Dear Referee,
Thank you very much for your helpful comments and suggestions. I have attempted to address each of your concerns to the best of my ability. If you would like me to implement further changes or iterations on a point please let me know. I appreciate your efforts to help me improve this paper.

The questions you raise about error estimation after removal of signal induced noise contributions are in my opinion very critical. I think that as a lidar community we really need to push the envelope on our data retrieval techniques and error estimates - particularly if we want to do new work in the Mesosphere/Thermosphere. The work in this paper is not perfect but it is an improvement to the commonly used (Hauchecorne/Chanin 1980) lidar temperature inversion. It's my belief that as a community we should continue investigating improvements to our techniques. A fully Bayesian approach such as the Optimal Estimation Technique presented by (Sica/Haefele 2015) might be a profitable endeavor. As an added benefit a Bayesian Technique produces full averaging kernels which would make lidar data much more attractive for assimilation to people in the satellite and reanalysis communities .
****

Response *Lidar temperature series in the middle atmosphere as a reference data set. Part A: Improved retrievals and a 20 year cross-validation of two co-located French lidars*: Referee #1

---

1) Page 6: Lidar equation (1) has dimension mismatch on the left and right hand sides. The first part on the right side has a dimension of energy, but the left side N(z) is claimed to be count rate per time integration per altitude bin. This equation is not acceptable for publication. Furthermore, beta (β) is commonly used to represent volume backscatter coefficient, not backscattering cross-section, as the cross-section symbol is usually sigma (σ). Authors are suggested to consult with a commonly referenced class lecture at the following link, and use the more commonly accepted lidar equations and symbols.

Good catch thank you. I've changed divided the right hand side by the photon energy hc/lambda and changed β_cross to σ_cross

---

2) Page 9: Please provide a reference to Turkey Quartile test, as this isn't a common practice for most lidar people. BTW, it should be "when the signal to noise ratio approaches 1".

Cited Tukey(1949)
"signal to noise" changed to "signal to noise ratio"

---

3) Page 12: Please provide a reference to the "one sided non-parametric MannWhitney-Wilcoxon rank-sum test" as it isn't common for the lidar field. BTW, what does "a scan" mean in Figure 6? Did you mean one profile?

Cited Mann and Whitney (1947)

I've always called a single level_0 or level_1 photon count time series a 'scan' and used the word 'profile' for level_2 things like density, pressure, temperature. Reviewer #2 and Reviewer #3 made the same point so perhaps it's a personal idiosyncrasy. In any case, I've changed all occurrences of 'scan' to 'profile'

---

4) Page 14, how are Si and Ni determined? Please provide a bit more details. Do you do this (equation (2)) for every altitude bin?

Line 268 - 272
The noise is always evaluated between 120 km and 155 km and the altitude range for the evaluating the signal is defined as the scale height below the point where the signal to noise equals one in the density profile. Each individual profile has a value representing the signal, S_{i}, and a noise, N_{i}. The profile values are compared to the nightly sum of the signal, S_{sum} and the nightly sum of the noise, N_{sum}.
Changed to:
***T*he noise of an individual profile, N_{i}, is expressed as the summation of photon counts in bins which fall between 120 km and 155 km and the nightly noise, N_{sum} is the summation of all N_{i} for the night. To determine a metric for the nightly average lidar signal, S_{sum}, we first calculate a quick density profile and determine the lowest altitude where the signal to noise ratio equals 1. Then we calculate the altitude that is one density scale height (~8 km) below this point. The lidar range bins which correspond to this altitude range are then summed to yield S_{sum}. A similar calculation, using the same range bins as in the nightly average calculation, is done to determine the signal of single profile, S_{i}.**
---

5) Page 16, notations are needed for equation (3).

N, τ, and Δt are described in lines 309-310. We have now replaced the definition of N with separate definitions for $N_{counted}$ and $N_{received}$, as they appear in Equation (3):
Replacement text:

The background theory and derivation of Eq. (3) is well described by (Donovan et al., 1993), where $N_{received}$ is the number of photons incident on the PMT per measurement time interval and $N_{counted}$ is the number of photons per measurement time interval which are actually counted by the system. In general, $N_{counted}$ < $N_{received}$ due to effects of the system deadtime.

