Reply to Referee No. 1
(referee comments are red; author replies black; new manuscript text blue)

1. I am confused by the statement on the value of Au 0.9 and 0.8, such as "... the height above which the temperature profile is more than 90% (0.9) and 80% (0.8) due to the measurements", which appears in several figure captions. This seems to contradict to the statement on "a priori" on page 8 line 14-16, "the choice of a priori has little effect" below the cut off height "on retrieved temperature. . .". I can understand the latter statement, but the former one, seems to me, would be other way around. So, I would appreciate if the author could clarify either one of the statement in the revision.

Thanks for catching that, you are correct. There is a typo in that sentence and we have changed "above" to "below".

2. The author argues the large "geophysical variability" in January, peaked at 41 km, is related to SSW (page 9 line 18-19). However January has the least amount of measurements based on Table 1. Could the lack of measurement contribute to this large "geophysical variability"? In addition, SSW would have impact all the way up to the middle latitude upper mesosphere (Yuan et al., 2012), but the variability is small between 70 km and 80 km in Figure 4 and the author contributes the large variability between 60 km and 70 km to MILs. So, I am not convinced the "geophysical" nature of these large variability as the author states.

The variability in January is most probably to some extent related to a lack of measurements in this month. Therefore, we would change the sentence:

There is a peak at 41 km in January which is related to sudden stratospheric warmings during winter (Argall and Sica, 2007).
to
There is a peak at 41 km in January which may be related to sudden stratospheric warmings during winter. However, the lower number of measurements in January will also contribute to the variability, and determining the extent of each contribution is not possible.

Reply to Referee No. 2

(referee comments are red; author replies black; new manuscript text blue)

Maybee changing the title to give more e

1. The purpose of this paper is not totally clear.

We have re-written the abstract to emphasize the purpose. The purpose of this paper is to demonstrate to the Rayleigh lidar community that our OEM retrievals can be used to reformulate climatologies over long time periods. The revised abstract describes what is new in the paper.

\cite{ch80} developed a robust method to calculate middle atmosphere temperature profiles using measurements from Rayleigh-scatter lidars. This traditional method has been successfully used to greatly improve our understanding of middle atmospheric dynamics, but the method has some shortcomings in regard to the calculation of systematic uncertainties and vertical resolution of the retrieval. \cite{sica2015} have shown the Optimal Estimation Method (OEM) addresses these shortcomings and allows temperatures to be retrieved with confidence over a greater range of heights than the traditional method. We have developed a temperature climatology from 519 nights of Purple Crow Lidar Rayleigh-scatter measurements using an OEM. Our OEM retrieval is a first-principle retrieval where the forward model is the lidar equation and the measurements are the level 0 count returns. It includes a quantitative determination of the top altitude of the retrieval, the evaluation of 9 systematic plus random uncertainties, as well as the vertical resolution of the retrieval on a profile-by-profile basis. Knowledge of the full random and systematic uncertainties for a temperature climatology is essential for its application for scientific studies of dynamics and atmospheric temperature change. Our complete calculation of the uncertainty budget is compared where possible to previous Monte Carlo simulations by \cite{leblanc2016T}, validating the OEM calculations. Our OEM retrieval allows the vertical resolution to vary with height, extending the retrieval in altitude 5 to 10\,km higher than the traditional method. It also allows the comparison of the traditional method's sensitivity to two in-principle equivalent methods of specifying the seed pressure: using a model pressure seed versus using a model temperature combined with the lidar's density measurement to calculate the seed pressure. We found that the seed pressure method is superior to using a model temperature combined with the lidar measured density. The influence of the \textit{a priori} on the retrieval is quantified and we set a reasonable cutoff height index for the OEM, which is validated by comparing our results to sodium resonance fluorescence lidar temperature measurements. The increased altitude capability of our OEM retrievals allows, for the first time, comparison of the Rayleigh-scatter lidar temperatures throughout the altitude range of the sodium lidar temperatures. Our OEM-derived Rayleigh temperatures are shown to have improved agreement relative to our previous comparisons using the traditional method, and that the agreement of the OEM-derived temperatures is the same as the agreement between existing sodium lidar temperature climatologies. This detailed study of the calculation of the new Purple Crow Lidar temperature climatology using the OEM establishes that it is both highly advantageous and practical to reprocess existing Rayleigh-scatter lidar measurements which cover long time

The statement "the for

Is "measure" the

The statement ".. allow

The abstract seems rather long and not focussed to this

periods, during which time the lidar may have undergone several significant equipment upgrades, while gaining an upper limit to useful temperature retrievals equivalent to an order of magnitude increase in power-aperture product due to the use of an OEM.

