Response to Anonymous Referee #1 comments

The authors would like to thank the referee for their detailed comments, which have led to significant improvements in the overall clarity of the paper.

General response to Referee's comments:

The aim of this paper is to establish a methodology, from the limited data available from sites with a high launch frequency, to see if a data correction factor could be established at these sites to guide launch schedules. This represents the first step in developing a general tool for calculating temporal correction factors for any ground based monitoring site.

The aim is not to physically explain or quantify the reasons behind the correction factors derived at each site.

The above has been added to the introduction to clarify the scope of the work contained in the paper.

The following text has been added to the conclusions section:

"Having established that the method presented in this paper is a viable one for estimating temporal variability it should be recognised that these results only directly apply to the radiosonde launch sites from which the datasets have been obtained. In order to generate appropriate correction factors for other sites the method will require further development, using additional data sources or model results for each site."

1) Referee's comment:

The first major comment is on the definition and discussions on "correction factor". First, there are two places where the authors used the term "correction factor", one is 8344 Line 1, and the other one in 8345 Line 15. The first time, the term is used as "The correction factor <between> the base set and the mean hourly rate of change in temperature for a single launch (i.e. no correction), 2 launches a day and 4 launches a day were calculated" (8344 Line1-3). The term of "correction factor" is misleading, because it seems to describe the "difference" in temperature changing rate (k/hr) between 3 hr launch schedule and 6 h, 12 h schedules (results shown in Fig 3). And the legend in Figure 3 says "4 launches/day, 2 launches/day, no correction". What is "no correction", and what is the ones being "corrected"? The second time, the authors defined the term "temperature correction factor" as "temperature changing rate", which is not the "difference" in temperature changing rate. Because of these confusions, the reviewer suggests the authors always use the term "temperature changing rate" instead of "correction factor", because the latter one is usually used to describe bias corrections on measurements, which is not the case here. This will also revise the term used in the title.

Author's response:

The authors agree that there can be confusion over the definition and application of the "correction factor". Where relevant "correction factor" and "change in temperature" has been replaced by "temperature change rate".

No change in article title has been made as data in Table 2 can be used as a correction factor for measurements taken at different times.

2) Referee's comment:

In addition, there are two analyses (Figure 3, and Figure 4, 5) that the authors did not explain exactly how the values are calculated in these figures. The reviewer suggests that the authors clarify the calculation for the values used in Figure 3, 4, 5 in both text and equation forms. For example, if using dT/dt_3h, dT/dt_6h, dT/dt_12h to define the temperature changing rates, how is “mean hourly rate of change in temperature” calculated? Is it the mean or absolute mean of the values in Figure 2? Does the “differences in correction factor (temperature changing rate)” in Figure 3 represent the difference of the mean of the absolute temperature changing rate, or the difference of the mean of the temperature changing rate, or mean of the difference of temperature changing rates? Also the difference is which dataset minus which one? The authors should explain which of the following is the way that “correction factor difference” was calculated (or other ways):

mean(abs(dT/dt_3h_i) – abs(dT/dt_6h_j)), i = 0 to 7; j = 0 to 3

or abs(mean(abs(dT/dt_3h_i)) – mean(abs(dT/dt_6h_j))), i = 0 to 7; j = 0 to 3

or abs(mean(dT/dt_3h_i) – mean(dT/dt_6h_j)), i = 0 to 7; j = 0 to 3

The term of “hourly rate change in temperature” in Figure 4 and 5 needs to be clarified in two steps. First as the question for Figure 3, how is “hourly rate change in temperature” calculated? Second, when averaged into different altitudes, how are the values averaged, and what are the bins of altitudes? Until the authors explain how exactly these terms are calculated, the discussions of Figure 3 and later on 6 h dataset for Figure 4 and 5 are not clear.

Author’s response:

We agree that the description of the calculations would be clarified by adding the equations used and additional text.

