Reply to Referee 1

My only major concern is on the comparison with sodium lidar climatology. While States and Gardner (2000) reported both climatologies based on daily mean and nighttime only measurements, Yuan et al., 2008 only listed diurnal mean results. So, the comparison between the Rayleigh nighttime measurements with CSU diurnal mean measurements are questionable.

You are correct that Yuan et. al. 2018 in their publication only listed the diurnal mean results. For this work, we contacted Dr. Yuan who gave us night time only results, which is what we used in this study. We mentioned this in the manuscript on page 18 (lines 2, 15) and we thank Dr. Yuan for providing us the nighttime climatology in the Acknowledgments.

All other corrections you suggested are done, and we thank you for your suggestions.

Reply to Referee 2

General comments:

The authors stress a lot on the extension of the altitude range and the comparison with sodium lidars. I have the feeling that this is somewhat biased towards the OEM. Regarding the altitude coverage, the H&C method often provided results up to 95 km, but is limited to 90 km for the present study (p. 7, l. 31). On the other hand, with the OEM the 90% limit is typically reached around 100 km. Therefore some extend of altitude coverage is acknowledged, and it may depend on situation whether it is 5 km or 10 km. The comparison with sodium lidars revealed differences that are slightly smaller for the OEM than for the HC data set. Nevertheless, the improvement is small and the remaining, larger differences between OEM and Na as well as within the Na data sets need to be examined first.

We agree with everything you say here concerning the size of the improvement, but we feel that science improves in small increments like those demonstrated in this manuscript. In the context of Rayleigh lidar and what improvement OEM brings to the community, consider that to get the same signal-to-noise ratio at the same temporal spatial resolution 7 km higher in the atmosphere (e.g. 1 scale height) requires a 2.7 times increase in the power-aperture product of a system (due to decreasing density with altitude). Consider a large power-aperture lidar like the Purple Crow system. For most of the measurements included in this study it used a 12 W laser and a 2.6 m diameter mirror. To get this factor of performance increase we would have had to use a 33 W laser or a 4.3 m diameter mirror. Current commercial lasers at 532 nm are about $120,000 CDN or more, and top out around 30 W. Developing a liquid mirror this size is feasible but would be challenging, and a traditional mirror this size is unaffordable for a university research group. With the OEM method described here and now shown to be useful for a large, heterogeneous data set, we offer users the ability to take the measurements they have already made and extend them up a scale height or further, basically for free. Small gains for sure, but significant when applied to networks of lidars such as in the Network for Detection of Atmospheric Climate Change (NDACC) lidars.

Figures 12-14 are only quickly introduced but hardly discussed. Maybe one should be kept, but the other seem to be redundant and mainly stress the difficulties of the comparison (valid for both Rayleigh methods). With respect to this, I wonder whether solar cycle effects on the temperature above 80 km are acknowledged. Four reasons for differences between the data sets are given, with the most important one unfortunately only in the last sentence of the fourth item: Large scale waves and natural variability. To improve the comparison in this altitude range, I suggest to try to do “closer” comparisons, e.g. with satellite data, instead of using data sets that only partly overlap in time and are separated in space. In order to have a more concise but better focused paper, I suggest deleting Fig. 3 and 4, as well as two out of Fig. 12-14. Instead, the authors should provide scatter plots with direct OEM-HC comparisons (or other appropriate figures) as well as direct comparisons with other instruments above 90 km.

We don't see this exercise as particularly useful. Figure 1 shows such a scatter plot. The two methods are in excellent agreement, and this can be verified on a case by case basis. We should also clarify what is being validated in this manuscript. We are validating the OEM method relative to the traditional HC technique, used in the community for almost 40 years. As far as comparisons with other instruments, our goal is not the same as a measurement validation. We most definitely agree that validation studies of upper air temperatures are extremely important, and hope that our OEM retrieval will be of help in this area. Validation is something we and other groups are working on, but measurement validation is not the goal of this study.