---

6) Page 16, after the quadratic fit to the background, how do you handle such background and data? Did you mean to subtract the quadratic fitted background from the raw data? In this case, how do you handle the noise term in calculating SNR? Are photon counts still obey Poisson distribution? Please clarify in the manuscript.

Yes, in the case of a quadratic background I subtract the quadratic function from the entire photon counts profile in exactly the same way I would treat a constant or linear background.

As you correctly point out, as soon as there is signal induced noise the profile is no longer Poisson as the count rate in each lidar bin is no longer fully independent of the count rates in the bins on either side of it. The Total counts are some combination of 'Real counts' and 'Contamination counts ' (T = R + C) with a common shot noise dT = 1/sqrt(T) with some contribution dC = 1/sqrt(C) coming from the Signal Induced Noise portion and dR = 1/sqrt(R) representing the noise from all other sources. When I'm using the linear or quadratic backgrounds I am making an assumption that I'm completely removing the signal induced noise, C and I no longer have to add dR and dC in quadrature. I'm approximating dN ~= dR and that the photon count profiles are now approximately Poisson.

On page 16 line 334, we have added a new sentence to the manuscript:
"...as our estimate of signal induced noise. The best background function is subtracted from the raw photon counts profile."

On page 16 line 341 we have added a new sentence to clarify about SNR:
"...than the simple quadratic approximation. For the quadratic case, as soon as there is signal induced noise the profiles no longer represent Poisson distributions as the count rate in each lidar bin is no longer fully independent of the count rates in the bins on either side of it. Therefore, precise calculations of the SNR would require the addition in quadrature of real noise (from sky background and signal photon counts) and contamination noise (from signal induced noise). Here, however, we make the assumption that the signal induced noise is able to be completely removed from the raw profiles with the subtraction of the quadratic function. We therefore interpret the background subtracted profiles to obey approximately Poisson distributions, thereby

approximating the total noise in the profile to the noise of only the real photons, which can be treated as uncorrelated."

---

7) Page 20, Figure 11, it's necessary to point out in the manuscript that satellite data aren't the real references as various satellites have their own calibration issues. Rayleigh temperatures around 90 km should be compared with ground-based resonance Doppler or Boltzmann lidar temperatures as these resonance lidars have much better signal to noise ratios at these altitudes.

Line 414 inserted text: **It is important to note that additional complications exist when comparing temperatures derived from ground based lidars to temperatures derived from satellite data which have their own calibration concerns. We explore the issues of lidar-satellite comparison in Part B of this paper. A co-located ground-based resonance Doppler or Boltzmann lidar would provide a better comparison data set as resonance lidars have high signal to noise ratios above 85 km (Alpers, 2004).**

---

8) Page 22-23, what do you mean by "misaligned"? A lidar beam was misaligned relative to its own receiver's field of view, or else? How were two lidars misaligned? Authors' writings here are confusing.

In both lidar systems the high gain Rayleigh channel has 4 mirrors, each of which needs to be aligned independently with respect to the laser in the sky and also the fibre optic with respect to the primary focus of the mirror. In LTA the low gain channel is a single independent mirror. So a total of 9 mirrors need to be aligned every night to make Rayleigh measurements.

Line 456 inserted text: **Internal misalignments happen when one or more of the five mirrors in LTA or four mirrors in LiO3S is not properly aligned with the laser or the fibre optic is not centered on the focal point of the mirror.**

Minor comments on English writing: As this is a very long paper, I strongly encourage authors go over the manuscript carefully to correct grammar and typo issues. For example, on page 24, near line 495, it should be "to initialize the inversion", not "initialized". The paper title doesn't have good English grammar, for which I suggest to change "a 20 year cross-validation" to "a 20-year cross validation

I will have an anglophone colleague read over the paper for grammatical errors.