This manuscript presents a one-dimensional Lagrangian firn model, TFM, that solves Richards' equation for meltwater transport and couples water flow with firn densification, grain growth, heat transfer, and refreezing. The authors apply the model to flowlines in the K-Transect and Helheim Glacier regions to investigate the formation and evolution of ice slabs and firn aquifers under idealized climate forcing. The topic is highly relevant because changes in firn hydrology can substantially influence Greenland Ice Sheet meltwater retention and runoff.

The manuscript has several strengths, including the physically based treatment of unsaturated water flow, explicit coupling between meltwater and firn evolution, benchmarking against RetMIP experiments, and an attempt to explore the interaction between accumulation and melt in controlling ice-slab and aquifer formation. However, I have substantial concerns about the experimental design, interpretation of the Lagrangian framework, treatment of permeability at high density, classification of ice slabs and aquifers, robustness of the proposed thresholds, and the extent to which the results can be generalized beyond the three modeled flowlines.

In its current form, I recommend Major Revision. Several of the major issues below should be addressed before the manuscript can support its strongest conclusions.

1. The manuscript constructs all climate scenarios from only RACMO 2018 and 2019 forcing, with 2018 representing a near-balance state and 2019 representing an extreme-melt state. The warming scenario then progressively increases the number of consecutive 2019 forcing years. This is a major limitation because 2018 and 2019 represent two particular meteorological years rather than distinct climate states. Repeating one individual year assumes that all relevant components of accumulation, melt, radiation, temperature, and precipitation retain the same temporal structure indefinitely. Consequently, the resulting "warming" experiment is not a physically realistic warming trajectory. More importantly, the manuscript subsequently derives broader conclusions from this artificial forcing, including the statement that three successive high-melt summers are required for ice-slab formation. This is not demonstrated as a general threshold; it is the outcome of one particular sequence of repeated 2018/2019 forcing.

2. The manuscript states that the Near-Balance simulation produces firn aquifers at Helheim and therefore suggests that these features existed during the 1990s. Similar reasoning is applied to ice slabs along the K-Transect. This inference is problematic. The Near-Balance scenario uses 2018 RACMO forcing, not reconstructed 1990s meteorological forcing. Therefore, the simulation

demonstrates that a climate resembling the selected 2018 forcing can produce these features under the model configuration; it does not establish that the actual 1990s climate produced them. I recommend removing or substantially weakening the historical inference unless the authors can provide an appropriate reconstruction of 1990s accumulation, melt, and temperature conditions. There is also an internal inconsistency: the text sometimes says these features existed during the 1990s, whereas the conclusion states that they existed before the 1990s. This should be corrected.

3. The central methodological innovation is the use of a one-dimensional Lagrangian model in which each point along a flowline represents approximately one year of downstream evolution based on the 2017–2018 velocity field. This is conceptually interesting but also introduces several assumptions that require much more rigorous justification. The authors use a mean 2017–2018 surface velocity field to represent the trajectory throughout simulations extending over many decades. This assumes that: velocity remains constant through time; the flowline geometry remains unchanged; firn profiles follow the prescribed trajectory; spatial changes in climate forcing can be represented by interpolation along a fixed trajectory; the relationship between distance and elapsed time remains constant. These assumptions could strongly affect the inferred inland expansion of ice slabs and aquifers. The authors should provide a quantitative sensitivity test using alternative velocity fields or at least discuss how temporal velocity variability could affect the results. In particular, because the paper's principal conclusions concern spatial expansion along the flowline, the assumed velocity field is not a secondary technical detail.

4. The classification requires an ice layer to be "horizontally continuous." However, the model itself is one-dimensional. It resolves the vertical structure at individual positions along a flowline but does not explicitly resolve the horizontal structure of an ice layer. This raises an important conceptual problem: how is horizontal continuity distinguished from simply observing a dense layer at adjacent Lagrangian locations? The authors need to define mathematically how horizontal continuity is evaluated and explain how this criterion differs from vertical persistence along the flowline. Otherwise, the term "horizontally continuous" is potentially misleading. This is particularly important because ice slabs are inherently two-dimensional or three-dimensional structures.

5. The authors explicitly demonstrate that the Shimizu (1970) parameterization produces relatively high permeability at densities above the pore close-off threshold, allowing meltwater to penetrate supposedly impermeable layers. This is not a minor numerical issue. The authors acknowledge that the model may overestimate aquifer thickness and ice-slab thickness because of this problem. Given that the central scientific objective is to investigate ice slabs and firn aquifers, I consider this one of the most important weaknesses of the manuscript. The authors should ideally perform at least one sensitivity experiment using an alternative permeability parameterization, even if that parameterization introduces numerical difficulties. Alternatively, a sensitivity test could impose a permeability cutoff above a physically justified density threshold. Without such a test, the manuscript should substantially reduce claims about quantitative ice-slab thickness, aquifer depth, and formation thresholds.

