The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2019-50-AC2, 2019

The Cryosphere
Discussions

Interactive comment on "Estimating Snow Depth on Arctic Sea Ice using Satellite Microwave Radiometry and a Neural Network" *by* Anne Braakmann-Folgmann and Craig Donlon

Anne Braakmann-Folgmann and Craig Donlon

anne.bf@gmx.de

Received and published: 25 June 2019

We thank the anonymous reviewer for his insightful comments and ideas to improve the manuscript.

The manuscript introduces the use of neural networks into the space of snow thickness retrieval on sea ice. The authors compare their results to previous methods including the most recent methods on this subject. In addition to the snow depth algorithm evaluations, the authors present the influence on an ice thickness estimate from CryoSat and compare it to the widely used Warren climatology. The authors also give an outlook towards the possibilities of the joint forces of the candidate missions CIMR and

CRISTAL using their methods. The research is well conducted and the manuscript is well written and is suitable for publication after minor copy editing and addressing the following comments:

General: 1. The neural networks are trained using a small amount of data. According to the text, the data was split into train, validation and test data sets. Did you somehow ensure that that similar values of snow depth occur in all three splits? In best case the histogram of snow depth in each of the splits should be similar. Did you try different splits and compare the results?

We had a look at the histograms of different splits and added a few comments on this in section 3.1 (p.10, ll.10-15, p.11 l.1):

"[To train the networks we temporally divide the OIB data into a training- (70%), a validation- (15%) and a test (15%) dataset.] **This is a common splitting in machine learning and ensures enough training data when the overall amount of data is small. We also verified that each of the splits contains a similar range of snow depth values and that their histograms look alike.** [Figure 3 shows the flight tracks and the measured snow depths from the 2013-2015 campaigns] **(the overall dataset).** [The top right box illustrates which parts are used as test data] **and the snow depth values occurring in this split.** [We end up with 755 valid snow depth measurements in 2013 and 2014 for training, 162 valid measurements in 2014 and 2015 for validation – meaning the identification of the best network architecture – and 162 valid snow depth measurements in 2015 for testing. When we train the AMSR2 + SMOS neural network, we have to discard all areas (especially the bigger hole at the pole), where no SMOS data is available. Again we split the remaining data into 70% for training and 15% for validation and testing each] **and confirm similar histograms of all splits."**

We found that a 60-20-20 split looks similar, but would give less training data and an 80-10-10 split for example yields histograms that are less alike. Attached (Figure 1) you can find the histograms of our 70-15-15 split. For the AMSR2+SMOS

NN we have less data available due to SMOS's bigger hole at the pole. However, a 70-15-15 split of the remaining data also gives histograms very similar to the ones shown. For the paper, we believe that figure 3 is sufficient to show which values occur in the overall and in the test data set.

2. It is unclear to me why the GRs and PRs are used as input. One would expect that the neural network would figure out the relations and adjusts the weights accordingly during the training process. Did you try higher complexity of the networks when you used brightness temperatures as input?

We also tried more complex neural networks when using brightness temperatures as input, but they have reduced performance. Our explanation for this is that a more complex neural network also implies that more parameters have to be learned with the same (very limited) amount of training data.

3. For comparisons with Models and also for the uncertainty values of ice thicknesses from CryoSat it is quite important to have uncertainty attached to each retrieved value. Can you think of a method estimating uncertainties for the neural network based snow thickness retrieval? It would be good to have a statement about this in the manuscript.

We divided the former chapter 4.1. into two subsections and added a third subsection on uncertainty estimation at the end of chapter 4.1 (p.18, l.10): "Finally we assess the uncertainty of our neural networks to enable usage of this snow product in models or for SIT calculation. The very complex and highly non-linear relationship between the input and snow depth output hinders a stringent variance propagation. Instead, to assess the uncertainty of our neural network approaches, we employ the Monte Carlo method and generate an ensemble of 50 samples for each input brightness temperature. We draw these samples from a normal distribution using the observed brightness temperature as mean and 0.5 K as standard deviation for AMSR2.

For SMOS we take the standard deviation provided in the L3 files for each obser-

vation and propagate them through our averaging process to obtain a standard deviation for each SMOS measurement used as input to the polarization ratio. The mean of these standard deviations is 0.76 K for V polarization and 0.79 K for H polarization. Further uncertainty arises from the tie points used both in the Nasa Team algorithm for SIC and to correct the open water part of the footprint (Eq. 2). Therefore we also create an ensemble of 50 samples for each tie point using the values from Ivanova et al. as mean and 3 K as standard deviation. The resulting mean uncertainty in SIC from the Nasa Team algorithm is 4 %.

We then estimate snow depth using each ensemble member as input to our neural networks. This yields an ensemble of snow depth estimates. The standard deviation of this ensemble is used as an uncertainty measure for the estimated snow depth value. Across all OIB data the resulting final uncertainty (mean standard deviation) is 0.05 m for the AMSR2-only NN and 0.02 m for the AMSR2+SMOS NN, indicating that the AMSR2+SMOS NN is less sensitive to noise in the input data. The error (repeatability) of the Monte Carlo simulation is 0.0005 m and 0.0001 m respectively. This approach however only assesses how robust the neural networks are to uncertainty in the input data and auxiliary parameters such as the tie points. Further uncertainty arises from training the neural networks with OIB data, which has its own uncertainty and limitations unlike a real ground truth dataset. "

In the conclusion we added a sentence (p.22, l.11): "From a Monte Carlo simulation we derive an uncertainty of 5 cm for the AMSR2-only and 2 cm for the AMSR2+SMOS NN."

