Response to comments from Referee #1 for TC-2021-384

Dear reviewer, Please see the response to your input below.

1 General Comments

- *This paper presents first a new data set of ice surface temperature (IST) merging different satellite products and assimilates afterwards this new IST data set to "improve" SMB estimations over the Greenland ice sheet. While the development of such L4 data sets is innovative and scientifically relevant as well as the assimilation into a SMB model, some improvements are needed before potential publication in TC.*

  We thank the reviewer for reading the manuscript and providing constructive comments.

2 About the L4 data set

- *Line 122: Why a radius of 75km? What is the sensibility of using a larger/lower radius?*

  The choice of 75 km search radius was a compromise between selecting a large search radius that ensured enough data to be included in the OI estimate and a computationally feasible search grid box. This search radius is also used in the CMEMS NRT L4 SST/IST product. The key point for this study is that the average number of observations found within the 75 km is very close to the maximum number of observations, which implies that the search radius is not severely limiting the number of observations available for the OI. This has been included in the text, Lines 360-365: "The choice of 75 km for the search radius of the Optimal Interpolation scheme was a compromise between selecting a large search radius that ensured enough data to be included in the OI estimate and a computationally feasible search grid box. The average number of observations found within the 75 km search radius is near the maximum number available of observations, thus the selected search radius is not severely limiting the number of observations available for the OI. The threshold of 75 km is being used for the operational production of the Near-Real-Time Arctic Ocean L4 SST/IST (Copernicus Marine Service, 2021)."

- *Fig 3. Like fig4, all the time series should be shown on the same plot to better allow to compare the different data. Here, in view of the used -60°C +10°C vertical axis, it is impossible to evaluate if a product is warmer/colder than an other one. Moreover, nothing is said about the considered area. Is it the same one for each product or is it an integrated value over available area in each data set (which could be very different). A plot using at least the same area for each day and each data set should be built. Finally, the differences in the passing time should be evaluated. The differences shown in Fig 2 could be due to the passing time which is different in each data set and therefore, these figures can not be compared for me.*

  We did combine time-series of mean daily values and associated standard deviation in one plot, yet it is actually harder to interpret as lines and shaded-areas overlay making it very hard to understand which is which (see figure below, which nonetheless excludes the L3S product). Therefore we refrain from modifying Fig.3 and we keep the y axis limits the same for all panels, so it can be straightforward to compare. For example, using the consistent y-axis limits one can see that for winter mean MODIS and AATSR values are below -40°C while AASTI and the L3/L4 are above. Such guidance is also provided in the text, see lines 180-183: "MODIS and AATSR (when available) show lower ISTs in particular during winter and late autumn compared to the other products, with minimum MODIS ISTs of about -50°C and AASTI and the L3S and L4 IST products reaching their lowest ISTs of -35°C to -40°C. All products, including the L4 IST, well represent the annual cycle with the warming that started in early March and peaked in July, followed by cooling and winter minimum at the end of December."

  Regarding the considered area used to derive mean daily temperatures in Figure 3, please see description in Lines 174-179:

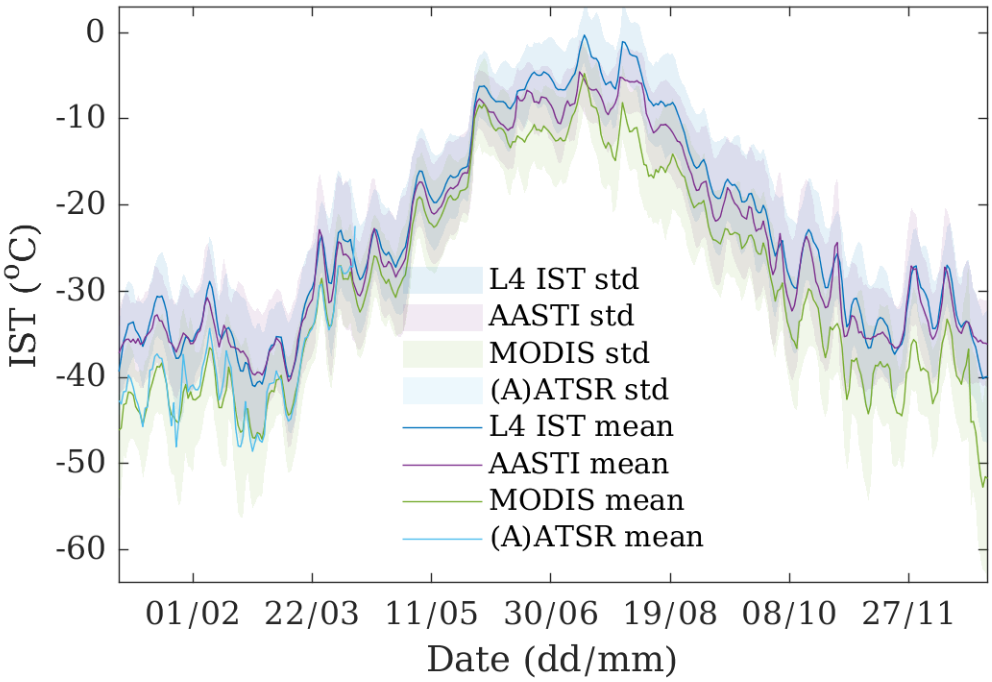

Figure 1: All mean daily time-series combined in one figure.