6. The manuscript classifies material above 830 kg m^{-3} as ice because this is approximately the pore close-off density. However, pore close-off is not equivalent to conversion of firn into impermeable glacier ice. This distinction is particularly important because the manuscript subsequently uses the 830 kg m^{-3} threshold to identify dense ice layers and infer their hydraulic behavior. The authors should clarify: why 830 kg m^{-3} is appropriate for the classification used here; whether the model explicitly represents pore connectivity; whether density alone is sufficient to define hydraulic impermeability; and how this threshold compares with observations of actual ice slabs. The manuscript itself demonstrates that the chosen permeability parameterization permits flow at densities above 830 kg m^{-3}, which highlights the inconsistency between the density-based classification and the hydraulic representation.

7. The manuscript concludes that three consecutive high-melt summers are required to produce a continuous ice slab and describes this as confirming the hypothesis of Jullien et al. (2023). I do not think the current experiments establish this. A different melt year or initial firn state could produce a different number. The authors should then discuss whether this behavior is robust.

8. The evolution of the modeled firn depends strongly on the initial state. The manuscript describes a spin-up procedure but does not provide enough information for the reader to assess whether the resulting initial profiles are realistic. This is particularly concerning because the authors themselves acknowledge that the Near-Balance Helheim simulation may retain an unrealistically thick firn layer extending toward the calving front. A sensitivity experiment with different initial firn states would greatly strengthen the manuscript.

9. The RetMIP benchmarking is useful, but the validation presented here is primarily qualitative. The authors state that TFM "successfully simulates" firn density evolution, while acknowledging substantial discrepancies in density magnitude and vertical structure. At present, statements such as "qualitative agreement is good" are insufficient to establish model skill. The authors should provide a quantitative evaluation of TFM against the available RetMIP observations and/or model ensembles.

10. The manuscript correctly notes that firn density alone is not sufficient to evaluate a comprehensive firn model and that temperature, grain size, and liquid water content are important. However, the actual validation focuses primarily on density. This is a major weakness because the central innovation of TFM is the explicit solution of Richards' equation and coupled meltwater transport. If the model is intended to make claims about: meltwater infiltration,aquifer formation, water storage, and ice-slab development, then validation against liquid-water observations is particularly important. The authors should strengthen this aspect using available observations from KAN-U, Dye-2, Helheim, or other relevant sites, or clearly state that the hydrological predictions remain insufficiently validated.

11. The model's use of Richards' equation does not by itself ensure a physically realistic representation of firn hydrology, particularly given uncertainties in permeability, hydraulic conductivity, preferential flow, and lateral drainage. These limitations may affect simulated water storage, aquifer persistence, and ice-slab development. In addition, the criteria used to define aquifers and ice slabs and the associated model parameters introduce substantial uncertainty. The authors should better justify

these assumptions and provide sensitivity analyses for the parameters and processes most critical to their main conclusions.

Minor Comments:

1. Section 2 does not clearly describe the forcing-data requirements or how the input data are incorporated into the model. A schematic model flowchart would substantially improve reproducibility and understanding. In addition, the densification scheme appears to be an important component of the model, but it is unclear whether its simulated densification rates have been validated against field observations. Given the complexity of the model and the number of tunable parameters, the authors should also explain how these parameters are constrained and quantify the associated parameter uncertainty.

2. The substantial differences between modeled and observed density profiles in Figures 1–3 require more explicit explanation. In particular, the model appears to underestimate or fail to reproduce some of the observed density variability, and the authors should identify the likely physical or parameterization-related causes rather than simply stating that deviations are inevitable (L252). The periodic up-and-down structure of the x-axes in Figures 1–3 is also difficult to follow and should be presented using a consistent coordinate system.

3. The evolution of the density profiles in Figures 3–4 raises important questions about the Lagrangian treatment of firn layers. For example, the density peak near 18 m appears to remain at nearly the same depth during 2013–2016, while other layers appear to merge, and the modeled profiles in 2015–2016 differ substantially from observations. Similarly, the apparent upward transition of density/ice layers during January 2014–2015 and after January 2016 in Figure 4 requires physical explanation. The authors should clarify whether these features result from advection, densification, ice-layer merging, or the model formulation.