The following additional descriptive text added to Section 3 along with equations to clarify calculation process:

Radiosonde temperature readings are amalgamated into altitude bins 500m high, labelled as the centre of each bin, i.e. 0 to 500m labelled as 250m. The temperatures in each altitude bin are averaged to provide a mean temperature, *T*, for that specific altitude. The rate of change in temperature between single launches 3 hours, 6 hours and 12 hours apart, at each altitude, were calculated according to Eqn 1. The mean rate of change in temperature between each launch separation and altitude, $\frac{dT}{dt_n}$, were then calculated according to Eqn 2.

$$\frac{dT}{dt_n} = \frac{T_n - T_0}{t_n - t_0} \qquad \text{Eqn 1}$$

Where *n* = 3, 6, or 12 hours separation between launch time.

$$\overline{\frac{dT}{dt_n}} = \frac{\sum \frac{dT}{dt_n}}{i} \qquad \text{Eqn 2}$$

Where *i* = the number of launch pairs

The mean rates of change in temperature ($\overline{\frac{dT}{dt_3}}$, $\overline{\frac{dT}{dt_6}}$ and $\overline{\frac{dT}{dt_{12}}}$) were used to define temperature change profiles over the day at different altitudes and are shown in Figure 2.

In order to quantify the difference between the different launch schedules it was assumed that 8 launches per day provided the best available measure of the changing state of the atmosphere. The mean hourly rates of change in temperature from these launches were therefore considered to be the base set. The difference in temperature change rates, $\Delta\frac{\overline{dT}}{dt_n}$, (K / hour) between the base set and a single launch, 2 launches a day and 4 launches a day were calculated according to Eqn 3. The results of which can be seen in Figure 3.

$$\Delta\frac{\overline{dT}}{dt_n} = \frac{\sum ABS\left(\frac{\overline{dT}}{dt_n} - \frac{\overline{dT}}{dt_3}\right)}{8} \qquad \text{Eqn 3}$$

Where n = 6 or 12. For single launches, $\Delta\frac{\overline{dT}}{dt_n}$ was taken as the mean of $\frac{\overline{dT}}{dt_3}$.

3) Referee's comment

Similarly, the term of "hourly rate of change of temperature" in Figure 6 and Table 2 should also be clarified. In fact, if this term stands for mean of the dT/dt_6h, then this mean value has averaged out the positive/negative dT/dt_6h values throughout 24 hours, and the final mean value (such as 0.081 K in the given example), does not represent the real dT/dt_6h at a given time of a day. Especially as shown in Figure 2, some parts of the day have dT/dt_6h as negative values, some parts are positive. The reviewer thinks that if the authors want to use the dT/dt_6h as a method to estimate temperature differences for dataset intercomparisons with time mismatch, this temperature difference factor should be given as different values for different times of the day.

Author's response:

The data for Table 2 is averaged over the whole year and as the referee points out the negative changes will cancel out the positive changes. The result in Table 2 gives the net positive or negative temperature change, for a specific site between 2 set launch times. This is a subset of the complete data set. The supplementary data has been expanded to give the data for all 3 data sets covering the separate 6 hour launch separations.

4) Referee's comment:

The second main comment is that when comparing temperature changing rate (difference between dT/dt_3h and dT/dt_6h), the statistical significant test should be done, such as showing whether the dT/dt of different launch schedules are within their 95% confidence interval. This is necessary if the authors later on use 4 times launch as adequate dataset: "Once a 6 h launch frequency was accepted to adequately represent the rate of change in temperature..:" (Line 13-14 8344).

Author's response:

The authors agree that showing the statistical significance will improve the data analysis. Statistical significance calculations have been performed and the following texted added to Section 3.1:

"In Figure 3, the 4 launches per day data set is statistically different from the single launch data set at all altitudes except 3250m, with a confidence level of 1σ (68%). At the 2σ (95%) level, 3 altitudes (9250, 12250 & 15250m) are statistically different."

5) Referee's comment:

In addition, it should be discussed if the 4 times launch schedule dT/dt is sufficiently showing temperature changing rate for different sites, seasons, and altitudes. Currently, the discussion of Figure 2 and 3, the correction factor differences between 2, 4, 8 times, are based on the dataset from Manus Island. The authors did not clarify if Fig 2 and 3 are based on Manus Island dataset for all seasons. If they are for all seasons, then it is unclear and also how this dT/dt comparison would change with seasons, and geographical locations. Especially since Figure 4 and 5 show that the temperature changing rate changes with season and sites.

Author's response:

The data for Figures 2 and 3 are for the complete 6 month dataset and have not been split into seasons. The Dynamo campaign is the only data set with 8 launches per day so analysis at other locations is not possible. The paper provides a disclaimer to cover this: "Clearly this result only directly applies to the Manus dataset, but it provides reasonable confidence in the use of 4 launches per day data for longer term analysis."

