The Cryosphere Discuss.,
https://doi.org/10.5194/tc-2019-258-RC1, 2020

The Cryosphere Discussions

Interactive comment on "Evaluation of long term Northern Hemisphere snow water equivalent products" *by* Colleen Mortimer et al.

Anonymous Referee #1

Received and published: 3 January 2020

The authors discuss the evaluation of three types of Northern Hemisphere snow water equivalent (SWE) products, including (i) four reanalysis-based products, (ii) two stand-alone passive microwave remote sensing products, and (iii) one product based on a combinition of passive microwave remote sensing data and in situ snow depth measurements.

The evaluation is primarily vs. a large number of independent snow course measurements from Russia, Finland, and Canada. The authors find that the performance of the stand-alone passive microwave remote sensing products is considerably worse than that of the other products, and only the passive microwave product constrained with surface observations provides comparable performance to the reanalsysis-based products.

Printer-friendly version

Discussion paper

Among the reanalysis-based SWE products, MERRA and the Crocus/ERA-Interim product perform best, suggesting that these products should be included in any multi-product ensemble estimate.

The manuscript discusses an important and still active field of cryospheric research. The manuscript is not ground-breaking and hews closely to the datasets evaluated in Mudryk et al. (2015). However, it includes the AMSR-E stand-alone passive microwave remote sensing products and, if I am not mistaken, the performance evaluation vs. the snow course measurements. These new elements provide, in my opinion, sufficient novelty to warrant eventual publication of the paper in The Cryosphere. However, before I can recomment publication, the authors would need to address the MAJOR issues outlined in the comments below.

Major comments: ————

1) Dataset selection and period

a) Why evaluate MERRA data when MERRA-2 has now been available for 3+ years (Gelaro et al. 2017), and MERRA has been discontinued since early 2016??? There are some differences MERRA and MERRA-2 SWE (e.g., Reichle et al. 2017). As it stands, the reader has to assume that MERRA was used because that is the dataset that was ready to use from the earlier Mudryk et al. (2015) publication. At the very least, the authors need to discuss the existence of MERRA-2, point to the relevant literature and differences, and justify their use of MERRA instead of MERRA-2.

Gelaro et al. (2017), The Modern-Era Retrospective Analysis for Research and Applications, Version-2 (MERRA-2), Journal of Climate, 30, 5419-5454, doi:10.1175/JCLI-D-16-0758.1.

Reichle et al. (2017), Assessment of MERRA-2 land surface hydrology estimates, Journal of Climate, 30, 2937-2960, doi:10.1175/JCLI-D-16-0720.1.

A similar comment applies to the paper's use of ERA-Land and "Crocus", which are both based on ERA-Interim, which has been replaced by ERA-5 and ERA-5/Land (albeit much more recently than the MERRA version change).

b) Why does the analysis stop in 2010 (Table 1)? As far as I am aware, all of the SWE products should be available for several years beyond 2010 (given that AMSR2 extends the AMSR-E record to the present, with only a modest gap). Being a few years behind real-time was ok in Mudryk et al. (2015), but by now 2010 nearly a decade behind real-time, which at the very least requires justification.

2) The discussion of the methodology needs to be improved.

a) As it stands, there are bits and pieces of the methodology in the Results section, and the Methods section is lacking a concise discussion of the various metrics. E.g., lines 193-196, 235-236, and 276-282 belong in the Methods section, and the Methods section needs a complete discussion of the metrics.

b) The temporal and spatial resolution of the metrics calculations is a bit unclear. Line 128 states that all SWE products were regridded onto a 1-deg grid, whereas the snow course measurements are on the 25-km EASE grid (line 161). How are the 1-deg grid cells matched with the 25-km EASE grid cells? And why introduce the 25-km EASE grid in the first place, given that the snow course data are not anywhere near that scale (transects range from 150 m to 4 km), and in any case the 25-km EASE grid is different from the 1-deg grid of the SWE products. Why not use the same grid for the SWE products and the (gridded) transect data? At the very least, this requires justification and clarification.

c) The snow course data are available from once every 5 days to once every month (Section 2.1), whereas the SWE products are available between hourly and daily (which requires better clarification!). Lines 158-161 state that the snow course observations were "converted into bi-weekly [or (over Russia) ten-day] periods". How exactly are the SWE products and snow course data matched in time for the computation of the metrics? Are the SWE products sampled on a single day (1st and 15th of each month), or are two-week (or ten-day) average SWE values computed from the hourly/daily products before the metrics are computed? This needs to be clarified.

d) Lines 138-141: Please clarify whether snow course data are measurements of snow depth or SWE. (The paragraph in question talks a lot about snow depth, but only in the context of the point-scale measurements used in GlobSnow.) Also, if snow course data are snow *depth* measurements, how are the measurements converted to SWE? Using local and contemporaneous snow density measurements? Or climatological snow density values?

e) Fig 2c: It is not clear how the correlation shown here was computed. Is this the spatial average of the temporal correlation coefficient at the individual grid cells? Or the spatial correlation of the time series average? Or all data points thrown into a single correlation coefficient calculation???