We also updated the sample code provided on github (mentioned under code availability) to include a standard deviation for each snow depth estimate (p.23, ll.18-19).

Specific: P1. L.2: .....it is fundamental climate.... -> .... it is a fundamental climate....

Changed

P2. L.34: acts as -> behaves like

Changed

P11. L 23: A few words about sea ice drift as a source for ice thickness variability would be nice.

We added the following sentence:

"In areas of mixed ice types and fast sea ice drift this assumption might not hold, but we want to avoid too many data gaps."

P15. L24: remove either "polar stereographic" or "EASE2". The EASE2 grid is actually not a polar stereographic projection but an equal area projection.

Polar stereographic was removed

P16. Figure7: The AMSR2 and the AMSR2+SMOS neural networks produce very different spatial distribution of snow depth and often by more than 20cm and even show inverse pattern (Canadian archipelago, East Greenland). To believe your statement that the combination of the two neural network would produce good results, a scatter plot between these two networks might be insightful, especially over a longer time span. Also P18 Figure 8 show partly anti pattern between the AMSR2 and theAMSR2+SMOS neural network. P23.

We have produced a scatter plot between the AMSR2-only and AMSR2+SMOS neural networks on all the OIB data (Figure 2). For comparison we also plot OIB snow depth against the AMSR2-only NN and find that both scatter to a similar extend around the AMSR2-only NN results. Indeed, a few estimates are quite far apart, but the same is true for the OIB data. Given the results presented in the paper (Table 2, Figure 5 and 6) we believe, that filling gaps in the AMSR2+SMOS NN with the AMSR2-only NN produces reasonable results. One main purpose

of the AMSR2+SMOS NN development, however, was to investigate how snow depth on sea ice could be derived from CIMR measurements and here we will not encounter the problem of 1.4GHz data gaps.

We added another sentence in Section 4.1.(1) (p.13, ll. 10-11) to stress this point: "[Combining the two networks could also be useful in a practical application to fill the hole at the pole and to still benefit from higher accuracies in regions, where SMOS is available.] **CIMR, however, would cover the whole pole at all frequencies and therefore, the AMSR2 + SMOS neural network would produce no gaps. "**

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-50, 2019.

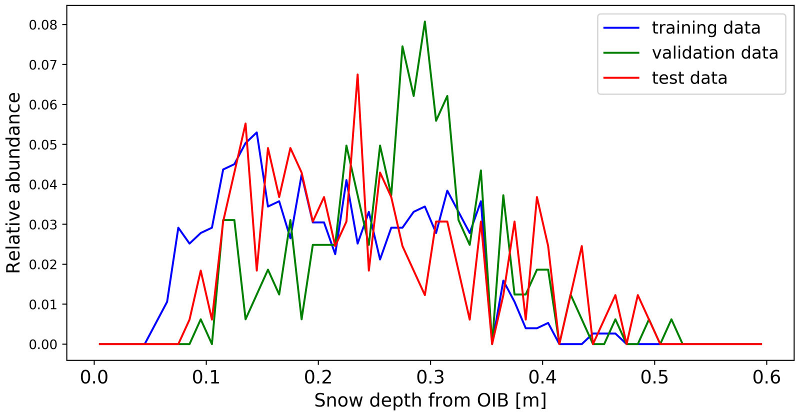

Fig. 1. Histograms of 70% training, 15% validation and 15% test data splits

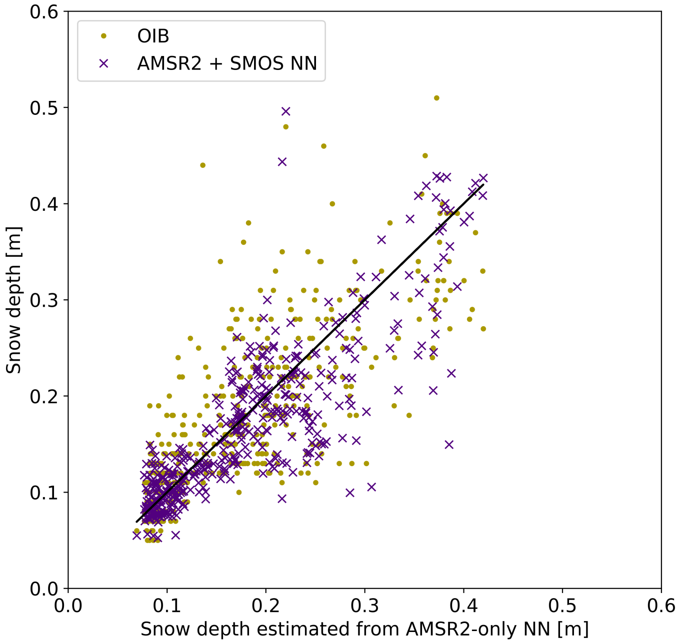

Fig. 2. Scatter plot between AMSR2-only and AMSR+SMOS NN snow depths with OIB snow depth for comparison