"To estimate the mean daily IST for each dataset, a mask defining the Greenland Ice Sheet was applied and all valid and available measurements were averaged to a daily value. Therefore, daily mean values shown in Figure 3 are means over the entire area of consideration and while the L4 IST product always has the same number of valid pixels used for the daily mean, the single-sensor products have a varying number of available measurements depending on data quality and cloud coverage. The L3S product is the combination of all single-sensor products so its average is based on all available points from all single-sensor products."

Regarding the differences in overpass times and differences in Figure 2, this is an example plot to demonstrate the transition from single-sensor products (upper 3 panels) to super-collated and optimally interpolated fields (lower 2 panels). Differences between the single sensor products shown in the upper 3 panels are associated to orbits, swaths, resolutions, instrument footprints, IR instrument characteristics, retrieval algorithms, etc. For the single day presented in Figure 2 differences are also associated with the number of available observations per sensor/dataset, e.g. see white areas for AASTI and AATSR and the almost full coverage for MODIS. As an example, for AASTI and MODIS which rely on the AVHRR instrument on multiple platforms and the MODIS instrument on the Acqua and Terra platforms, there are more than 10 overpasses during for a given date. To the contrary, AATSR on Envisat had a revisit time of multiple days so full coverage of the GIS was very limited.

Please see clarification in Lines 169-172: "The coarser spatial resolution of AASTI (top left) compared to MODIS (top middle) is visible, resulting in AASTI grid points with missing information while MODIS daily aggregated L3 data offer superior coverage over the GIS. The sampling of AATSR (top right) with its narrow swath and lower temporal resolution results in characteristic artefacts resembling the ENVISAT platform orbit. Such artefacts do not appear nor for the MODIS neither the AASTI products."

- *Figs 5-6: biases are systematically negative as LST (observed mostly during day) is compared to a daily mean (including night). Therefore, both product are not comparable for me because the passing time of LST is not representative of the daily amplitude of observed temperatures (This issue is moreover mentioned in the conclusion). The passing time should be considered to have a fair comparison or at least, only the day hours should be used to compute the PROMICE average.*

The satellite observations are based on infra-red instruments and not optical, which would result in a day/night bias. Studies have shown seasonal cloud cover dependence over Greenland but not a daily one, see Nielsen-Englyst et al. 2019. In Figure 5, mean values from the single-sensor products, L3S and L4 are compared against mean values from the PROMICE stations. PROMICE mean values are systematically higher, hence the negative biases. As already mentioned in the manuscript, Lines 126-133, the main difference between the IR datasets and PROMICE is the clear-sky vs all-sky conditions, as in the former case colder temperatures are experienced compared to the latter: "The satellite products used in this study represent the clear-sky IST as the IR satellite sensors cannot observe the surface through clouds. As a result, a clear-sky bias is usually observed when comparing averaged clear-sky surface temperatures against averaged all-sky temperatures (Koenig et al., 2010; Comiso et al., 2003). Nielsen-Englyst et al. (2019) used

PROMICE observations to estimate the clear-sky bias introduced when averaging using different temporal windows. Using a 72-hour averaging window, they found a clear-sky bias of -0.96°C when PROMICE stations located in the middle/upper ablation zone and the accumulation zone were used. Here, the clear-sky bias of 0.96°C has been added to the satellite products in order to provide an estimate of the corresponding all-sky daily IST fields, which can be compared to the all-sky ISTs observed by PROMICE and IceBridge."

- *Section 4.2: the mean LST from satellite products over 2012 should be compared in 2D by considering all the available data and by considering only the pixels present in each data set. The differences between the products are due to the passing time? the cloud mask which is different in each data set? or for the same area at the same time, observed LST is different? Moreover, the amount of missing data (still in 2D) in each data set should be compared instead of showing the number of aggregated observations in Fig 10.*

  The number of aggregated observations in Figure 10 aim to demonstrate the number of available observations used to generate the L4 product and to compute the monthly means shown in Figure 9. Nonetheless, we have removed it and included the information in the text, which was also a request from reviewer 2.

  We have now estimated mean IST from MODIS, AASTI and L3S over 2012 along with the associated number of obs used for each dataset. This is now figure 5 and the associated text in Lines 197-202. "The spatial variability of mean annual IST values over the Greenland Ice Sheet for 2012 from AASTI, MODIS, the L3S and L4 OI products are shown in Figure 5 along with the number of observations used to derive the means. For the MODIS and AASTI datasets, the intermediate L3 gridded products were used for the estimates, i.e. the original 1-km and 4-km L2 observations were re-gridded to the 5-km final grid. MODIS mean IST was significantly lower over the entire Greenland Ice Sheet compared to the AASTI estimates, although significantly more observations were available for the former. Mean IST from the L3S and L4 OI products appear more similar to AASTI mean estimates."