6) Referee's comment:

The third question is on the radiosonde instrument uncertainty and precision. The authors need to provide references and previous validations/lab calibration results on the radiosonde temperature measurements. In particular, the radiosonde temperature measurement is shown to have differences in "T error characteristics" for daytime and night time, as mentioned in Sun et al. 2010 (JGR). In addition, Sun et al. 2010 also suggested that the radiosonde day-minus-night T differences increase with height: "For most radiosonde types, daytime mean(delta_T) tends to be larger than night time for the upper troposphere and stratosphere... Average (over all types) day-minus-night mean(delta_T) increases from about zero at 350 hPa, to 0.10 K at 50 hPa, to 0.20– 0.30 K at 10–20 hPa." Then this leads to the question, if there are different biases in T between day and night, the temperature changing rate would have unevenly distributed biases throughout the 24 hr of the day as well. Thus the authors need to take this part into consideration and discuss the influences on the results.

Author's response:

The data analysis has been performed on publically available data sets. As noted by the referee and commented on in Section 3.2, changes in radiosonde type and post processing routines have not been taken into account in this work, but could be addressed in future work.

7) Referee's comment:

Fourth, for the discussion on how many measurements are needed to have temperature uncertainty within a certain range. Here the authors only used the dataset at altitude of 5km, in spring, and in one time range 13:00-19:00 hr local time. But it is unclear if the minimum number of measurements needed would vary with altitudes, seasons, and times of the day. The authors should provide more information on these factors and show if 10 measurements are sufficient for all these conditions. In addition, it is unclear how the authors chose 1, 10, 100 measurements, like randomly or other ways? Please explain how the standard deviation is calculated for 1 measurement?

Author's response:

The standard deviation for a single measurement corresponds to the standard deviation of all the temperature change rates calculated for that specific launch time, altitude and season. It is therefore the random noise at 1σ for a single measurement.

The supplementary data in support of Table 2 gives the mean rate of change, standard deviation of 1 measurement and the standard error of the mean of 10 and 100 measurements for the 2 Lindenberg and Southern Great Plains Data sets at 4 altitudes (5, 10, 15 & 20km) for each season for all 6 hour launch separations.

The 1, 10 and 100 measurements were chosen to give an indication of the reduction in uncertainty due to repeat measurements, it is a simple reduction by root n of the random noise associated with a single measurement. Figure 6 gives a good visual indication of the uncertainty compared with the magnitude of the correction factor.

8) Referee's comment:

General comments on all figures, the datasets used for each figure should be clarified (at least in the figure caption). That includes, whether the datasets are for all seasons, all times of the day, all altitudes, given that some of the information cannot be guessed from the figure itself. Also for Figure 2, 3, 4, 5 and Supplementary Figure 3 and 4, the number of measurements need to be provided. These numbers are not provided in the supplementary material either.

Author's response:

The authors do not agree that recording the number of launches in each Figure caption would aid clarity, however text recording the number of launches per data point has been added to the text discussing figure 6.

Additional text:

The standard errors of the means for 100 measurements in Table 3 are similar to those for the Manus Island results in Figure 1 (0.038), which were typically made up of 970 launches per result. The number of launches per data points for the Lindenberg 1999 to 2008 data set is 889 and for the Lindenberg 2009 to 2012 data set is 227.

Detailed comments

9) Referee's comment

8340 Abstract. Line 1-4. This first sentence in abstract needs rewording. Suggest: "Radiosondes provide one of the primary sources of upper atmosphere temperature data for numerical weather prediction, the assessment of long-term trends in atmospheric temperature, the study <add: of> atmospheric processes and [delete: provide a source of] <add: the> intercomparison [delete: data for] <add: with> other temperature sensors e.g. satellites."

Author's response: Implemented

10) Referee's comment:

8340 Line 13: "analysed" should be "analyzed".

Author's response: Not implemented, UK English has been used throughout the text.

11) Referee's comment:

8340 Line 14. The sentence needs rewording. Suggest: This provides the appropriate estimation of temperature differences for given temporal separation and the uncertainty associated with them.

Author's response: Implemented

12) Referee's comment:

8340 Line 17. What is "standard uncertainty"? Do the authors mean standard deviation, or uncertainty?

Author's response: Replaced uncertainty with standard error of the mean

13) Author's comment:

8340 Line 20-23. Please put a citation for: "…used globally as input data for numerical weather prediction {ref}." Also please put citations for "Radiosonde data can also be used to assess long-term trends in atmospheric temperature {add ref}, study atmospheric processes {ref} and provide a source of intercomparison data for other temperature sensors e.g. satellites {ref}."

