Response to the referees comments on the Earth System Science Data Discussion paper Combining Data from the Distributed GRUAN Site Lauder-Invercargill, New Zealand, to Provide a Site Atmospheric State Best Estimate of Temperature by Tradowsky et al., 2018.

We would like to thank the two reviewers for their useful comments and have included the suggestions into the text and added some figures.
In the following, the comments will be addressed one after another and the answers are indicated in blue.

Reviewer 1:
The authors undertake to address the challenge of how to optimally combine periodic measurements from a distributed GRUAN site to create a continuous estimate of the atmospheric state for a target ECV of interest at the central location. The work is novel and interesting in terms of coming up with a potentially robust approach to addressing the challenge. The problem is also clearly in scope for the journal. As such the work is likely eventually publishable in ESD. However, I have a number of concerns which I would suggest be addressed prior to eventual publication. As far as I can tell, these are all concerns which the authors could address in revisions.
Before getting on to matters of a scientific nature, although well written there were times when I had a very strong sense of déjà vu in that entire passages were often repeated almost verbatim twice or more in the text. I would urge careful proofing to ensure that things are either said only once (obviously in the most appropriate place) or said sufficiently distinctly that the reader is not getting such a sense of déjà vu.

We have worked through the text and taken care that these similar passages where removed or the text was changed.

My biggest concern regards the results section which is presently in my view an inadequately in-depth assessment of the SASBE product quality. Substantial additional analysis is warranted here to justify publication. I would suggest:
1. Combining the current 4 figures into one single figure that covers the four day Sequence.
2. Repeating that analysis for different levels and seasons so that you can satisfy the reader of the verity of the SASBE by showing a range of case studies.

We have combined the figures and presented the results in addition at two more pressure levels (150hPa and 70hPa). Furthermore, we have also added three plots showing the SASBE at the same levels, but in a different season of the year.

3. Further, producing summary statistics of performance via the leave-one out type approach currently employed but for all levels and the entire SASBE series. i.e if you

sequentially leave the Lauder ascents out how well does the SASBE reproduce the launch series given preceding and following Lauder launches and all Invercargill launches? Summary statistics should consider bias and spread based statistics.

Figure 10 shows the yearly mean residuals between Lauder radiosondes and (i) ERA5 diurnal cycle and (ii) a denial study for which the inferred temperature anomalies plus the diurnal cycle are used, but not the Lauder radiosonde temperatures. This figure shows that including the Invercargill radiosonde temperatures into the SASBE improves the fit to the measured temperatures in all years.

4. The comparator of climatology is a necessary but not a sufficient benchmark. Comparison could also be made to the ERA-5 timeseries estimate interpolated to Lauder. As I understand it ECMWF did not, yet, incorporate Lauder ascents into the reanalysis (although that can and should be checked and verified). Thus the ERA-5 analysis would be formally independent of the series of Lauder ascents. As such it would be of great interest to ascertain how the SASBE approach stacks up against a state-of-the-art reanalysis.

While all indications are that Lauder measurements have not be assimilated into ERA5, this can not be ensured at this stage. Those radiosonde sites that are operationally assimilated by ECMWF are used as well as sites that are part of the Integrated Global Radiosonde Archive. Lauder has started submitting data into the Global Telecommunication system in 2016, and ECMWF started to assimilate them operationally in July 2017 (Bruce Ingleby, personal communication, 2018). We are not entirely sure when ERA5 has started or will start to assimilate Lauder ascents.
As of September 2018, Lauder is not part of the IGRA archive as can be seen in https://www1.ncdc.noaa.gov/pub/data/igra/igra2-station-list.txt.
We agree that it would be interesting to compare the SASBE with the ERA5 climatology. However, we decided to leave this out of the paper due to two reasons. (i) as described above, we are not entirely sure if Lauder accents were left out of ERA5, (ii) ERA5 is currently only available from 2008 onwards which would allow the comparison to be made for only a small part of the SASBE timeframe.
The goal of this paper is to present a method and a thorough comparison with ERA5 could take place once ERA5 is available for the entire time frame of the SASBE.