2. Comment: it does not seem to add substantial new work regarding the methods compared to SH2015.

The revised abstract clarifies the new results in this work. SH2015 presented the new method, along with technical details of the retrieval. They showed the results for 9 nights of measurements in a period of a few months in 2012. This paper uses the OEM to re-generate the PCL climatology with the entire database, 519 nights, providing for the first time a full uncertainty budget and vertical resolution. The new climatology also brings the retrievals upward a full scale-height over the previous climatology due to the use of the OEM method. To accomplish this increase in altitude via hardware would require the power-aperture product of the system to increase by an order of magnitude, which would be prohibitively expensive. Our NDACC lidar colleagues are very supportive of our efforts with OEM, and urged us to apply the method to a large dataset to thoroughly validate it, as well as demonstrate its advantages and flexibility. This manuscript is the result of that effort.

3. Comment: The dataset used is extended from AS2007, and the methods applied therein are repeated.

AS2007 presented the methodology used to create the composite temperature climatology in detail. We used two paragraphs to summarize the AS 2007 methodology in order to discuss the retrieval grid size and the vertical resolutions for the OEM and HC.

4. The fact that an additional 10-15 km can be gained must be attributed to SH2015. The scientific value therefore lies in the temperature climatology in this 10-15 km (which I think is rather 5 km looking at Table 4) wide altitude range in the mesopause region (and the fact that finally the gap between Rayleigh and sodium temperatures is closed).

This manuscript shows that gaining 10-15 km for the entire climatology is possible using the SH2015 method. The important point about 5 km rather than 10-15 km is that, if the seed pressure for the HC method is not accurate, between 10 to 15 km of the calculated temperature profile is affected by the seed pressure (introducing 20 to 30 K uncertainty) and should be removed. Because there is not a mathematical tool in the HC method to measure the uncertainty due to the seed pressure (except the Monte Carlo methodology presented by Leblanc et. al. 2016a), the lidar community typically removes 10 km from the top of each profile. Our paper demonstrates that with the OEM, one does not need to throw out these measurements in an *ad hoc* manner.

5. However, no detailed geophysical interpretation and discussion on the implication of these newly retrieved data is included, however.

The purpose of this paper is not to focus on the geophysical variability; the geophysical variability was calculated to show the similarity to AS2007. The paper is focussed on the new results discussed previously to allow this climatology to be formed; future studies will use the climatology for geophysical studies. We feel that the level of interpretation and discussion is reasonable for the scope of *Atmospheric Measurements Techniques*.

6. The demonstration that a method valid for single profiles is also valid for a long-term dataset does not make a very strong point, in my opinion.

We agree; results for a few nights are not sufficient to prove our OEM retrievals are valid for a full dataset, and that is one of the important purposes of this work. The PCL was upgraded during the time the data were taken (in 1999 and 2012), and the OEM forward model changes each time that the hardware properties change. SH2015 showed retrievals for a single configuration of the system. Our forward models are first principle models, that is they completely describe the instrument and work from the raw measurements. Hence it is possible, but not trivial, to test and implement the forward model for different types of detectors, numbers of channels, data grids, etc.

This contradicts the stat

It is important to show that the OEM can be applied to lidars with different properties (like in SH2016), and with different detector configurations. We spent a long time determining how to apply the OEM to the measurements from previous years to have reasonable results, particularly when only a single channel was available. Each different situation had to be verified, but now we are able to process the entire set of measurements. One purpose of this paper is to demonstrate to the Rayleigh lidar community that our OEM retrievals can be used to reformulate climatologies over long time periods where instruments or instrument specifications change, in addition to being applied to new measurements.