Author's response:

Not implemented, these are well accepted uses of radiosonde data and we consider that specific sources do not need to be referenced.

14) Author's comment:

8341 Line 4. Please explain "major void": "…designed to meet climate requirements and to fill a major void in the current global observing system (Thorne, 2013)."

Author's response: "Major void" replaced by "large gap".

15) Referee's comment:

8341 Line 26-28. Need to add some details, what satellite was used, and the 0.15 K difference for what pressure level? In addition, the authors need to point out what is the new contribution of this work compared with the Sun 2010 work, and other previous work on temperature variabilities in time/space.

Author's response:

The authors agree that additional details of the comparison with satellite data are required. The following additional text has been added to Introduction:

"Previous work [Sun 2010] has found that the mean temperature difference (all altitudes) between across 13 types of radiosonde and the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) Global Positioning System, Remote Occultation (GPS RO) satellite measurements for a global network to be 0.15 K. For Vaisala radiosondes, whose data was analysed in this paper, an increasing warm bias (from 0 – 0.4K) with altitudes above 19km (50 hPa) was found. The effect of the difference in radiosonde launch time and satellite overpass was also examined. The comparison standard deviation errors (all radiosonde types) for temperature were found to be 0.35 K per 3 hours difference."

16) Author's comment:

8342 Line 18. Please spell the full name of "DYNAMO".

Author's response Implemented.

“To determine the frequency of launches needed to have an acceptable understanding of the atmosphere’s temperature stability over short-time periods (<24 hours), temperature measurements from radiosonde flights made by the upper-air sounding network for Dynamics of the Madden–Julian Oscillation (DYNAMO) at Manus Island, Papua New Guinea were analysed.”

17) Referee’s comment:

8343 Line 19. To statistically demonstrate the 3h and 6 h launches have “similar” profiles, besides “within the error bars (standard error of the mean)”, the authors should show a statistical significant test for their differences.

Author’s response:

Statistical differences between launch separations has been added to this section, see point (4) above.

18) Referee’s comment:

8343 Line 12-14: For Fig 2, the time stamp, Local time (hr) should be mentioned in the figure x-axis legend, and also should be mentioned when describing the Fig 2 in Line 12-14. Otherwise there is no unit for “Time” and readers could assume either UTC or local time by just reading the figure.

Author’s response: Implemented in figure 2 and in supplementary data.

19) Referee’s comment:

8344 Line 25. Figure 4 should be Figure 5 in the text.

Author’s response: Implemented.

20) Referee’s comment:

8345 Line 1-5. The discussion on the differences between sites at different seasons. The authors should explain if the significant differences for different seasons/sites are due to different reasons. For example, the summer time differences between Southern Great Plain and Lindenberg, are they due to the differences in their climate, typical weather condition? And for the comments on the differences in Lindenberg data as a result of changes in radiosonde type and analyses procedures, the authors should add references to these documented changes in instrumentations, and maybe previous work on the calibration / intercomparisons of different instrument of radiosonde.

Author’s response:

The aim of the paper is to establish a method to correct for temporal mismatch. As noted by the referee we have not considered the effect of change in radiosonde type on the two Lindenberg data sets, however each set of data was taken using a single radiosonde type so should not affect the specific correction factor. The analysis of the effect of radiosonde change on the datasets could be the subject of future work.

The following summaries of atmospheric near surface temperatures (0 to 500m) for the three radiosonde launch sites have been added to Table 1.

Lindenberg 1999 to 2012, median temperature: 9.4°C, range of daily temperatures: -17.6°C to 30.2°C.

SGP 2006 to 2012, median temperature: 17.1°C, range of daily temperatures: -14.3°C to 40.9°C.

Manus 24/09/2011 to 31/03/2012, median temperature: 26.4°C, range of daily temperatures: 15.9°C to 34.3°C.

21) Referee's comment:

8345 Line 14. "10 and 100 repeat measurement", suggest use "repeated measurement". Please revise.

Author's response: Implemented.

22) Referee's comment:

8345 Line 20. "Figure 6... in Spring and for 13:00 and 19:00 local time…".

Author's response: Implemented.

23) Referee's comment:

8345 Line 23. Please explain why use temperature <= 0.1 K/hr. This might be derived from Sun et al. 2010, where temperature is found to have 0.35 K/3 hr changing rate. Yet, besides the work of Sun et al. 2010, which acknowledged that the radiosonde biases in T change magnitude/sign with day/night, altitudes, sites/instrumentations, the authors should also add citations of other datasets, for example, higher resolution dataset of temperature variability in time, for the definition of "acceptable bounds" of temperature changing rate.

