Dear Dr. FitzGerald,

Thank you for reading and commenting on our study.
Please see our response in green below.

Sincerely,
Baptiste Vandecrux on behalf of the co-author

This report is significant in that it compares the performance of multiple firn models in Greenland, run using the same forcing data and boundary conditions, to field observations. The importance of this work, however, is obscured by the poor organization of the report, particularly in the discussion section. Improvements to the organization of the report through the implementation of an interpretation framework and systematic discussion would increase its relevance to the modeling community.

The discussion currently reads as a list of findings, instead of taking the reader through the results in a systematic way. Organizing the discussion using a framework (e.g. in terms of model type, site, result type) similar to that used in the results section would improve clarity. Additionally, it is unclear why some sections are included in the discussion section. For example, section 5.1.10 discusses the importance of the value used for fresh snow density in the model; however, this section cites only previous studies and does not discuss how this is demonstrated by the results of this work.

As suggested, the Discussion Section has been reorganized to cover each of the four sites/climate zones: dry firn, percolation, ice slab and aquifer regions.

Additionally, with the many abbreviations used for model and site names, keeping track of the properties of each model without constant reference to Table 2 is difficult if the reader is not familiar with all the models. Having a framework in which these models are referenced, and a more systematic discussion of results, would improve readability.

We understand the low readability of the model name. Unfortunately it was not possible to change them at this point. Nevertheless, we now always mention the characteristic of the model that we want to discuss, along with its name: e.g. *"only the models with explicit deep-percolation schemes (CFM-Cr, CFM-KM and UppsalaUniDeepPerc) simulate water below 10 m depth"*.

This report is of particular interest to those deciding between firn models to use and those interpreting results from such models. As such, the results of this paper could be presented in a way to highlight the effect that these findings have on choosing or interpreting these models. These findings are touched on in the abstract but are difficult to parse from the discussion. Better organization of the discussion, as previously suggested, would remedy this issue.

In addition to the restructuration of the Results and Discussion Sections we now also summarize and highlight the strengths and weaknesses of different models in the Summary remarks and perspectives Section.