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:

$$S_f = G_y \underbrace{K_b \; S_b \; K_b^T}_{} \; G_y^T$$

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:

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

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 ± 0.05 K temperature difference between the OEM and HC climatology (Jalali, 2014). Also, using just the HLR channel allowed directl comparison with AS2007.

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.