Reply to Anonymous Referee #1

This manuscript reported a dataset of daily and monthly averaged Antarctic air temperature at the resolution of 1km for the 2003-2017 period. This dataset is constructed using MODIS LST and automatic weather station (AWS) air temperature observations by means of machine learning algorithms, which are proved to be effective for reconstructing Antarctic air temperature in the authors' early work published in 2016. The uncertainty of the dataset is also estimated based on a cross validation method. This dataset is an important supplement for the remote sensing based on global dataset of air temperature by Hooker et al. (2018), also useful for estimating Antarctic climate change, and validating the simulation of regional climate models. Generally, this manuscript is well organized and written, and figures are appropriate. It deserve publication in ESSD. However, before publication, the manuscript still requires some following revisions.

Thank you for your positive feedback. We hope that our response will meet your expectation.

Page 3 10-15: What about the quality of AWS observations? Can the data be trusted and is the quality the same at all of the observational stations? What kind of quality control is made for these observations? The availability of observed data is greatly inconsistent in time between weather stations. How large can the errors in reconstructed air temperature be due to this inconsistence?

We used the best available data for this region but certainly maintenance is challenging in such a remote area. However, as of the inconsistencies in time we are not expecting any problems because we assume that relationships between LST, further predictors, and air temperature don't change. So it is not required that we have complete observations for each AWS, as long as the gradients in the predictors are well covered in the data. We will add a section to the manuscript on this topic:

"It is extremely difficult to maintain Antarctic remote stations like these. With the data used here we rely on the utmost professional ability of the programs within the very constraining international level logistics and costs of Antarctic operations. There has been no apparent inter-agency harmonization of sensor types we are aware of and the data quality check of the data used here was limited to the effort of the individual providers. The data are not consistent in terms of their time series. However, due to the machine learning approach that is used for the spatio-temporal model of air temperature, it is not required here that a full time series at a respective location is acquired: The task of the algorithm is to learn the relationships between satellite-based LST as well as environmental properties and air temperature. Since long time series are used, we are confident that the general gradients in the data are covered"

Page 3 20-30: Also please add the description of the quality of MODIS LST. It is also necessary to discuss the suitability of MODIS LST for Antarctic air temperature estimate. See Wang et al. (2013) reported robust correlation between MODIS LST and air temperature over the Lambert glacier drain.

We will add the following description on the MODIS LST quality to the manuscript.

"The data are cloud-masked using the MODIS Cloud Mask algorithm (Ackerman et al., 1998) that applies typical thresholds in the visible and infrared channels. Though the MODIS LST product is cloud masked, the "white on white" and "cold on cold" effect is a challenge for cloud detection in Antarctica (Allen et al., 1990). This holds especially true for cirrus clouds that could in parts not reliably be detected in the used LST product. This is an ongoing challenge and further research effort on this will certainly improve the presented AntAir dataset in the future.
The MODIS LST data are reported with a quality of better than 1°C in the range from -10 to 50°C (Wan et al., 2004). However this did not involve an extensive validation for Antarctica. For the

antarctic McMurdo Dry Valleys, Wan (2014) reported a mean error of 1K. Note that a general bias is not problematic for this study due to the applied machine learning based regression approach, but that robust relationships is relevant. Here, previous studies have indicated robust correlation between MODIS LST and measured air temperature, e.g. Wang et al., 2013 over the Lambert glacier drain or Li et al. (2016) between measured snow surface temperature and MODIS LST."

Page 5 5-10 I think that the authors should use the DEMs with higher accuracy, such as DEMs from Bamber et al. (2009) or Slater et al. (2018), rather than RAMP DEM

The RAMP DEM might have an accuracy not sufficient for e.g. glacier drainage basin delineation for mass balance analyses (Cook et al., 2012), however has been indicated to be suitable as a surface topography dataset (Cook et al., 2012), which is what we need it for.