Is the "how to" sufficient

7. Perhaps by design, the current paper shows very strong resemblance to SH2015 and AS2007, the latter being cited eleven times.

You are correct, this is by design. Repeated citation from these papers is necessary to avoid repeating material in the two earlier papers, or adding unnecessary length to the paper.

8. Almost all of the work seems to be reproduced from these two papers, either repeating equations, lists or arguments, or reproducing the same figures with updated datasets. For example, equation 1, 2, 5, 6 are taken from SH2015 (eqn. 1, 3, 8, 9), Table 1 and 2 are updated from AS2007, Fig. 3, 4, 5 are reproduced matching AS2007 (Fig. 1, 3, 6). The numbered list in 3.1 is taken from SH2015, the numbered list in 4.2 is rewritten from AS2007 (section 8), section

4.2 is very similar to AS2007 (section 7). Even the same nightly profile (24 May 2012) is used as by SH2015.

We tried to keep repeating equations to a minimum, but at some point it become too difficult for the reader if the equation is not in front of them. The lidar equation and its component parts are often repeated in lidar papers, and in the OEM community it is typical to clarify the basic method employed. The origin of equation 1 is from SH2015, however, the origin of equations 5 and 6 is from Rodgers 2011, whose method was cited in line 14. However, we will make the citations more obvious so that the sources are clear. It is necessary to use these equations to adequately introduce the material so that the reader can follow the rest of the paper.

The best way of comparing the two methods was to compare the calculated climatologies using each method with independent measurements, here the Na lidar climatologies, which are based on entirely different physical assumptions (e.g. kinetic temperature determinations from Na line widths). There are few available Na lidar temperature climatologies in the same latitude region as the PCL, therefore we used the same Na climatologies that were used in AS2007 as well as a new one which was published recently (upgraded CSU).

The same PCL temperature profile was used to show the effect of the *a priori* in order to avoid showing the entire temperature retrieval again but we can refer back to SH2015 for the temperature retrieval details. In general, we did not want to sacrifice clarity in our own discussion solely to avoid citing or never reproducing from these two closely related earlier papers.

9. A lack of structuring also becomes obvious in section 2.2, whose titles read:
2.2.1 HC Method
2.2.2 Optimal Estimation Method
2.3.1 OEM Methodology
2.3.2 HC Methodology

Good point. Sections 2.2.1 and 2.2.2 have been changed to HC theory and Optimal Estimation theory for clarity.

10. Citing Khanna et al. in the HC section (p. 5, l. 15) is confusing. Khanna et al. also claim to gain 10-15 km altitude range, so it would perform equally well to your algorithm? If this is so, this should be discussed.

We agree this should be clarified in the paper, as Khanna et al. did not have a regularization term in their retrieval. We have added this to the paper:

Khanna et al. (2012) used an inversion approach to retrieve the temperature using a grid search method and Jalali (2014) applied the grid search method to calculate the PCL temperature climatology. Jalali (2014) then compared the results with the HC temperature climatology. The grid search is a least-squares approach applied to a non-linear forward model. The main difference between the grid search method and the OEM is first, the lack of a regularization term in the grid search forward model. Second, the grid search method uses a Monte Carlo technique to calculate the statistical and seed pressure uncertainties, which computationally is extremely slow. The grid search method gained 10\,km in height over the HC method, but it does not provide the same advantages as the OEM does. For example, the grid search method does not provide the full uncertainty budget, vertical resolution, and averaging kernel. Additionally, the grid search method cannot use several channels of measurements to retrieve a single temperature profile, but requires gluing of photocounts or merging temperature profiles, which introduces additional uncertainties which are difficult to quantify.

11. However, I think this whole section 2.2 is much better explained in SH2015, so it is unfortunate to repeat here in this way. In all this repetitions it is hard to see what part of this work is original or goes beyond the two former papers. The original achievements should have been highlighted more clearly, and larger parts of the text substituted by citations.
The SH2015 paper focused on methodology and this paper focuses on application of their method, and its practicality for use on a large data set. In fact, this work was inspired by requests from our NDACC colleagues to see a demonstration of the method on a large data set before committing to trying a new processing technique. The revised Abstract and Summary statements clarify what work is new.