The results section would require a substantial and comprehensive redraft in light of this expanded analysis.

Thank you very much for your helpful suggestions and we updated the results sections which now discusses several more cases and the overall performance of the SASBE.

In regards of the choice to consider the manufacturer processed data, after carefully

checking the GRUAN website I see presently available data streams apparently from both Lauder and Invercargill. These may well be available only for the most recent past meaning that a SASBE processed using GRUAN processed data would be shorter. Nevertheless I think it would greatly benefit the paper were the analysis to be able to be redone using these data-streams even if the resulting SASBE were much shorter. This would reduce the number of assumptions necessarily encoded in the current SASBE product.

While the authors agree that it would be useful to have a SASBE based on the GRUAN data, the main aim of this paper is to present a method. At the time the code was written and the analysis was performed, only a limited amount of GRUAN data was available (the latency for GRUAN data has has been larger than it currently is, due to a change in the ground station software that required updating the GRUAN processing software). Due to the low amount of available data, we decided to use the data prior to the timeframe of GRUAN processed data. However, cognizant of the increasing availability of GRUAN data, we point the readers towards the possibility to use GRUAN data instead. While it would be possible to extend the timeframe of the SASBE now and use the GRUAN data for this, funding for this project has seized and a lack of resources permits the extension of the SASBE to take place at this stage.

This method itself is unaffected by the choice of input data (i.e. vendor provided or GRUAN data). Instead of using assumed uncertainty estimated for the radiosonde, GRUAN estimated uncertainties would be used if the SASBE would be calculated based on the GRUAN data. Once such an update of the dataset takes place, this paper will be referenced.

Temperature given its large spatiotemporal scales is arguably the easiest ECV to perform such an analysis upon. I would suggest at a minimum discussing potential extensibility / challenges in considering other variable.

We have added the following text to discuss the applicability to other ECVS.
*The described method is likely to be applicable to other ECVs that have a large spatio-temporal scale, such as ozone above the boundary layer. However, for ECVs that vary rapidly over distance and time, such as water vapour, the described method is unlikely to provide useful results. Therefore, prior to applying the method, the correlat*ion between the time series of a chosen variable at both locations should be evaluated.

I would suggest being much clearer in Section 3 that the SASBE is being calculated on 16 levels and the justification as to why. Currently, the reader is first made formally aware of this as far as I could tell at the start of Section 4. The fact it is on standard levels and why (for the non-radiosonde experts) should be made much more explicit in building the model through Section 3.

We added the following text in section 3 to clarify:

The SASBE is calculated at 16 vertical levels, i.e. surface, 925hPa, 850hPa, 700hPa, 500hPa, 400hPa, 300hPa, 250hPa, 200hPa, 150hPa, 100hPa, 70hPa, 50hPa, 30hPa, 20hPa, and 10hPa, which are the so-called standard pressure level of radiosondes transmitted in the alphanumeric TEMP format. Currently, the global operational upper-air network is migrating towards high-resolution radiosonde reports in the BUFR format (Ingleby and Edwards, 2015).

Section 5 really constitutes a discussion section rather than conclusions. I would redraft current Section 5 explicitly as a discussion and add a short Section 6 which highlights concluding remarks.

We followed the suggestion of the reviewer.

While it is great to see the SASBE data archived in a long-term sticky archive, given that the primary aim is a tool being developed, arguably the greater value will be in the code. Is the code being archived and made available via e.g. Github? This should be considered and if it is shared how and under what conditions should be documented within the data availability section (would become section 7 if prior comment actioned). If it isn't openly shared then how it can be obtained should be outlined.

The section title was changed to "Code and data availability and the following sentence was added:
The code used to produce the SASBE has been developed in R and is available on request from the first author.

