Supplement to author comment on *RC1 'review' by Xavier Fettweis*

The manuscript has been revised, incorporating changes from all Interactive Discussion comments. Versions of the manuscript and supplement with tracked changes are included alongside this document. Major changes to the manuscript include:

1. What was Figure 3 has been moved to supplementary material. Two extra stations have been added, and lines made clearer.
2. Two new datasets have been added to the analysis: Box (2013) data, and MAR v3.5.2 forced by ERA-20C.
3. Time series of summer ice-sheet average surface air temperature have been added as a new panel in what was Figure 7 (now Figure 6).
4. Time series of annual surface air temperature at several long running coastal stations have been added to supplementary material.
5. Maps of grid point trends have been added to supplementary material.
6. Annual mean of monthly bias and monthly mean absolute error has been added in a new table.

Point-by-point responses to referee comments are given below.

Referee's comments are reproduced in black.

Authors' responses are in blue.

Interactive comment on "Evaluation of Greenland near surface air temperature datasets" by J. E. Jack Reeves Eyre and Xubin Zeng

X. Fettweis (Referee)

Received and published: 23 January 2017

This paper presents an interesting evaluation of the ability of a lot of reanalyzes, RCM and observations based datasets to represent the current near-surface temperature variability over Greenland. The paper is well written and deserves to be published in TC.

In addition to the Jason Box's suggestions (that I fully support), I have also several recommendations before publication:

- pg1, line 8: The MARv3.5.2 model should be explicitly cited in the abstract.

Due to the large number of datasets used, specific examples are only mentioned in the abstract in relation to key conclusions. Citing MAR would require citing all other datasets too, which we feel would be inappropriate for the abstract.

- pg2, line 9-11: the sensitivity of the MAR results to the reanalysis used as forcing has already been discussed in depth in

Fettweis, X., Box, J. E., Agosta, C., Amory, C., Kittel, C., and Gallée, H.: Reconstructions of the 1900–2015 Greenland ice sheet surface mass balance using the regional climate MAR model, The Cryosphere Discuss., doi:10.5194/tc-2016-268, in review, 2016.
and in particular for the simulation of the near-surface temperature (see Fig 7 of Fettweis et al., 2016). This works should be cited here and one of the goal of the present work is rather to extent the analysis of Fettweis et al. (2016) because a part of the proposed aim has already been made in Fettweis et al. (2016).

We thank the referee for their important work on this subject and for making model outputs publicly available. Fettweis et al. (2016) has been cited and the results and conclusions have been factored into our work.

- pg 2, line 19: it is the version 3.5.2 which is used here. Fettweis et al. (2016) must be cited here.

Done.

Why do not include also MARv3.5.2 forced by ERA-20C in this evaluation ? Outputs are also available on my FTP.

This model run has now been included in Figs. 4, 5 and 6, and Tables 1, 3 and 4. Discussion of these results has also been added.

In Table 1, change MAR-20CR by MAR-20CRv2c.

Done.

- pg 6, lines 14-20. I agree with the authors that Summit is a good candidate to evaluate the near-surface temperature. But, afterward, evaluation is done over all stations. Therefore, I don't see the interest of this small paragraph and of Fig3 which could be put in Supplementary material.

This section has been re-written and the relevant Figure has been altered (in line with J. Box's comments) and moved to the Supplementary Material.

- pg 9, lines 17-25. This paragraph must be reformulated. Firstly, it is true that MAR is worse than reanalyses at the coastal stations but these DMI observations have been assimilated into the reanalysis and not into MAR! Secondly, MAR has been developed for well representing the ice sheet conditions and not the oceanic conditions. Sea ice cover and SST are forced into MAR because their impact on the SMB is low (Noel et al., 2014, TC). When there is sea ice, the sea ice thickness is prescribed to 1m in MAR and MAR computes itself its surface temperature. This explains why the results of MAR at the coastal stations (fully impacted by the neighborhood ocean conditions) are particularly bad in winter because using a fixed sea ice thickness of 1m overestimates the real sea ice thickness in most of the places and therefore allows a extreem cooling of the surface temperature in MAR. In summer, there is less sea ice, coastal near surface temperatures is more impacted by the SST (which is prescribed into MAR) and therefore the MAR results are better (because it is more constrained). This issue with sea ice in winter has been corrected in part in MARv3.6.

This paragraph has been re-written to relay the explanation of coastal biases. However, the finding (that MAR performs worse than reanalyses at coastal stations) still stands and, we feel, should be reported. Further, to the best of our knowledge, it is generally not true that DMI (or any other network's) SAT observations are assimilated in reanalyses, with the exception of ERA-Interim (in which they are used to update surface properties as part of its land surface analysis scheme; this may explain ERA-Interim's relatively low MAE at DMI stations). In addition, SAT observations *are* used in the gridded SAT analyses.

- pg 12, lines: 13-14: Done in Fettweis et al. (2016).

This sentence has been removed.

- All the datasets should be evaluated also at the summer (JJA) time scale as already suggested by Jason Box. It should be good to have the equivalent of Fig 7 for JJA.

A summer (JJA) panel has been added to Fig. 6 (old Fig. 7).

- The mean biases for all CMIP5 model should be listed in Supplementary. What are the best CMIP5 models ?

The ice sheet annual mean SAT from CMIP5 EMSs are already listed in Table S2 and, alongside the same quantity from an observational-based dataset, we think this gives an adequate idea of which CMIP5 ESMs are most and least biased. We refrain from naming best CMIP5 ESMs as this is dependent on which comparison dataset is used – indeed, this fact is one of our conclusions.

- In addition to Table 3, mean bias as well as RMSE over 1980-2010 of all data sets used here should be listed for both ice sheet and coastal weather stations.

A new table has been added to showing mean bias and mean absolute error.