Detailed response to Anonymous Referee #2

Note: reviewer comments are given in bold and our response is given in normal type.

1. Using the GC net stations as representative samples of ablation, percolation and dry snow zone is problematic. Rather than being representative of a zone, they can be considered representative for a geographic region of varying sizes. The three stations from the ablation zone are in close proximity of each other on the west coast. All percolation zone stations are in the south, all dry snow zone stations are in the north. Some more rigorous analysis is needed before these stations can be assumed representative of the three regions if at all. I suggest another approach that focuses on model and station comparison rather that generalizing over the three zone. If the authors want to generalize about the three zone a more rigorous analysis of the representability of the stations for each of the zones are needed.

We agree with the reviewer that our station groupings are geographically proximal, however these groupings do also coincide with the ice sheet's melt zones; particularly the dry snow zone, which is mostly in the North of the ice sheet, and the percolation zone, the largest expanse of which is in the South. We have added contours to Figures 1 and 5 to show the coverage of these zones with respect to the distribution of stations.

We appreciate that there are only two stations in the ablation zone, albeit located in the region where a majority of ice ablation occurs (McMillan et al., 2016). We acknowledge the sparsity of the Gc-Net stations with respect to the size of the ice sheet on line 204 of the original manuscript and have added the following in line 220:

'Because we only have data for two ablation zone stations which are located in close proximity, further work is required to assess whether this is a general property of the ablation zone or restricted to this location; temperatures in general are much warmer here…'

2. Provide information about time span in addition to time series length for each of the GCnet stations and discuss implications of varying time series length and time period span on the extreme value statistics.

We have included a new figure (the new figure 1) to show the data coverage at the GC-Net stations used here, and amended Table 1 such that it shows the total amount of data (excluding gaps). We have also added text to introduce the new figure, refer explicitly to the table and discuss the significance of the missing data beginning on line 70 as follows:

'Our analysis focuses on 14 of the 18 stations; we found the remaining 4 stations to have temperature time series which were either too short or too patchy for robust statistical analysis. Figure 1 shows the data coverage at each of the 14 stations studied here and Table 1 gives the total number of years of data available, when gaps are excluded. We attribute these missing data to equipment failure and assume that it is unrelated to the occurrence of extreme high temperatures. As such we treat these data gaps as `missing at random' and ignore them in our analysis. Since most of the missing periods cover whole years, rather than just a summer- or winter-period, this assumption is reasonable. '

It is true that varying time series length and span may have implications when comparing extreme event characteristics between sites; for example it is difficult to pick out a trend in extremes in shorter time series. In this study we did not find any evidence of a trend in extremes at any of the stations and in any case the main focus of our paper is on the comparison between GC-Net and MAR, both of which were commonly sampled at each station and subject to the same EVA procedure. We are hoping to investigate temporal trends in extremes in future work however and so we have amended line 260 to read:

‘further work is needed to determine the relative contributions of potential physical drivers of extreme events at different locations and over different time periods’

3. The analysis of melt energy and extreme temperature events needs some work because extreme temperatures at several of the stations appear to not be associated with melting at all (i.e. Figure 1).

We are not sure what the reviewer is asking for here as Figure 1 does not consider melting; it presents extreme event characteristics at each of the stations.

Figure 5 shows PDDs at each station and attributes them to extreme events (blue portion of stacked bar) or normal conditions (orange portion of stacked bar). Extreme events produce PDDs at all of the GC-Net stations we consider in this study. The MAR model variants do not produce any melting at some of these stations. This is discussed in the manuscript on lines 199-201 and 250-256.

4. A clear presentation of the analysis behind the conclusions that MAR simulates duration of extreme temperature events but not frequency or magnitude of those events are not well supported with figures and tables. It seems that Figure 2 and 4 are indented for this purpose, but they are not clear (see more comments on the figure design below).

We have edited the figures in line with the reviewer’s suggestions (see specific comments below) and added an additional reference to Figure 2 at line 143. This should be clearer now.

5. The analysis discussed in lines L242 to L248 belongs in the result section and needs some more elaboration to be convincing. First, a figure showing the amount of melt energy during extreme versus average conditions would be really nice to see. Second, you have to address the fact that some stations have extreme temperatures that are not above freezing and therefore no PDDs.

These points are discussed on lines 186-201 in the results section and the corresponding data is presented in Table 4. The stacked bars in Figure 5 also illustrate the PDDs during extreme (in blue) vs non-extreme (in orange) conditions. Hopefully the edits we have made to this figure will make this point clearer.

Minor comments

L14: Clarify that you are examining extreme positive temperature anomalies (as opposed to positive and negative)

Edited to read ‘extreme positive temperature‘

L92: Clarify what MAR grid cell elevation that lower than the AWS, e.g. the center point?

