is well chosen and reflects the content well. The abstract is short and concise and contains key information of the paper. Due to the topic of the study, the paper contains a larger number of mathematical formulae, but these are balanced in scope, well presented and explained in the text as well as are useful for understanding the content. Results are clearly sufficient to support the interpretations and conclusions, which are clear and useful for the research community. In general, the paper is well embedded in appropriate international references (see also my minor comment below).

I suggest only minor revision before the paper can be accepted in the Cryosphere.

We appreciate the supportive comments.

Minor comments:

Supercooled talus slopes are an interesting phenomenon in the context of permafrost processes. This is well described and introduced in the paper, but I suggest to mentioned that supercooled talus slopes have been known for a long time (e.g. since the end of the 18th century).

We will include a sentence in the introduction referring to that:

“Abnormally low ground temperatures at relatively low elevations, often indicative of isolated permafrost patches, have long been documented in porous debris accumulations such as talus slopes and, in some cases, relict rock glaciers \citep{Wakonigg1996, morard2010, sawada2003}.”

I also wondered to what extent the distinction between summer and winter situations for air circulation in the talus slopes is too general. In principle, the difference between day and night *often has a similar effect, especially if, for example, long-wave radiation at the surface (radiation night) causes a strong cooling of the upper talus sediment layers near the surface at night. Then a corresponding density gradient of the air in the talus body should also occur. Of course, the time available to drive this process is very limited, but at least it can counteract the daytime situation in summer. Perhaps this process is taken into account in the paper, but I couldn't find it in the text.*

The distinction between summer and winter air circulation regimes in the manuscript is intended to illustrate the model's capability to reproduce the dominant seasonal patterns of convective airflow in talus slopes, where the direction of flow and its thermal impacts are more clearly defined. We agree that similar convective mechanisms can also occur on shorter timescales, such as the diurnal cycle. To clarify this point, we will include the following text on line 318:

" Similar convective patterns can also occur on a diurnal timescale: surface cooling by longwave radiation at night can drive upslope air movement, while daytime heating by shortwave radiation can reverse this flow as surface temperatures rise."

While these short-term processes are not explicitly analyzed in the summer and winter regime simulations, they are indeed accounted for in the developed model through the full surface energy balance. This effect is taken into account in the long-term simulations that are forced with half-hourly meteorological data. We will indicate it in the relevant section of the revised manuscript.

It also appears that the winter and summer simulations shown are driven by constant temperature values and do not take into account daily temperature variations.

Yes, constant external air temperatures were used in the summer and winter simulations, which were designed as simplified test cases to isolate and illustrate the direct effects of higher or lower air temperatures on airflow patterns. This setup was intended to demonstrate the model's behavior under contrasting thermal conditions. However, this is not a limitation of the model itself; it is fully capable of incorporating time-varying boundary conditions. We will clarify this point in the revised manuscript.

In the long-term simulation in Chapter 4.3.1, it is also not clear at which temporal resolution the driving data (e.g. of air temprature, radiation etc.) are used. This should be clarified.

For the long-term simulations the model was forced with half-hourly meteorological data from the nearby (~0.5 km) Opabin weather station. This was mentioned in line 376 in the original manuscript, and we will make it more explicit in the revised manuscript

I am not familiar with the study area presented, but it seems to me that there are simpler situations with supercooled talus slopes (e.g. in the European Alps) where the temperature effects are measurable (and have actually been measured). These conditions would be much more suitable for evaluating the model, as the temperature could be evaluated directly (and air convection could actually be measured). In the supercooled talus slopes of the Alps, the temperature effect is also *very clearly visible in the vegetation. Is there such evidence in the vegetation of the study area in Canada?*

Thank you for this valuable suggestion. In this study, we chose to focus on permafrost dynamics in the Canadian Rockies, an area that has received less attention in the literature compared to regions like the European Alps. We agree that the supercooled talus slopes of the Alps, with their more available data, would provide an excellent context for further evaluation of the model.

Regarding the use of vegetation as an indicator of ground temperature, we note that the study site in the Canadian Rockies is predominantly composed of coarse, blocky sediments with very limited vegetation cover. As a result, it is difficult to delineate thermally anomalous zones based on vegetation patterns. We will add a sentence in the revised manuscript to indicate this.

Equation makes the important assumption that the circulating air is assumed to be incompressible and the temperature is only a consequence of the density (if I understood it correctly). However, if the air in the soil/talus layers sinks downwards parallel to the slope (as indicated in Figures 5, 7 and 9), then it reaches regions of higher air pressure and undergoes compression, which is cannot be neglected over larger vertical distances (or conversely expansion as the air rises). Figure 5 also shows that the air in the talus body is in equilibrium with the outside air (i.e. corresponds to the respective air pressure). Was the assumption made here that the vertical movement is not large enough (maybe 10-20m) and therefore the vertical pressure change is not large enough and is neglected? However, this would limit the applicability of the model.