12. I feel unhappy with the seeding of HC by CIRA-86 (p. 16, l.3), knowing that CIRA is way off as you have shown before. I see that using CIRA for both methods gives them the same starting conditions, (and CIRA is guaranteed to be available in the future..) but this makes it a more academic endeavour, which is fine. In practice however one would seed HC with more realistic SABER temperatures, of which you might have > 10 yrs available, or with the sodium temperatures themselves, of course. This probably means that you could have created the PCL climatology with HC just the same.
We are not sure about the statement that “CIRA is way off”. In what sense? For every temperature profile we determine? On parts of a night if we analyze 10 min averages? Monthly? The point is you never can be sure at a particular time how representative the model or an individual measurement is. The problem with the tie-on pressure uncertainty is you don’t know on an individual retrieval whether the model is spot-on or off by 20% or more, as we showed by using different estimates of the seed pressure. This unknown difference is why the community in general cuts 10 km or more off each profile, because even if the tie-on pressure is 20% or more off by on the order of 1.5-2 scale heights the integration in the HC method converges to the same place within the statistical uncertainty of the measurement at that height. So even if you used SABER if wouldn’t help, you still wouldn’t know on a given night (where the overpass coincidence may be off hours in time and 100s of kilometers or more) that the individual SABER

measurement was better than CIRA. We would recommend someone who uses the HC method cut off 10-15 km from the top of the profile. This recommendation was supported by the Monte Carlo simulations of LeBlanc et al. 2016b.

It is again important to stress that our retrievals are highly insensitive to the choice of the *a priori* (please see for example Fig. 1, which demonstrates that for very different choices of the *a priori* the retrieval changes less than its statistical uncertainty), and that our results, summarized in the revised abstract, are about the advantages of using an OEM method for temperature retrieval, not the effect of tie-on pressure on the HC method. The advantages of OEM go far beyond just this single systematic uncertainty.

13. Fig. 3 is a replication of AS2007 Fig. 3. Both show lidar minus CIRA, however the sign looks to be inverted, am I correct on this? E.g. Nov, 80 km, there is -15 K in AS2007 but +15 K here.

Thank you. Figure 3's caption has been corrected.

14. Fig. 4: this seems to be an unfortunate colorbar. In Apr-Sep, 30-70 km the contour line says 4 K while the colour definitely says < 0.5 K. As it is the same plot as in AS2007, also in Aug-Sep 80-90 km there seems to be more than 3 K difference to AS2007. What's the explanation for this?

Thank you for pointing this out. We changed the colorbar and contours to be clearer than before. The figure colorbar choice is intentional and appropriate to show the temperature changes.

I think the main concern was:

15. p.1, l. 10: "our new retrieval": Is this algorithm different from SH2015?

The revised abstract does not contain this phrase.

16. p. 2, l. 32: it would be beficial to the reader to expand the summery paragraph to include the name of the instrument, the reference to the published climatology and a quick walk-through through the sections.

The last paragraph in the introduction is now:

We have created a new climatology with measurements from The University of Western Ontario's Purple Crow Lidar (PCL) using the Optimal Estimation Method (OEM). Section 2 summarizes the Rayleigh temperature retrieval methods including the HC method and the OEM, as well as the procedure for generating the climatology. Section 3 compares the OEM results with the HC results. Section 4 presents the comparison between the PCL temperature OEM climatology with other sodium lidar climatologies.

17. p. 7, l. 6: "methodology of Argall and Sica (2007)" -> please state again what method this refers to

Argall and Sica (2007) used the HC method. This sentence has been added:
The climatology using the methodology of Argall and Sica (2007), who used the HC method, was formed as follows.

18. p. 7, l. 9: “Unlike..” this sentence is hard to follow. What is “our altitude”, which method? Which decrease of initial height? What comparison with OEM climatology in a paragraph under the title “HC Methodology”?
We have modified the manuscript as suggested as follows.
Old:
Unlike Argall and Sica (2007), our highest altitude with minimum signal to noise ratio of 2 was 90\,km rather than 95\,km because the decrease in the initial height of integration led to having more nights to have a better comparison with the OEM climatology.
New:
Argall and Sica (2007) used the nightly averaged measurements with minimum signal to noise ratio of 2 at the initial height of integration of 95\,km, however, this height is reduced to 90\,km in this study because the decrease in the initial height of integration led to having more nights, which allowed a better comparison with the OEM climatology.

