The Cryosphere Discussions

The Cryosphere Discuss., referee comment RC3
https://doi.org/10.5194/tc-2021-396-RC3, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Review of tc-2021-396, can the 2 networks be merged into a single regression network?

Guillaume Jouvet (Referee)

Referee comment on "Predicting ocean-induced ice-shelf melt rates using deep learning" by Sebastian H. R. Rosier et al., The Cryosphere Discuss., https://doi.org/10.5194/tc-2021-396-RC3, 2022

This paper proposes a novel method to substitute a physical model that predicts ice-shelf melt rates from geometry, temperature and salinity fields by a deep learning emulator. The strategy is to use a state-of-the-art ocean model (NEMO) in order to generate a large variety of input/output pairs of data, and to train an Artificial Neural Network (ANN) (two ANN here). Using ANNs instead of NEMO permits to save huge computational time with moderated loss in accuracy.
I have no doubt that the method proposed by the authors will be of high interest for the community considering the current need for physically complete and computationally efficient ice sheet models -- especially to design the new generation of models for glacier evolution and sea-level rise prediction.
Deep-learning surrogate models have already proved their worth in other disciplines (several orders of magnitude speed-ups compared to their instructor models are often reported in the literature). This approach has been shortly used in glaciology to emulate the ice flow dynamics (Brinkerhoff and al., 2021, JOG), (Jouvet and al., 2021, JOG). The proposed application by the authors sounds therefore relevant.
I found the paper overall well written and convincing. I have mostly one major comment about the machine learning approach, (my first one below), which does not question the overall relevance of the paper. I give below specific and minor comments that I hope will help the authors to improve their manuscript.

Main Comments:

As you design an ANN mapping 2D to 2D fields with continuous variables, the most logical and intuitive to me would be to use a standard Convolutional Neural Network (CNN) trained as a regression problem with a L1 or L2 loss (e.g. similarly to the CNN I use to learn ice dynamics). You may have also considered a U-NET architecture as well to better capture underlying multiscales, if any. Therefore, my main point is: \textbf{why do you split in two networks?} -- a first segmentation/classification and an auto-encoder (AE) -- I just do not see what this brings except unnecessary complications (and probable loss of information!). Unfortunately, I could not find any line of justification for this choice, namely transforming the problem into a classification one, and then afterwards recovering the lost information (or 'corrupted' as you term it) by an AE. \textbf{To me, the final

paper should either i) try to simplify their approach using a single and simple regression network if this proves to be as efficient OR ii) clearly justify the choice of going to a more complex network sequence and explain why the simplest approach was unsatisfactory}. In case of i), consider revising the paper title and removing references to segmentation.

The most convincing to me is the fidelity result of the ANN to the instructor NEMO, but then I think it would be good to clearly give clear numbers and report it in abstract and conclusion. You may choose a metric and state how far (in %) is MELTNET solutions from NEMO? By contrast, I unsure that comparisons with other simpler models should be too elaborated. E.g. Fig. 4 is useful as it shows that the loss in accuracy between MELTNET from NEMO is small/negligible compared to the discrepancy between low and high complexity models (PICO vs PLUME). I think that is enough as I expect the paper mostly to focus on the accuracy of the MELTNET to reproduce its instructor model -- the in-between model comparisons being a substantial task to make sure this is done fairly (I don't have the expertise to assess this). From Fig. 4 I retain that comparing MELTNET with other models is roughly the same as comparing NEMO with others as the two are (hopefully) very close to each other (as the ANN makes a very good job). This also means that the rest is a pure comparison of models no longer involving deep learning, and this may go beyond this scope of the paper. In conclusion, I would probably keep the comparison with PICO & PLUME rather concise, and favor MELTNET/NEMO comparisons. \item The main point of using deep learning emulators is the huge computational gain versus minor loss in accuracy. While you have quantified the accuracy (Fig. 4), it is a pity that you do not do it for computational time. What speed-up? I expect several orders of magnitude. Quantifying the computational time is essential for your paper. You may also comment on the fact ANNs run extremely well on GPU (which is not the case of CPU), giving another important advantage of your method (compared e.g., to NEMO which may not take the same advantage on GPU).

I think the paper can be made more efficient by moving technical machinery in Section 2.3.1 to appendix. The generation of synthetic geometries is necessary, but of lower interest. Moreover, using a GAN is an elegant strategy, but this is probably nonessential.

I think Section 2.2 should come first for the sake of clarity. It sounds more logical to first describe the physical model, and then the ANN you design to learn from the physical model as the choice of the ANN architecture is motivated by the type of emulated physics.

Why not using Antarctica and Greenland real topographies to generate ice shelf geometry? This would avoid to generate synthetic geometries?

Minor comments:

For clarity, I think you should call MELTNET like NEMO-trained MELTNET or at least include NEMO as you may train MELTNET with other models.

l76: suggest "the inputs and NEMO resulting melt rates ..."

l79: fix the typo with "paramterisations"

"These filters are learnt ..." not sure this is understandable for who is unfamiliar with ML vocabulary.

l127-128: why not cropping the area of interest instead of weighting? Anyway, you normally can feed your ANN with different frame dimensions.

l129: You can state clearly you do data augmentation.

l169 2000 m

l273 Figure 2 presents THE main result of the study, ...

s the AE a U-NET architecture? If yes, you should say it.

l 352: you may call it PINN for Physically Informed Neural Network.

l 345: ~ 502 km^2