The model employs the Oberbeck-Boussinesq approximation, which assumes constant air density in the air mass balance equation, while allowing for density variations in Darcy's law (Barletta, 2009; Landman and Schotting, 2007). This assumption is commonly applied in systems where pressure or temperature changes are not significant, simplifying the numerical solution of the problem.

Regarding external pressure, the model assumes that while external pressure does increase with elevation, it does not affect air density. For the relationship between external air pressure and internal air pressure, the model calculates the internal air pressure at each time step based on air density (derived solely from temperature) and external air pressure (Eq. 6), which depends on elevation and acts as a boundary condition.

We acknowledge that neglecting flux compressibility effects over large vertical gradients could limit the model's applicability. However, we believe this assumption is appropriate for the typical scenarios considered in our study. This simplification ensures that the model remains computationally efficient while providing reasonable results for the scale of the problem. We will add a sentence to indicate the limitation of our approach more explicitly.

On page 18 the performance of GEOtop in comparison to the UEB model is described. It is stated that the GEOtop is well performing and show similar snow cover patterns as UEB. However, the difference of simulated SWE is rather large and also the time of maximum SWE is quite different.

It seems that GEOtop is calibrated towards UEB in a way that it represents snow depletion well. I don't believe that differences in snow simulation change results in Figure 12 significantly. But I would rephrase the description of the performance of snow simulations of GEOtop more accurately according to the simulations.

We rephrase the description of the performance of snow simulations to "Figure 12B compares SWE from CE-GEOtop and UEB. Although there are differences in the magnitude and timing of peak SWE, both models capture similar snow depletion patterns, coinciding in the timing of snow-free ground. This suggests that CE-GEOtop can reasonably reproduce SWE dynamics in the Babylon talus slope, supporting its applicability in permafrost-related studies.."

Also the discharge simulations shown in Figure 12 are with significant deviations from observations (could add some skill measures here). This is explained in the text but it also stated that the GEOtop simulations align with observations in June and July. Though it is not easy to see exactly from Figure 12, it seems that also for June und July deviations are significantly (but the daily cycle is strong and give the impression of high coincidence).

We computed the Nash–Sutcliffe model efficiency coefficient (NSE) based on daily values and obtained an overall NSE of 0.48 for the scenario without permafrost (CG-AO) and an overall NSE of 0.55 for the scenario involving permafrost (CG). According to Moriasi et. Al 2015, NSE values over 0.5 for daily streamflow values are considered satisfactory; therefore, following this value, the overall CG simulations give a satisfactory simulation of the observed daily streamflow during this period. If we calculate the NSE every 2 weeks. The NSE is only >0.5 for the period of last week of June and first week of July for the CG model.

To reflect the suggestions and we will change the paragraph to:

"The Nash–Sutcliffe Efficiency (NSE) coefficient was calculated using daily discharge values, yielding an overall NSE of 0.48 for the CG-AO case and 0.55 for the CG case. According to Moriasi et al. (2015), daily NSE values greater than 0.5 are considered satisfactory, indicating that the model incorporating permafrost (CG case) provided a more accurate representation of basin discharge. Particularly, both cases aligned more closely with observed discharge during the last week of June and the first week of July. However, from mid-July onward, contributions from late-lying snowpack in zones not considered in the model became more significant (present in the upper part of the Babylon talus slope as depicted in Fig. 3) . As a result, the simulated discharge showed lower magnitudes than the observed values, matching the observed discharges only during precipitation events. Additionally, when comparing the spring temperatures to the measured data, it's evident that the temperatures from the model with permafrost closely resemble the measured values, while the case without permafrost overestimates the spring temperatures. These findings support the idea that the low water temperature during the summer may be related to the presence of coarse sediments and permafrost occurrence."

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. Transactions of the ASABE, 58(6), 1763-1785.

Equations (13) and (14): wn is not used consistently here (does not appear in equation (13), but is stated afterwards)

We will clarify this in the revised manuscript.

Figures 7 and 9: Please add a x-axis scaling, the arrows of air flow are hard to identify

We will include a x-axis scaling on both figures

The labelling of Figures 6, 8 and 11 could be improved to make it easier for the reader, e.g. indicate the profile name above the figure. A short description instead of the acronyms of the model configurations would also help (depending on the space available).

We will move the profile name to the top of the figure. These figures were thought to be included using only one column in a two-column format, so a short description does not fit on the figure

Figure 12: Please explain with more text

The caption of the figure will be changed to:" Effects of permafrost on discharge rates and spring water temperatures. (A) Hourly simulated and observed discharges. (B) Hourly simulated and observed temperatures. Pink lines show simulation results from the CE-GEOtop case with the air convection module turned off (CG-AO), while blue lines represent the CE-GEOtop case with air convection enabled (CG). In panel A, black lines indicate measured discharge; in panel B, black stars indicate measured water temperatures."