Review of **Emulating the future distribution of perennial firn aquifers in Antarctica**
Veldhuijsen et al. 2024

In this work, the authors develop an emulator that mimics the IMAU-FDM firn model to predict firn aquifer presence in Antarctica. They apply their emulator to an ensemble of different RCM-GCM pairs to evaluate how firn aquifer presence may change in different future emission scenarios. Most of the text and figures throughout the manuscript are clear; however, there are several important methodological details that are not well-explained. I have several major comments mainly related to the methodology. Because there were several things I did not fully understand, I found it difficult to assess the results and discussion. I believe this is an interested and exciting topic and support publication if these concerns are addressed.

Major comments

First, L80 reads: "The bucket method is computationally efficient but does not allow for saturated pore spaces, preferential flow, standing water over ice layers, or horizontal flow". Since a firn aquifer is quite literally "saturated pore space", I think a more in-depth explanation about why the bucket scheme can be used in this work is required.

The paragraph in lines 115-119 is very confusing. I don't understand the reasoning for why it is necessary to introduce a negative value to counterbalance the positive value. Also how are these values 'of the same order of magnitude'? In Figure 2a,b there are many points with LWC above 1000; however I would think the minimum possible value for the number of days without LWC is -365. When you say "we introduce a negative value…", what is meant by this? Do you average this value with the perennial LWC? Or is your model then predicting 2 separate target variables?

Also, why can you not just predict the average LWC during the winter months, instead of the annual minimum? As winter is the least-likely month for liquid water to exist, maybe this would be a simpler way to predict PFA presence?

Finally, from the target variable, what do you actually consider to be a PFA? This is unclear from the text.

Further, in section 4 (L184): It does not make sense to me to report the RMSE as 86 mm or days. I think I just don't understand the target variable but I believe it to be some combination of LWC and days without PFA presence it which case the RMSE would not strictly be 86mm or 86 days but some combination of the two?

Finally, a look into feature importance would be a very valuable addition to this work. Are there features that are more heavily utilized by the model than others for predicting LWC? Which features of the model are most important in cases of firn aquifer presence?

Minor and techical comments

Abstract

As you redefine all acronyms again later in the text, I think it would be better if the abstract did not contain acronyms. This will help make the abstract more clear and approachable.

L5 – "to approximate a firn model" I would recommend being more specific here since your emulator does not approximate the entire firn model but merely a part of it.

L11 – "For SSP5-8.5, PFAs expand to Ellsworth Land in West Antarctica and Enderby Land in East Antarctica" – This is only true for some cases of SSP5-8.5 right?

L12 – "For climatic forcings from RACMO and MAR, we find that liquid water input (melt and rain) and snow accumulation are good predictors for PFA occurrence." – How do you quantify this? I would recommend computing some feature importance metrics to quantify which features are the most important in your emulator.

L14 – specify "air" temperature.

Introduction

L20 – "reduces their buttressing effect" → "reduces *ice-shelf* buttressing effect"

L20 – "Another process to reduce the…" → "Another process that can potentially reduce the…". I think the phrasing of this sentence should be modified a bit because the process of hydrofracture does not directly impact ice-shelf buttressing. Hydrofracture has more of an indirect effect because it can impact the ice-shelf volume (i.e by causing ice-shelf disintegration). This change in ice-shelf volume then reduces ice-shelf buttressing.

L25 – Maybe it would be better to write this as: "Currently, 60% of the ice-shelf area buttresses upstream ice and…"

L28 – "Perennial firn aquifers…, potentially causing ice shelves to break up". Please cite Montgomery et al 2020 or the Firn on Ice Sheets review paper (https://doi.org/10.1038/s43017-023-00507-9) here.

L42 – Fix "climaticx"

L43 – I think the statement "The absence of ice shelves along most of the north-western AP coast also suggests that the combination of ice shelves and firn aquifers is not viable" is misleading… There is no evidence that suggests that the NW AP has no ice shelves because of PFAs (which is what is indicated by this statement). Instead, the atmospheric and oceanic conditions on the NW side of the AP contribute to the lack of ice shelves. It may be that the climatic conditions which make ice shelves unviable are the same as those which allow PFAs to form. I find this statement misrepresentative as currently written.

L60 – add “properties” after “firn”

Methods

Some details in the IMAU-FDM firn modeling section are missing. What sort of spin-up do you do for the firn model? How thick are the firn layers in your model? Are they add steady state when you begin your simulations?

L95 – “see next section” → “(see section 2.2.1)”

L123 – “These are the most…” → “These input features describe the most…”

L125 – “For the summer, we consider snowfall instead of snow accumulation due to the likely presence of evaporation along side sublimation, which has different implications for PFAs”. I don’t follow this reasoning for including DJF snowfall, but annual and MAM snow accumulation. Why does this need to be different for different periods of the year?

L130 – How many total input features are there then for your model?

Section 2.2.3 – Which parameter is optimized? R2? RMSE?

Emulator tuning and evaluation

L 178 – This reference should be Table 2
L179 – Do you also leave out the 30-year rain input feature?

L184 – I think the R2 value here should be 0.89.

L187 – What is meant by “the individual locations” in this sentence?

L188 – “The performance is poorest on and around the Larsen C and D ice shelves…” This makes sense to me because you are asking the model to extrapolate to warmer climate conditions that are likely not seen during the training which utilizes lower emission scenarios.

Results

L226 – “Notably, HIRHAM-EC-Earth predicts at least twice the initial (2015)…” Is EC-Earth much warmer in the present climate? Higher precipitation?

L242 – “For example, on Larsen C ice shelf most aquifers are predicted by MAR…” What is meant by “most aquifers” here. “Most” compared to what?

Section 4.4 – Why are some firn aquifers transient? From the snow modeling output, what happens to make the aquifers disappear?

L291 – “As future warming leads to increased melt accumulation, the emulator is expected to produce more accurate PFA predictions…” This seems highly speculative and I don’t fully understand the logic behind this statement.

Discussion

L325 – CESM2 also has a high precipitation bias

L331 – “Thus, the importance of including PFAs when assessing the timing of ice-shelf vulnerability also decreases” Can you elaborate on this? I don’t fully understand.

Figures

Figure 2 c/d – The x and y-labels are “Years with PFAs” but the units are “mm”.