‘data’ changed to ‘cell centre’ to clarify.

L105: Explain PDD, the concept may not be widely known outside glaciologist circles.

Edited to read: ‘Melting is most appropriately calculated as a function of the surface energy balance; however measurements of variables required to calculate the surface energy balance (e.g. net radiation, wind speed) are not consistently available at the Gc-Net stations. Positive Degree Days are an estimate of the magnitude and duration of above-zero temperature events and are typically well-correlated with melting (e.g. Braithwaite (1995), Huybrechts et al. (1991)). Here we calculate positive degree-days (PDD) for both observed and modelled temperatures and take this to be a reasonable approximation for melt energy'

L108: Provide more background for equation 1. Typically PDD are a function of mean daily temperatures and a temperature to melt conversion factor.'

We choose to calculate PDDs by integrating a model of daily temperature variability because we are interested in potential melt that occurs during unusually warm conditions. It may be that the mean temperature for a day is 'normal' but the daytime has been unusually warm and the night time has been unusually cold, for example during persistent high pressure conditions in summer. As such we could 'miss' melting that occurs during the warmer part of the day if we use an equation based on mean daily temperatures. This concept is not new as a similar approach has been applied to higher resolution (i.e. sub-daily) data in studies of Antarctica (e.g. Vaughan, 2006, Barrand et al., 2013)

L135: Please write more about Figure 2. It is difficult to interest and draw conclusions from it beyond that MAR data are colder than GC-net station data.

Figure 2 is also discussed in the remainder of this paragraph, we have added an additional reference to the figure on line 143 to clarify.

L140-149: Is this data shown in figures or tables? If so clarify how, if not considering adding a figure to clearly illustrate these findings.

The text on lines 140-149 refers to Figure 2. A citation to the figure has been added on line 143, and more information has been added to the figure to clarify.

L154: Remove the gaps from the total count of data points to give an accurate count of the number of years with data (in table 1).

Done.

L170: Explain "raw Era-interim" and how it is different form "Era-interim"

We use the term raw Era-Interim in order to clearly distinguish from MAR forced by Era-Interim. Have edited to clarify as follows: '…raw Era-Interim output <u>(i.e. not MAR forced with Era-Interim)</u> captures…'.

L180: Provide units for PDD here and elsewhere

PDDs are in ^{o}C, edited throughout to clarify

Figure 1: Show the spatial distribution of the ablation, percolation and accumulation zones

This information has been added to Figure 1 and Figure 5.

Figure 2: This figure needs work. First, a scale is needed to be able to interpret the height and width of the boxes.

Scale added.

Second, it is unclear why the bars are centered in each box.

The bars are centred in each box the better to contrast the difference with the observations.

Third, units on the colorbar are needed.

Units added

Forth, five shades of blue are displayed in the figure, extend the colorbar to capture the darkest blue.

The darkest blue is the bottom out of bounds colour. Apologies; this fell off the submitted version.

Finally, the figure captions says that the figure shows model results, but the colorbar text says it is showing deviations from observations.

Yes, it is the modelled magnitude – the observed magnitude, i.e. the model anomaly with respect to the observations. Caption text edited to be: 'of the modelled magnitude' to clarify.

Figure 3. Clarify the meaning of the black dotted line. Add units to each axis. It would be useful to also show the RMSE in the plots.

Text added to figure caption: 'Black dotted line denotes a 1:1 fit.'
RMSE added as annotation. Caption edited to read: 'Pearson's correlation co-efficient and the root mean squared error (°C) between the data is annotated'

Figure 4. Same comments as Figure 3.

We assume that you mean the same as Figure 2 and have edited accordingly.

Figure 5: Add units to x-axes

Edited as requested

Table 3: Explain the "nyrs" variable. Clarify what spatial domain the data values are calculated for.

Text added to figure caption: 'The number of years of data in total (i.e. the sum of the number of years of data at each station) is also identified (nyrs).'

The spatial domain for each melt zone is taken to be that defined in McMillan et al., 2016, this is specified in lines 112-115 in the 'methods' section.

Table 4: Provide units for PDD

Edited as requested

References

Barrand, N. E., Vaughan, D. G., Steiner, N., Tedesco, M., Munneke, P. K., Van Den Broeke, M. R. & Hosking, J. S. 2013. Trends in Antarctic Peninsula surface melting conditions from observations and regional climate modeling. *Journal of Geophysical Research-Earth Surface,* 118**,** 315-330.

Vaughan, D. G. 2006. Recent trends in melting conditions on the Antarctic Peninsula and their implications for ice-sheet mass balance and sea level. *Arctic Antarctic and Alpine Research,* 38**,** 147-152.