f) Lines 235-236: "seasonality [metrics...] were computed at a bi-weekly time step for 2002 through 2010". This is unclear. Based on this statement, the metrics could have been computed in one of the following ways: - subset time series at each location, then throw all values into the metrics computation - subset time series at each location, then compute (temporal) metrics at each location, then spatially average metrics - subset time series at each location, then compute time-average SWE values, then compute (spatial) metrics Which is it?

g) How were zero SWE values treated? Are SWE values excluded from the metrics computations if the snow course and/or SWE product indicated zero SWE? How about cross-masking

h) The number of grid cells ("locations") with snow course data is unclear. According to section 2.1, there are 517 snow course locations in Russia, 200 in Finland, and >1000 in Canada. However, the y-axis scale in Fig. 4d suggests that at most ~100 locations are used for Russia. The discrepancy between 517 and ~100 needs to be discussed explicitly. Is this reduction due to insufficient length of time series, or because the snow course data are ultimately averaged into 25-km EASE grid cells (or 1-deg grid cells)??? How many sites (or grid cells?) were used for Finland and Canada?

i) Lines 276-279: The text here is unclear. Is the metric discussed in line 277 different from that discussed in line 278? In Line 281, in which way are the "anomolous SWE fields" different from the "anomalous snow mass"??? Is not "SWE" synonymous with "snow mass"?? (Or does "snow mass" here refer to the spatially integrated SWE? If so, that is not clear.)

Also, throughout section 3.3 I was confused whether there were two different temporal correlation metrics (one using raw data including the seasonal cycle, and another using data with the mean seasonal cycle removed).

3) ERA-Land and Crocus similarities, and dependence on snow measurements

a) ERA-Land and Crocus use the same forcing data. Including the correlation of the two datasets in Fig 6 therefore artificially elevates the "R4" correlation. Should the ERA-Land/Crocus pair not be excluded from the correlation coefficients contributing to the "R4" value?

b) Perhaps more importantly, ERA-Land and Crocus are *not* fully independent of in situ snow measurements. Both datasets rely on ERA-Interim surface meteorological forcing data. ERA-Interim includes a snow analysis that is based on snow cover data and on in situ snow depth measurements, which impacts the ERA-Interim surface meteorology estimates through, at the least, surface albedo feedback. This needs to be pointed out. (Note that there is no snow analysis in MERRA or MERRA-2.)

4) Lines 75-76 (implicitly) motivates the present study by saying that "[t]o date, these ensembles have relied heavily on models driven by atmospheric analysis and include only a single dataset (GlobSnow) which utilizes remote sensing." However, Line 263 states that "[t]he two AMSR-E products were excluded from this comparison because of the low correlation with the snow course data [...]" That is, the present study is not

really different from previous studies in this regard. This particular motivation of the present study seems therefore invalid.

Minor comments: ————

i) Line 52: Please add a reference for the "temporal inconsistencies" in reanalysis datasets, e.g., Robertson, F. R., M. G. Bosilovich, J. Chen, and T. L. Miller, 2011: The effect of satellite observing system changes on MERRA water and energy fluxes. J. Climate, 24, 5197–5217, doi:10.1175/2011JCLI4227.1

ii) Lines 53-61: Recent results using Sentinel-1 (active) radar data suggest that at least for deep mountain snow much higher spatial resolution snow depth estimates are achievable (Lievens et al. 2019). This should at least be pointed out here, and a clarification should be added that the present study focuses on passive microwave data only. The Lievens et al. (2019) results also suggest that the text in Line 77 may need clarification. Lievens et al. (2019), Snow depth variability in the Northern Hemisphere mountains observed from space, Nature Communications, 10, 4629, doi:10.1038/s41467-019-12566-y.

iii) The nomenclature "NASA Historical" and "NASA Operational" is a bit unfortunate. First, MERRA is (or rather, was) also a *NASA* (quasi-)operational product. Second, the use of *Historical* and *Operational* suggests that "Historical" is only for the retrospective period while "Operational" is for the present and future. However, if I understand the manuscript correctly, "Historical" is really an older version of the NASA AMSR-E retrieval product, and "Operational" is a newer version of that same product. Two of the authors of the present paper are also authors of the "NASA AMSR-E" product. They should know the appropriate version numbers of the NASA AMSR-E products discussed here, and these version numbers should be used in the paper.

iv) In the context of Figure 2 or the corresponding Methods discussion, the number of grid cells with snow course measurements contributing to the metrics should be provided. See also comment 2h) above.

v) Line 346: replace "idealized" with "ideal"

vi) Lines 369-370: The term "NASA AMSR-E *operational* dataset" appears twice, once in each line. Should one of the two be the "historical" dataset?

vii) Line 82: replace "to evaluation" with "to evaluate"

viii) Lines 123-124: The paper should make it clear whether the SWE output from the reanalysis data was used or whether the snow depth output was used (with subsequent conversion to SWE using ancillary snow density values). This is a bit unclear.

Interactive comment on The Cryosphere Discuss., https://doi.org/10.5194/tc-2019-258, 2019.

Printer-friendly version

Discussion paper