The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2019-126-RC1, 2019
© Author(s) 2019. This work is distributed under
the Creative Commons Attribution 4.0 License.

The Cryosphere
Discussions

Interactive comment on "Deriving Arctic 2 m air temperatures over snow and ice from satellite surface temperature measurements" *by* Pia Nielsen-Englyst et al.

Anonymous Referee #1

Received and published: 23 October 2019

The authors present a remote sensed 2m temperature product for the Arctic ocean and the Greenland Ice sheet. For this aim, they use Arctic and Antarctic Ice Surface Temperatures from thermal Infrared satellite Sensors (AASTI) data set and apply a correction to convert surface temperatures into 2 m temperatures. The derived temperatures are compared with in situ observations and the data set performance is compared to the performance of T2m from ERA-Interim.

First of all, I apologize for the long time I needed to complete this review, which added to the long time that the editor needed to find reviewers. Sadly, I'm not convinced that

the authors properly resolve all the challenges that occur when remote sensed surface temperatures are converted into t2m temperatures.

My largest objection is that the authors fail to properly resolve the "cloud problem". Infrared satellites can only measure surface temperatures in cloud free conditions. However, cloudy conditions lead totally different weather than cloud-free conditions, especially in winter. Too little attention is paid to this problem in the methodology and results section. A publishable data set must resolve the cloud-data-gap problem and it should be explicitly shown that reasonable estimates can be provided for cloudy conditions, even if these conditions last for days or more. As the authors do not aim to present a clear-sky T2m product, which would be of limited value, this shortcoming in the methodology and discussion must be resolved prior acceptance can be considered.

Specific major and minor comments:

P3L13: which "data"? I presume satellite data, as AWSs do not move. This must be clearly worded.

P4L10: Snow on sea ice have a major effect on the measured temperature as snow is a very good insulator (e.g. Graham et al, 2019, https://doi.org/10.1038/s41598-019-45574-5). Hence, if the buoy thermistor has a smaller diurnal cycle as the T2m sensor, snow cover is affecting the observation and buoy thermistor should be discarded as valid surface temperature observation. Unless "unrealistic data artifacts" includes damped daily cycles - which then should be stated explicitly -, I believe data scarcity should not be an excuse for retaining incorrect data.

P5L4: All three citations listed to introduce the AASTI refer to technical documents; they do not refer to peer reviewed papers. This is not extremely relevant in itself, but it raises, in my humble opinion, the necessity that the authors restate briefly the methods to compile this dataset. Furthermore, from the title of Høyer 2019, this dataset is only providing clear-sky ice surface temperatures. This must be restated when the AASTI dataset is introduced.

P6L10: It does not become clear to me how these 3-hourly bin averages are aggregated into one daily value. The procedure should be added and described plainly.

P7F3: Although the figure is somewhat instructive, I would be more interested to see (also) **a)** the ratio between cloudy and cloud free observations, as the current figure is clouded by the variations in observation density and **b)** the percentage of days with one or more valid observations within every time interval. Increasing from 1 to 25 observations in a 3-hour interval improves the measurement accuracy, decreasing from 1 to 0 leads to a data gap.
By the way, I am puzzled by the fact that even far North (>75 N), where the polar night and midnight sun periods are long and the daily cycle weak, such a strong daily signal in the mean number of observations is found.

P8L2: Here we are at the end of the description of the skin-temperature data treatment and there is nothing about treatment of data gaps introduced by cloud cover. Please correct me if I am wrong, but if I am right, that is a major omission. Given this absence, my presumption is - I cannot find any clarification in this manuscript - data gaps are left open; if one has no method to fill data gaps these gaps remain gaps. In favor of my presumption, Figure 6 has also regions with no data. Introducing gaps when your method fails positively bias your method performance and introduces an unknown bias in the final result. Again, please correct me if I am wrong; but if I am not, again, this data set cannot be used as all-weather T2m dataset and the paper cannot be accepted for publication until the cloud problem is resolved.
Furthermore, no comments are made in how sub-tile temperature variations due to topography are dealt. I thus presume it is ignored, fine, but state explicitly. It

does affect your correction procedure of 3.1.

P8L6: The discussion of the comparison with in situ observations is a missed chance. It allows you to understand why remote sensed skin temperatures are deviating. Does the correlation improve if the exercise is repeated using valid 3-hourly estimates? If so, then it is a data gap (= cloud) problem. Furthermore, as surface temperatures are very elevation dependent, this must be discussed as many of the PROMICE AWSs are close to the ice sheet margin, thus in terrain in which the elevation potentially varies more than 1000 m in a 0.25° degree tile.

P9L11: It is very common in comparable studies to cut your dataset into 3. In that case, you can perform the training-validation cycle three times; all three data subsets are used once in a training-validation cycle for validation and the remaining two are used for training in that cycle. Why is this approach not applied here?

P10: Equations 4 to 7 provide an elegant approach to evaluate rather simple correction functions, Eqs. 8-12. Still, this is not the best you can do. Why is not a state-of-the-art method like a neural-network approach used? Furthermore, as there are data gaps due to clouds, how are they dealt here? Are these days neglected?

P17F8: The paper gives me no real clue how these data gaps are filled.

As you can see, I am not convinced of the scientific soundness of the approach to convert clear-sky remote sensed skin temperatures into a continuous daily T2m data set. If I would have to do this, I would have taken the following approach: Take the discontinuous 3-hourly dataset of clear-sky skin temperatures, the continuous dataset of the cloudy/clear sky observation ratio, and other sensible remote sensed data products (like cloud properties, shortwave fluxes, atmospheric temperatures, sea ice state, local topography) and put it all into a neural network method (or any other AI) to order to produce a continuous 3-hourly T2m time series, trained with and evaluated against the in situ dataset. In a final step the 3-hourly data are averaged to daily means.

I have read the results section as if the authors had produced a sensible daily T2m dataset, as it is possible that they indeed did this, but failed to convey that to me. At the other hand, since my objections to the method (description) are so major, it makes no sense to do detailed suggestions how the results, discussion and conclusions sections may be improved.

The results section analyses if there are systematic biases as function of the estimated T2m. In a renewed submission, the authors should analyze separately the performance for clear-sky, mixed sky and fully cloudy conditions. It should be proven that a reasonable method is found to estimate T2m for all conditions.

Table 8 and Figure 11 show in my humble view that the data set presented here is not good enough to be used. As e.g. Batrak and Müller (https://doi.org/10.1038/s41467-019-11975-3P) demonstrate, ERA-Interim and ERA5 and other reanalyses do a very poor job over the Arctic ocean due to missing snow cover over sea ice and misrepresented sea ice thickness. For Greenland, I have no paper at hand that does a similar analysis and I am neither aware that ERA-Interim is doing an extremely poor job there too. As you did not mention anything about applying an elevation correction on the reanalysis data – which is essential for a fair comparison, I suspect that overlook might be part of the poor performance of ERA-Interim over the ice sheet. Nonetheless, a useful t2m product derived from remote sensing should be able to beat easily a flawed model product – and yours does not.
Furthermore, as these reanalyzes fail to represent Arctic T2m, they should not be used as benchmark. The data set should be benchmarked against reanalysis results of RCMs optimized for either the Arctic or Greenland. I know there are several colleagues at your institute that can help you in selecting appropriate RCMs and retrieving the data.

Finally, the discussion leaves me puzzled by the fact that the authors are aware of the cloud-gap problem, but try to publish a data set in which this problem is not fully solved (as there are data gaps in the presented data set) and failed to present properly the measures they have undertaken to mitigate the “cloud problem”.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-126, 2019.