19. p. 9, l. 2: this sentence could be improved, e.g. the expression “using the OEM were used”. Which nightly mean profiles? Temperature? Then why use nightly mean profiles to calculate nightly average profiles, isn’t it the same?
We rewrote the sentence to be more clear.
Old:
The nightly mean profiles for each day of the year using the OEM were used to calculate the nightly average temperature profiles to create the temperature climatology (Fig.~\ref{fig:OEM_CIRA}).
New:
To create the temperature climatology, we used the nightly OEM temperature profiles to calculate an average temperature profile for each day of the year (Fig.~\ref{fig:OEM_CIRA}).

20. p. 15, l. 22: why don’t you use the same ozone profiles?
The temperature differences between the OEM and HC climatology is on the order of 0.5 K and the effect of the ozone profile is 0.05 K. Therefore, the ozone profiles cannot cause a 0.5 K difference. The main reason has been added to the paper.

Old: This bias is likely due to differences in the ozone profiles used for the two climatologies, which causes temperature differences on the order of+0.05 K.

New: The bias due to differences in ozone profile between the two climatologies is only +0.05 K. The OEM used two Rayleigh channels (HLR and LLR) measurements after 1999 to calculate the OEM climatology while only the HLR channel measurements were used for the HC method and the OEM before 1999. The effective LLR signal is up to about 60 km altitude. The temperature difference in the bottom range of measurements is because of using a two-channel retrieval in the OEM and comparing it with a one-channel (HLR) retrieval in the HC method. The 2-channel OEM method retrieves the dead time for each profile while the dead time in the HC method was an empirically determined constant based on count measurements using a pulsed LED source. In order to compare the OEM with HC temperature climatology, we could have merged the calculated LLR and HLR temperature profiles in the HC method. However, the temperature uncertainty induced by the merging will be more than the \pm0.05 K temperature difference between the OEM and HC climatology (Jalali, 2014). Also, using just the HLR channel allowed directly comparison with the AS2007 work.

21. p. 17, top: you might explicitly name the difference between the CSU and upgraded CSU dataset and why you chose to use both instead of the upgraded one only.
The difference between the data sets is the years that are used in the climatology, which is in the end of the Introduction. We added the following sentences to explain our choice.

The URB and CSU climatologies are among the best data sets for validation of upper mesosphere and lower thermosphere temperatures, plus they allow direct comparison between our new climatology and \cite{argallclimatology}. The upgraded CSU \citep{Tao} provides additional years of overlap with our new climatology for validation of our OEM-derived temperatures.

22. The discussion of the quality of the sodium lidar datasets does not belong to the summary, p. 21, l. 18, but to the earlier section where the difference of PCL to the sodium lidar climatologies is shown. Otherwise the large differences remain unexplained. It then looks like a rather bad comparison.
We in no way commented on the quality of the Na lidar datasets, which are some of the best upper mesosphere-lower thermosphere temperature profile measurements available. We believe them to be of high quality. The important point is that the agreement between the Na lidar and the Rayleigh lidars can't be better than the differences between the Na lidars themselves. The difference between the sodium lidars is approximately the same as the differences with the OEM derived temperatures, meaning the temperatures derived using the OEM retrieval are basically the same as measured by the Na lidar, contrary to many other Na vs Rayleigh comparisons which showed differences between the two techniques. Furthermore, the OEM temperature retrievals allow valid retrievals to be obtained in the 95-100 km altitude region, where the systematic uncertainty of the tie-on pressure on the temperature is too large for the temperature to be useful.