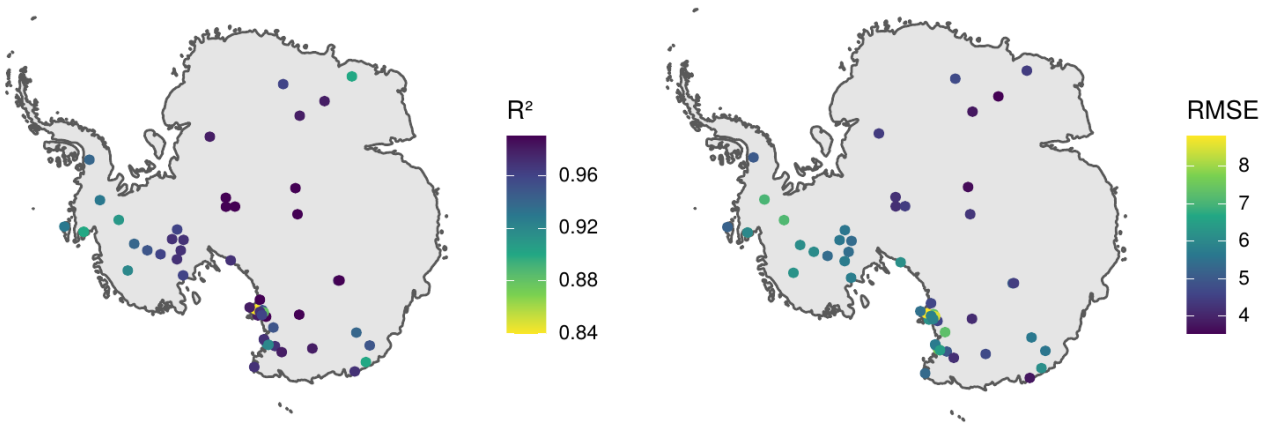
Please note that not all the characteristics of the DEM were used as predictor (see results after variable selection), so the only terrain related relevant information is daily maximum hillshade. We compared the results for daily maximum hillshade for the DEM of Bamber et al. with hillshade derived from RAMP and found that the differences are small (R^2 for the entire study area = 0.98 and with focus on the DryValleys where hillshading is most variable= 0.84). Therefore we are convinced that no change in the results must be expected.

We will justify our choice of the DEM in the manuscript: "*The Radarsat Antarctic Mapping Project (RAMP) digital elevation model (DEM) (Liu et al., 2015), version 2, was used, which has been indicated to be suitable as a surface topography dataset (Cook et al., 2012)*"

Cook, A. J., Murray, T., Luckman, A., Vaughan, D. G., and Barrand, N. E.: A new 100-m Digital Elevation Model of the Antarctic Peninsula derived from ASTER Global DEM: methods and accuracy assessment, Earth Syst. Sci. Data, 4, 129–142, https://doi.org/10.5194/essd-4-129-2012, 2012.

In my opinion, spatial distribution of errors of reconstructed air temperature at each AWS location should be shown.

Agreed, this will add valuable information. We will include the following map to the manuscript, that shows the error/performance for each station.



To further estimate the accuracy, the constructed dataset should be compared with the previous temperature reconstruction for their overlapping period by Steig et al. (2009),

O’Donnell et al. (2011), and Nicolas and Bromwich (2014).

We appreciate this comment and we’re sorry that we cannot fulfill the request for the following reasons: A comparison based on statistical validation metrics as reported in the publications is not possible because of different data, validation strategy etc being used. So only a direct comparison using the data is an option. Unfortunately, for all three data sets that you mentioned only anomalies can be accessed (http://polarmet.osu.edu/datasets/Antarctic_recon/, http://faculty.washington.edu/steig/nature09data/data/), which cannot be compared to our absolute temperature values. If you’re aware of e.g. monthly temperature data available for the mentioned papers, please let us know and we will be happy to perform a direct comparison.