3 About the assimilation:

- *The passing time needs to be take into account for me. It is particularly relevant in May when melt occurs sometime only at local noon. By assimilating a daily average, this smooths the daily amplitude in the energy balance model and then the production of melt. It is particularly relevant at the beginning and at the end of the melting season.*

  The L4 IST dataset would ideally be produced in at least 6 hourly timesteps for the purposes of assimilation into our SMB scheme. Unfortunately, this is not possible within the current study and with the limitations of available data, which is why we have focused on using the daily means which have already been produced as an initial test of the system using the L2 data from the ESA CCI LST phase 1. The reviewer is quite correct that this introduced smoothing affects the energy balance calculations and therefore biases the derived SMB. Improving this scheme is part of planned future work. We have made this more clear by adding a few sentences in the text describing the assimilation in the methods section, see lines 158-163: "The L4 IST product is available once daily, yet modelled IST is dependent on the full surface energy balance and therefore highly variable in space and time, assimilating the L4 IST product inevitably introduces some biases. The aim of this study is therefore to act as a proof of concept for the assimilation of satellite derived data into the model. It is also for this reason that we focus on the month of May, a period with highly variable IST and surface melt where the addition of satellite observations is likely to give the most added value in identifying surfaces close to the melting point."

- *During the peak melting season, melt occurs every time and as IST is limited to 0°C, assimilating of not IST does not change the SMB results explaining why the focus has been made here over May.*

  The reviewer is correct that as melting surfaces are forced to be 0°C, there is no change in SMB over surfaces where both model and IST data suggest it is melting. However, it is very rare for the entire ice sheet to melt, so there can in fact be changes in the positive direction if the assimilated IST gives a colder surface than the model during the melt season. In May, the ice sheet typically goes through a period of rapid change as melt starts. We have added this, see lines 158-163 and the reply above.

- *While the aim is to improve SMB and surface melt, nothing is said about the differences in the cumulated melt amount, runoff, SMB, . . . between the control and assimilation run. Only IST is compared between both simulations. But, in term of SMB and melt, it is not clear if this improves or not the results. As the temperature is forced in the SMB model, it is not very relevant for me to evaluate IST only. The melt extent should be compared with a microwave derived product for example to see the interest here of assimilating a daily product. Moreover, integrated over the whole season, what is the impact on the production of melt, runoff and refreezing?*

  Response: The aim of this paper is to evaluate a remote sensing product and show examples of how can it be utilised to analyze the IST over Greenland and to interpret regional climate and surface mass balance models of the Greenland ice sheet. Therefore, as the focus is not on the melt and the manuscript is already extended, according to comments from reviewer 2, we refrain from adding new analyses/figures. Please also see clarification in the Discussion, lines 330-334.

  Future work will focus on this aspect and in refining the assimilation scheme to take into account some of the mentioned features. Nonetheless, we have done a quick comparison with the melt area between the control and the assimilation run compared with the melt area derived from passive microwave observations and published by NSIDC in order to assess the modelled melt area (see figure below). The assimilated IST run shows a closer match with estimates of melt area in western Greenland as well as in the NE and SE, but areas around the NW of Greenland are better represented in the control run. The extra analysis required to understand the biases and implications of the current rather simple assimilation scheme is beyond the scope of this paper but will be included in future work.

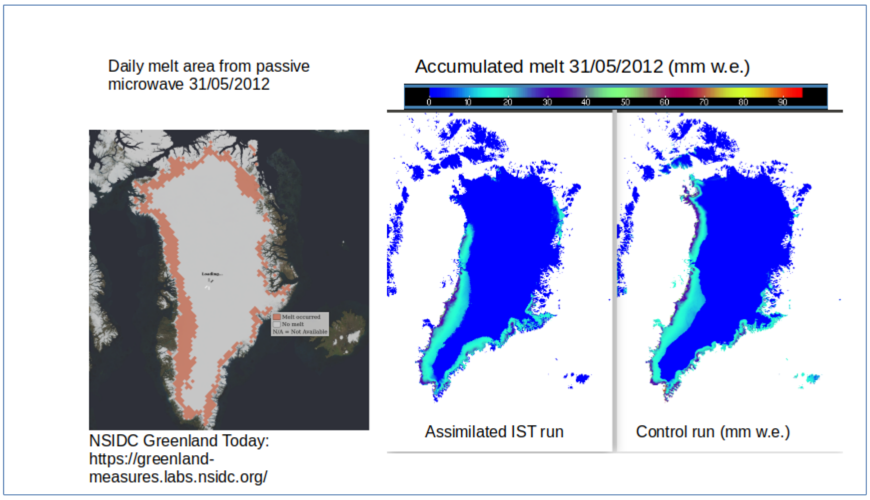

Figure 2: Quick comparison of melt area between control and IST assimilation models with NSIDC melt area derived from passive microwave.