Supplement to author comment on *SC1 'review of Eyre et al. TCD 2017' by Jason Box*

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

JEB Box

Received and published: 19 January 2017

The study brings together a large set of climate modeled surface air temperature output and compares the data with in-situ field observations at monthly temporal resolution.

The writing is clear.

major comments
a.) The fact that the AIRS data are clear-sky means biases the more cloudy the all-sky retrieval is, as represented by field data. How large is that bias? It's seasonal range is low bias in summer and how large is the bias in winter, deg. C units?

The goal of including the AIRS data in this work are to assess its overall suitability for air temperature monitoring over Greenland. We feel that attributing the bias to different sources (e.g., clear-sky versus all-sky; local time of retrievals; biases from retrieval algorithms) is beyond the scope of this study, and that the assessments presented and comparison with previous results (Koenig and Hall, 2010) are valuable as is. Nonetheless we extend the discussion of AIRS in Sect. 3.3.

b.) "Two regional data sets are also included", the the study neglects to compare field data with the Box (2013)* data that are worthy to compare because the 1.) span decades earlier than compared data; 2.) are in similar class of compared data GISTEMP, BEST. 3.) p. 9 line 25 “the benefits for SAT of RCM downscaling are not clear” comparing with Box should help since Box mimics RACMO2. The Box data are available at:
http://research.bpcrc.osu.edu/~jbox/Box_2013_Reconstruction_data/Box_Greenland_Temperature_monthly_1840- 2014_5km_cal_ver20141007.nc
* Box, J. E. 2013. Greenland ice sheet mass balance reconstruction. Part II: Surface mass balance (1840-2010), Journal of Climate,Vol. 26, No. 18. 6974-6989. doi:10.1175/JCLI-D-12-00518.1
Along this line, the paper should compare these data along side the those in Figure S2. p. 8, line 1 could include the Box data next to these others.

We thank the referee for making this data available, and include results from the dataset in Figs. 4, 5 and 6, and Tables 1, 3 and 4. Discussion of the results is also included.

c.) Figure 3; selected sites are 3/4 W Greenland. Comparison more meaningful to select 1 site from NE Greenland, 1 SE, 1 Central, one SW or NW. So, recommend to keep Swiss Camp, use a KPC_ site, use a TAS site, use Summit, may be also use a THU_ site. recommendation: Figure 3, scatter plots are desirable with dots represented by the month name

We thank the referee for this recommendation and alter the stations included in Fig. 3 accordingly. However, we retain the figure format as time series, in order to give an impression of the seasonal cycle and the effect that dataset biases have on their representation of seasonal cycle. This figure has now been moved to the supplementary material.

d.) you compare AIRS SAT retrievals, but why not also MODIS MOD11 as in the Hall et al work? You could justify not comparing with in-situ data because Hall et al already did. Then again, if you’re after a comprehensive comparison and ‘beauty contest’, would be worth knowing which were more accurate AIRS SAT or MOD11.

We use AIRS and not MODIS for the reason that, nominally, AIRS includes a near-surface air temperature (SAT) product, while MODIS MOD11 is a surface temperature (LST) product. While this difference might seem somewhat artificial, we feel the introduction of further biases by comparing SAT with LST will reduce the value in including MODIS MOD11.

e.) for more impact, seems worth more analysis of July or June through August temperatures since the headline-grabbing melt issue is more strongly tied to this part of the year. In discussion and conclusions, what novel melt season findings you deduce could get some attention through the frame of Greenland melting as a societal risk factor.

Time series of summer mean SAT have been added to Fig. 6.

f) try to more clearly distinguish spurious trends from real trends in 20CR and ERA 20C vs long term coastal DMI observations

Time series of annual mean SAT from long term coastal DMI observations and (nearest grid point of) SAT datasets have been added to Supplemental Material.

minor comments
more clearly introduce NANSENSAT; appears abruptly p 7 line 14 without introduction

Brief introduction in Sect. 2.2.

p. 3, line 26 remove "heavily"; adverbs are vague and unnecessary

Done.

4, line 17 remove "around"; unnecessary

Done.

p. 7, line 7... “we are interested in changes in bias”... please more clearly frame why that is in your methods section.

We have clarified in Sects. 2.2 and 3.1 that stationarity of bias over time is relevant to the credibility of long term trends and variability.

p. 11, line 6 “improved performance”; quantify that statement with a number, i.e. an x % reduced MAE and/or a x increase in correlation and/ or x reduction in bais. consider to rank tabulated MAE values to more clearly display which datasets are most accurate

Improvements due to elevation corrections have been quantified in Sect 4.