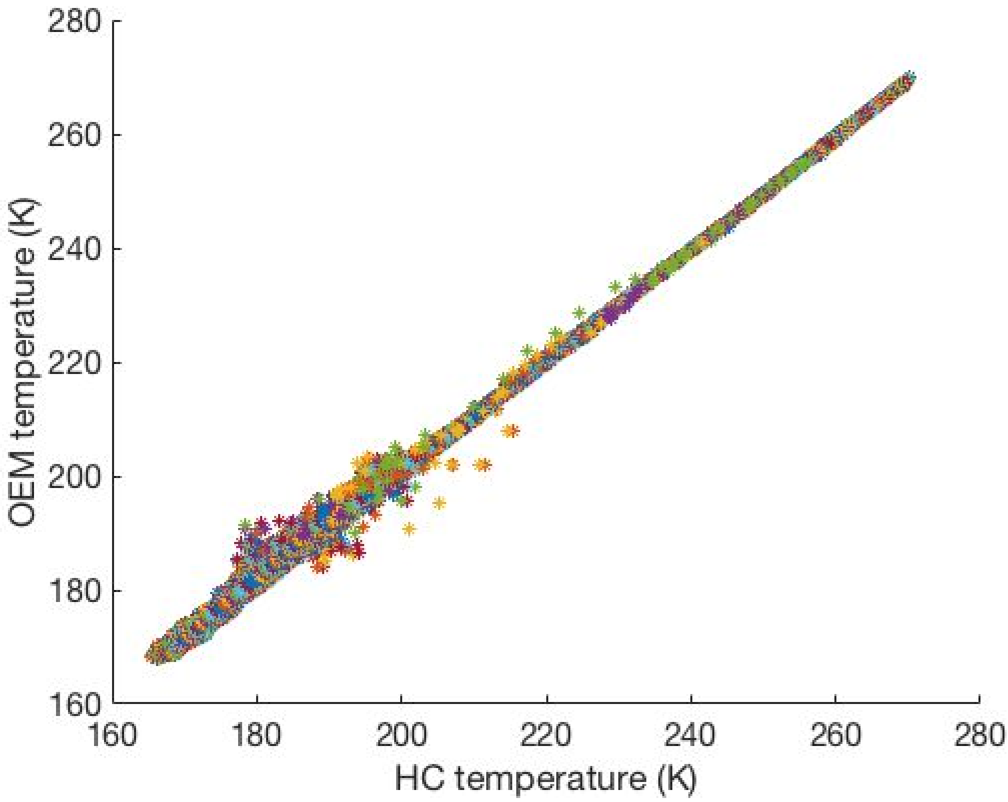

Figure 1. HC versus OEM temperature for all nights in the climatology fit to a line constrained to pass through zero. The average slope is 1.0013 ± 0.0017, meaning the match between the HC method and the OEM is excellent. We are currently extending these comparisons to individual nights using coincident data sets with multiple lidars and satellite-based temperature profile measurements.

Specific comments (l. x/y means page x, line y): l. 2/14-17: I suggest to write “resonance lidars”. Beside sodium lidars there are also powerful temperature lidars using iron or potassium.

Agree, good point. Done.

l. 2/21-3/3: Here is a long section introducing different climatologies and the seasonal variation of temperatures in the MLT. This erroneously draws the attention of the reader to geophysical problems rather than the intended methodic comparison. If different mid-latitude lidar temperature data sets shall be introduced, I suggest adding Gerding et al., ACP, 2008, because there Rayleigh and resonance data are combined, overcoming the limitations from unknown seed temperatures.

Thanks for reminding us of the Gerding et al work, which is an important study offering an alternative way to overcome the seed pressure uncertainty. For Rayleigh lidar systems with modest power-aperture product co-located with resonance temperature lidars, this method is the preferred option to using the Rayleigh technique only. However, for a larger power-aperture product lidar like the PCL (which no longer has a co-located resonance lidar) this method would limit the greatest heights obtainable by the system, as demonstrated in this study. We have added a description of the Gerding et al. work on p 18/l 5 of the revised manuscript in the discussion of seed pressure uncertainty.

l. 2/31: I assume that the two beam technique is mainly for wind measurements rather than for daylight capability.

Correct, it is used for measuring wind as well as measuring during daytime.

l. 7/31: I do not understand how (later) data can be compared to Na data above 90 km, if the HC method is limited here to altitudes below 90 km. On the other hand, also Fig. 10 and 11 show data above 90 km. Maybe I do not understand the method here.

In order to calculate the temperature from Rayleigh lidar measurements it is necessary to have an initial seed (“tie-on”) pressure at the highest altitude of the measurement to initialize the downward integration. The altitude for the seed pressure is chosen based on signal-to-noise ratio (SNR). We have found an SNR=2 is a good choice for the seed altitude in the HC method (in the OEM method we seed at around 120 km altitude). However, the altitude on a given night corresponding to an SNR = 2 can change greatly due to sky conditions or system performance. In AS2007 we decided to required that the maximum height with SNR = 2 be above 95 km altitude. For the current study we wanted more nights available for comparison, and thus took the minimum height to be 90 km altitude, which allowed 20% more nights to be included.

l. 9/19: I do not see any white spots in Fig. 4. They seemed to be filled blue (0-3 K) now. Please clarify.

You are correct, there is no white color in the colorbar and the color between 0-3 K is purple. This sentence is now removed from the paper.

l. 10/3: I do not see the maximum between 80 and 90 km. The largest variability in this period is observed above 100 km, which maybe should be ignored (but the 80/90% lines are missing). Even then, the variability is generally increasing with altitude (with a lot of temporal variability). You should further repeat which type of waves you expect to see in this plot. But as written above, this figure is not essential for this paper.

We have amended the paper as you suggested, on 9/17 changing the sentence to:

> 2 to 33 days, *encompassing the scale-range of planetary waves*.

l. 16/15: Maybe I missed the point, but please explain why the LLR channel is not used for the HC method.