Grammar remarks: Thank you. We have made all of your suggested grammar changes.
p.1, l. 4-6: grammar. understanding “of” the connection of temperature “and” ??. to change -> to changes: done

p. 2, l. 8: wrong expression: “satellite resolution” -> the resolution of satellite measurements: done

p. 2., l. 18: (Bills et al., 1991, Krueger et al., 2015): done

Fig. 1: “using the a priori profiles shown in Fig. 1a” -> Fig. 1a does not show an a priori profile, but the difference between two a priori profiles: “Shown” changed to “used”

Fig. 1: more than 90% -> less than 90% and all other figure captions: done

p. 9. l. 5,6: repetition “There is a .. . There is a ...”: Done

p. 10, l. 3: “is it provides” -> is that it provide? : Done

p. 12, l. 6: “almost less than 3 km”: just give the number: Done

p. 7, l. 12: what’s a 3’s and 5’s filter?
It is a low-pass digital filter that removes high frequencies from the signal. This filter uses seven data points for the signal filtering purpose. The convolution coefficients of this seven-point filter are yn = 1/15 [1 2 3 3 3 2 1]. A reference describing this filter has been added to the paper (Hamming, 1989).

p. 23, l. 19: iversion -> inversion: Done

Reply to Referee No. 3

(referee comments are red; author replies black; new manuscript text blue)

1. Page 3, Table 3, 0.2% for uncertainty on air number density taken from CIRA-86 seems unrealistic. Due to the variability of the atmosphere, it should be in the same order than the a-priori uncertainty on pressure profile, around 5%.

We agree, we had intended to use the Leblanc et al. value of 5%. We have recomputed the uncertainties and updated Figures 5, 7 and 9 accordingly. Due to the small contribution of this uncertainty relative to other uncertainty terms, this change has no impact on our results or conclusions.

2. Page 11, Figure 5: I do not understand why all error terms (except the gravity) are increasing with height above 60 km. The proposed explanation on page 18 for the increase of the uncertainty due to ozone cross-section is via the upward integration of the transmission integral but above 60 km the transmission is very close to 1 and I do not expect any effect. Please clarify.

The error covariance on the retrieved quantity (here temperature) due to a model parameter is:

$$S_f = G_y\ K_b\ S_b\ K_b^T\ G_y^T$$

Section 1 | Section 2 | Section 1

where G_y is the gain, S_b is the error covariance matrix for the model parameter and K_b is the model parameter Jacobian (Equation 3.18 in Rodgers).

The error covariance S_f is comprised of 2 sections, the gain matrix (Section 1) and the multiplication of the model parameter Jacobian with the error covariance matrix of the model parameter (Section 2). The S_b values are usually small (e.g. a few percent of the ozone density in the mesosphere). Also, K_b is small at higher altitudes as the retrieval is not sensitive to the small amount of ozone at these heights, therefore Section 2 is small and decreasing with altitude. The gain matrix is the sensitivity of the retrieval to the measurements:

$$G = \frac{\partial \hat{x}}{\partial y} = (K^T S_\varepsilon^{-1} K + S_a^{-1})^{-1} K^T S_\varepsilon^{-1}$$

inversely?

The gain matrix is proportional to the measurement noise (S_ε). The measurement noise for a Rayleigh-scatter lidar using digital detection is that due to photon counting, which means the measurement noise decreases with height and S_ε^{-1} rapidly increases at higher altitudes. Large values of gain means the retrieval is more sensitive to measurement noise (Rodgers, page 10). In other words, the gain matrix can be thought of as amplifying the retrieval's uncertainty at the greatest heights. Thus, the increase in measurement uncertainty (Section 1) increases at a much faster rate than the model parameter error decreases (Section 2), and the uncertainty on the retrieved temperature due to ozone (and many of the other model parameters) increases with altitude.

We have changed:
3. The other parameter that has significant uncertainties at higher altitudes is ozone cross section, whose uncertainty propagates upward via the transmission integral. It reaches a maximum of 1 K at 100 km.
To
3. The other significant contribution to the temperature uncertainty budget at higher altitudes is ozone cross section, whose uncertainty increases with altitude due to increasing measurements uncertainty (as do many of the retrieval's uncertainties due to the model parameters). The uncertainty on the retrieved temperatures due to ozone reaches a maximum of 1 K at 100 km.

3. Page 2, lines 24-25 : Something is missing on the sentence "They also discovered . . .. than the models". Please rewrite.
The sentence is reworded as below and added to the paper:

The lidar measurements showed that the mesopause altitude was lower in the summer than in the winter. The models did not predict the observed seasonal behavior, showing little difference in altitude.