Other minor comments
I assume p.3 lines 4-5 is a stub sentence the authors meant to complete but did not do so. Please edit in revisions accordingly.
In p.5 line 5 I assume that the time difference extends in both directions (prior to and after) yet this is ambiguous as written. I would suggest being explicit here as the alternative explanation of providing information only post-measurement could also have been a logical choice.
In introducing the SASBE method in Section 3.1 I would be more explicit from the outset that the diurnal component is purely climatological in nature.
In p.6 line 7 good -> well

Thank you very much for the comments and we clarified/corrected those.

Reviewer 2:

This paper describes a statistical method for computing a best estimate of the atmospheric state over Lauder, New Zealand, using a combination of observational data sources: weekly sonde launches at Lauder and twice daily sonde launches from Invercargill, New Zealand; and model/data assimilation data from the ERA5 reanalysis.

The method used is the Site Atmospheric Best Estimate (SASBE) technique, here specialized to observational data obtained from spatially disparate locations. The method computes a diurnal climatology for Lauder from the ERA5 reanalysis, and the computed atmospheric profile converges to this climatology in the absence of data. The model uses a regression approach to compute the temperature anomaly at Lauder
with respect to the Invercargill observations. A weighting scheme utilizing computed temporal autocorrelations is used to weight Invercargill- and Lauder-derived observational anomalies with respect to the diurnal cycle. Uncertainties for measurements, the diurnal cycle climatology, and uncertainties in regression parameters are propagated through to the atmospheric state estimate via standard techniques.
This paper is definitely within the scope of the journal, and should be publishable after attention to some issues.
The first significant issue is that a key issue for the SASBE approach investigated here is how useful a site based on spatially distributed measurements is. It seems that in the process of preparing the manuscript, the authors have developed the appropriate quantitative framework for answering this question, but the manuscript as written does not directly address this essential point. In particular, lines 2-5 of p. 14 note that the residuals between the Lauder radiosonde measurements and the diurnal climatology plus the Invercargill regression anomalies and the residuals using the diurnal climatology suggest that there is added value for adding the spatially distant Invercargill measurement. However, the figures that support this analysis are not shown.
It seems that this could be one of the most significant findings from this manuscript: how much are the uncertainties in the atmospheric state estimate improved, as a function of seasonal, atmospheric level, and time of day.

A figure showing the change in residuals when including Invercargill radiosondes has been added to the manuscript for all pressure levels (Fig. 10). Furthermore, more individual cases are discussed in the results section.

Some of the technical aspects of the paper could use better elaboration. For example, it is important to note that in eq. (1) on p.5, the diurnal cycle climatology includes the higher harmonics of the diurnal cycle. These are an important part of the temporal variability of the upper troposphere and stratosphere due to atmospheric tides, and are undersampled by sounders in sun-synchronous orbit. Another example is the Fourier series expansion of the wind direction in eq. (6) on p. 8. What is the purpose of this Fourier expansion? Is it meant to capture some kind of lead/lag behavior between the temperature anomalies at Invercargill and Lauder as a function of wind direction? It was unclear to me what vertical and temporal dependence the regression coefficients have. I assume they are computed independently for each vertical level, but are fixed in time. In general, I think the values of the regression coefficients, and their standard errors, are an important part of the paper.

We have included more detail about the Fourier expansion and their purpose in section 3 of the paper, i.e.:
Inclusion of the higher harmonics of the diurnal cycle, which represent the temporal variability due to atmospheric tides (Haurwitz, 1964), is possible due to the hourly resolution of ERA5. The ability to resolve atmospheric tides is typically limited by the temporal resolution of e.g. satellite-based instruments (Huang et al., 2010).
And:
Term 2 is expanded with a Fourier series for the wind direction to account for the dependence of the correlation between Lauder and Invercargill temperature anomaly on the meteorological conditions, such as a lag/lead in the temperature anomaly based on the wind direction.

Indeed, the regression coefficients are calculated separately for every pressure level and do not depend on the time. This has been clarified in the text.
The number of regression coefficients used in the SASBE is quite large (approx 120) and these can not be used for the analysis of other measurements. Therefore, we believe the regression coefficients will be of little use to the reader and they are not included in this paper.

Revisions that address these issues, and a careful proofreading, should lead to a manuscript that will be appropriate for publication.