Since the purpose of this paper was method validation, the HC method calculations were only done with HLR channel, consistent with AS2007. However, the method simultaneously uses both LLR and HLR measurements to retrieve a single temperature profile from the complete range of measurements. We discuss this in the appear on page 16, line 20.

l. 16/22: Please explain how you derived the +/- 0.05 K difference and why merging the data will provide in a larger error.

We calculated the average from all temperatures between 40 and 60 km for all months and we calculated their standard deviation, which is ± 0.05 K. The merging uncertainty is the topic of Jalali's MSc thesis and discussed in detail there (Jalali, 2014).

l. 16/26: Why is HLR data lacking in the winter months?
The PCL operates needs relatively clear nights for operations. The region of southern Ontario where the PCL is located is typically cloudy due to its location between Lakes Ontario and Erie. It can also be difficult to access the facility in the winter, particularly in its original location at the Delaware Observatory.

l. 17/14 and l. 22/26: Please explain why Na lidars are most sensitive between 90 and 100 km. I would expect their best performance around the maximum of the Na layer at 90 km (i.e. at 85-95 km).

We have revised this sentence as follows:

Using the OEM to extend the PCL Rayleigh lidar temperature climatology to above 100\,km provides the opportunity to validate the PCL results against sodium lidar climatologies, which have their highest signal-to-noise ratio in a few kilometer wide region between about 90 to 95\,km altitude, and obtain sufficient high quality measurements to calculate climatologies from 85 to 105\,km.

l. 18/14-15: I am not sure that I understand this sentence. Do you mean the >10 K difference above 100 km? a) This occurs also in summer, b) this is above the 90% line, c) it is even smaller in Fig. 12 (URB). Reference to AS2007 is confusing, because there the winter mesopause is explicitly not covered ("The winter mesopause is above the top altitude of the PCL climatology").

We are referring to the temperature difference at 90 km (Figure 13, CSU climatology) which is around 15 K, whose value drops to around 10 K (Figure 14) in the upgraded CSU climatology.

Section 4.2: I see the whole comparison with the Na climatologies problematic. If I understand correctly, the Na data are defined as "truth" and a smaller deviation of the OEM data compared to the HC data suggests the OEM data being better. To some extend this is OK, even if the larger differences between the Na data sets limit the validity of this approach. Any closer discussion of differences is – from my point of view – problematic, because the natural variability of all data sets needs to be acknowledged first. Day-night differences in Yuan 2008; covered years disagree in all data sets (but overlap); differences due to incomplete sampling, … On the other hand the authors mainly aim for the comparison of the RMR retrievals, i.e. the problems of comparing climatologies are outside of their scope and the reader gets side-tracked.

If you assume (as we do) the best measurements for comparison we have in London, Ontario in the upper mesosphere and lower thermosphere are sodium lidar temperatures, then we want our climatology to be in agreement with these measurements as closely as the sodium lidar measurements agree to each other (with the tacit assumption that these differences are primarily geophysical). Our OEM climatology passes this crude, but reasonable, metric for agreement.

l. 19/4-5: I do not see that the AS2007 comparison is worse than the OEM-Na comparison shown in Fig. 12-14. Please explain!

For clarity and showing the difference more clearly we listed the temperature differences inside Table 4 since it is difficult to tell these differences from the plots. The numbers themselves show overall an improvement when OEM is used.

l. 22/22: I do not agree that OEM-Na is within Na-Na differences. Often it is larger. Sometimes OEM-uCSU is smaller, but the daytime data of uCSU has an unknown effect.

We do not use daytime temperatures here, as discussed on page 18 (lines 2, 15). We agree with you that there is variability, but overall the Rayleigh-scatter temperatures, with the seed pressure quantified using the OEM, are consistent with the Na lidar temperatures. Our next step, as you suggest, is to perform validations of the temperatures against other available measurements such as satellite-borne instruments, and when averaging kernels are available from these instruments use the lidar averaging kernels for these comparisons.

Thank you for your thorough and thoughtful readings of our manuscript, and for your comments and suggestions to improve it.

Reply to Referee 3

All of the minor corrections are done as you suggested, thank you.

Response to the Editor: amt-2018-117 (Jalali et al.)
24 October 2018

We have made the following changes to our manuscript.

1. The sentence (p17, l6) "The differences between the OEM and HLR are not calculated below 40 km due to the lack of HLR data in winter months." was changed to:

> The differences between OEM and HLR are not calculated below 40 km due to the lack of HLR data in some time periods.

2. We have changed the confusing sentence about the sensitivity of the sodium temperature measurements with altitude on (p18, l11) and in the response to Referee 2 to:

> Using the OEM to extend the PCL Rayleigh lidar temperature climatology to above 100 km provides the opportunity to validate the PCL results against sodium lidar climatologies, which have their highest signal-to-noise ratio in a few kilometer wide region between about 90 to 95 km altitude, and obtain sufficient high quality measurements to calculate climatologies from 85 to 105 km.

3. The x-axis titles in Fig. 10 and 11 have been changed from "Temperature" to "Temperature difference." In Figs. 10, 11, 12, 13, and 14 the captions have been changed as suggested to show the formula for the difference used.