Author's response:

The authors consider that a standard error in the mean of 0.1 K/hour is an acceptable level of uncertainty for this measurement and therefore show that 10 or more repeated measurements would need to be taken to achieve this level of uncertainty. The method could be applied to other uncertainty criteria if required. We therefore feel that the text does not need a qualifying external reference.

24) Referee's comment:

8345 Line 24. Missing Table 3.

Author's response: Change reference from Table 3 to Table 2.

25) Referee's comment:

8345 Line 25. "…similar to those…in Fig. 1 (0.038)." It is unclear what 0.038 value the author meant in Figure 1, please explain where this value comes from.

Author's response: Reference to Figure 1 changed to Figure 2.

26) Referee's comment:

8346 Line 1. "Correction of 0.081 K should be applied, ." Please explain in text or equation how this 0.081 K is applied, like plus or minus? In fact, as mentioned in the general comments, some parts of the day have positive dT/dt, and the other parts have negative dT/dt. The "correction" should be given for different times of the day.

Author's response:

The authors agree that the application of the correction factor should be given. The following text has been added:

"This correction should be subtracted from the radiosonde measurement to adjust for the temporal mismatch."

Tables 2 a, b & c in the supplementary data have been expanded to give the summary data for all three data sets covering each 6 hour launch separation. The text in Line 1 related to the SGP data set which is not shown in the example in Table 2. in the main body of the paper, this has been changed to give the results for Lindenberg. Line 1 changed to a correction factor of 0.036K with an additional random uncertainty of 0.265K…

27) Referee's comment:

8346 Line 4-5. "The overall results presented here enable such an evaluation to be made for any altitude, time of day and season." The reviewer does not agree with this sentence, because the results presented here only represent 13:00-19:00 hr local time, and only in spring, and the authors haven't shown any analyses whether the factors such as season and time of the day would matter to the "correction" (such as 0.081K) and uncertainty (0.316K). If the authors address the question for 8346 line 1, that is, how the correction is applied with 0.081K, then maybe this method can be applied to specific location/time/season, but the results presented here are not covering all conditions.

Author's response:

The data is presented in the supplementary data, the following sentence has been added to cover this:

"The supplementary data gives a summary of results for all three data sets over the separate 6 hour launch separations."

28) Referee's comment:

8346 Line 16-18. Rewording. "…provide appropriate estimation of temperature differences for a given temperature separation and the uncertainties associated with them."

Author's response: Implemented.

29) Referee's comment:

Table 2. Altitude range needs to be clarified. _5km, is it 0-5km, 5+/- 5km? First row, "Mean rate of change K/hour", is this the mean of (dT/dt), or absolute mean? As mentioned in the general comments, if it is the mean, then positive and negative values cancel out; if it is the absolute mean,

then which direction should the dataset be adjusted? Similarly for Figure 6 caption, please define “mean rate of change of temperature”.

Author’s response:

The authors agree that the altitude range needs to be clarified.

The additional text in Sections 2 & 3 (Point 2) describes how temperature data is averaged into 500m bins labelled as the centre of each bin, i.e. 0 to 500m labelled as 250m. The Figure 6 caption has been modified as in point (8) above.

Supplementary material

30) Referee’s comment:

Complete Figure 3 and 4. Number of measurements should be provided.

Author’s response:

Document title “Complete set of charts for Figures 3 and 4” is mislabelled and should read Figures 4 & 5. The number of launches per data point has been added to introduction.

“The number of launches per data point:

Lindenberg 1999 to 2008	889
Lindenberg 2009 to 2012	227
Southern Great Plains	572”

31) Referee’s comment:

Table 2a and c, the altitude “= 5 km” is not consistent with “_ 5km” in the Table 2 of the manuscript, please clarify if 0-5 km, 5+/-5km or otherwise.

Author’s response:

the approximately sign “~” has been removed from Table 2 altitudes. The data included in each altitude bin are now defined in the main body text, see point 2.

32) Referee’s comment:

In the excel sheet for “Data for Figure 2”, the date “1/1/1900” at the bottom of each data column is wrong should be corrected or removed.

Referee’s response:

This data drives the x-axis for the charts and additional labelling has been added accordingly. The display format has been changed to “hh:mm” for consistency.