4. Page 4 equation (1): B may depend on altitude if there is some signal induced noise and should be written B(z).
You are correct, and we have rewritten *B* explicitly as a function of altitude, as the background can be height-dependent. The PCL has a constant background, thus, we use a constant background in our forward model.

5. Page 6, line 26: Please define what is the "lidar constant" for non-specialists.
We agree but suggest the explanation be on page 4 line 5 where we introduce the lidar equation and lidar constant.

We would add:

> Various lidar system parameters and physical constants affect the photocounts independent of altitude. The combination of these parameters is called the lidar constant and in our definition includes: the number of photons emitted by each laser pulse, the optical efficiency, the detection efficiency of the photomultipliers, atmospheric Rayleigh scatter cross section and the speed of light.

6. Page 7, line 4: The a priori variance for CIRA-86 is expected to increase with altitude. Climatology is based on less information at higher altitude.
The variance of CIRA-86 is considered constant for the entire temperature profile. SH 2015 investigated the choice of variance in detail, and found no difference or advantage to varying this quantity with height to a maximum of (35 K)^2, as given by Fleming et. al. (1988). For the temperature retrievals here, the *a priori* has essentially no contribution up to around 90-95 km altitude, where (35 K)^2 is a reasonable choice for the variance.

7. Page 15, lines 20-23: The proposed explanation for the warmer OEM temperature than HC temperature from 40 to 60 km is probably not the differences in ozone profiles that contribute only to 0.05K, one tenth of the observed bias. Is it possible that the smoothing procedure has an impact on the retrieved temperature at the stratopause?
The following sentence in line 21 is incorrect and has been changed:
Old: This bias is likely due to differences in the ozone profiles used for the two climatologies, which causes temperature differences on the order of +/-0.05 K.
New: The bias due to differences in ozone profile between the two climatologies is only +0.05 K. The OEM used measurements from two Rayleigh channels (HLR and LLR) after 1999 to calculate the OEM climatology, while only the HLR

channel measurements were used for the HC method and the OEM before 1999. The effective LLR signal is up to about 60 km altitude. The temperature difference in the bottom range of measurements is because of using a two-channel retrieval in the OEM and comparing it with a one-channel (HLR) retrieval in the HC method. The 2-channel OEM method retrieves the dead time for each profile while the dead time in the HC method was an empirically determined constant based on count measurements using a pulsed LED source. In order to compare the OEM with HC temperature climatology, we could have merged the calculated LLR and HLR temperature profiles in the HC method. However, the temperature uncertainty induced by the merging will be more than the \pm0.05 K temperature difference between the OEM and HC climatology (Jalali, 2014). Also, using just the HLR channel allowed directl comparison with AS2007.

If this comparison was actually done the result sho

8. Page 17, line 3-17: I am not sure that the better agreement between sodium lidar and OEM is significantly better than between sodium lidar and HC. First the differences are not so large, HC difference is 1.2 K warmer than OEM difference on average, and second part of the difference may be due to the distance between the sodium lidars and the Rayleigh lidar.

You make two interesting points. First, that the differences between methods is not large, and second, there are geographical differences in the sodium lidar locations. As you said, when you consider the numbers in Table 4 in the 85-90 and 90-95 km bins where both Rayleigh temperature methods and the Na lidars have good measurements, the OEM-calculated temperatures show 20% better agreement with the Na lidars than they do with temperatures calculated with the HC method (5.0 K versus 6.3 K). How significant is this improvement? To determine the significance consider the variability between the Na lidars themselves, which is 4.5 K, and primarily due to geographical differences. The difference among the sodium lidars is approximately the same as the differences with the OEM derived temperatures, meaning the temperatures derived using the OEM retrieval are basically the same as measured by the Na lidar, contrary to other Na vs Rayleigh comparisons which showed differences between the two techniques. Furthermore, the OEM temperature retrievals allow valid retrievals to be obtained in the 95-100 km altitude region, where the systematic uncertainty of the tie-on pressure on the HC method-derived temperatures is too large for the temperatures to be useful.

Please give a concise answer that should be include in the manuscript as