A middle latitude Rayleigh-scatter lidar temperature climatology determined using an optimal estimation method

Ali Jalali[1], Robert J. Sica[1,2], and Alexander Haefele[2,1]

[1]Department of Physics and Astronomy, The University of Western Ontario, London, Canada

[2]Federal Office of Meteorology and Climatology, MeteoSwiss, Payerne, Switzerland

Correspondence to: sica@uwo.ca

Abstract. Hauchecorne and Chanin (1980) 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. Sica and Haefele (2015) 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 Leblanc et al. (2016b), 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 *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 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.

1 Introduction

Improving middle atmosphere temperature climatologies is a priority focus of programs such as those led by the Stratosphere Reference Climatology Group, part of the World Climate Research Programme (WCRP) Stratospheric Processes and Their Role in Climate (SPARC) project. Defining middle atmosphere temperature trends, including those in the stratosphere, mesosphere, and lower thermosphere (MLT), is important for understanding the connection of temperature variations in the middle atmosphere to change in the lower atmosphere. Ramaswamy et al. (2001) and Randel et al. (2004, 2009, 2016) discussed the effects of the middle atmosphere temperature trend over time using different instruments. The MLT region is too high for weather balloons to measure the temperature and satellite resolutionthe resolution of satellite measurements is on the order of 2 km or greater in this region. Rocketsondes were used for studying this region but high cost and discontinuous measurements were two large deficiencies. Nightglow imagers and hydroxl imagers are other instruments that are used to investigate the mesopause, but it is difficult to access their vertical resolution.

One of the best instruments for high spatial and time resolution temperature measurement is lidar. The primary lidars that operate in the stratosphere and lower mesosphere are Rayleigh lidars, while Rayleigh and sodium lidars are best in the upper mesosphere and lower thermosphere. Sodium lidars measure the height-dependent kinetic temperature in the sodium layer (roughly 83 to 105 km). Rayleigh lidars measure relative density; by assuming hydrostatic equilibrium between layers and applying the Ideal Gas Law, a temperature profile can be calculated from the relative density measurement. Sodium lidars use the resonant scattering of the transmitted laser pulse from the sodium layer; here temperature accuracy is limited by our knowledge of the received photon noise and transmitted wavelength and line width (Bills et al., 1991; Krueger et al., 2015).

Typically, Rayleigh lidars don't measure as high in altitude as sodium lidars. Several Rayleigh lidar temperature climatologies have been calculated e.g. Leblanc et al. (1998), Argall and Sica (2007) and have been compared with sodium temperature climatologies such as those in She et al. (2000), States and Gardner (2000a) and Yuan et al. (2008). They found significant temperature differences between the measurements and atmospheric model, in particular between 84 and 104 km, with the coldest temperatures occurring during July and the hottest temperatures occurring during November. They also discovered that the mesopause is at lower altitudes during the summer than the winter than the models. 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.

Diurnal and nighttime temperature climatologies were published by States and Gardner (2000a) from Urbana, Illinois (40 ° N, 88 ° W) (URB) using measurements between 1996 and 1998. She et al. (2000) used eight years of nighttime measurements of Colorado State University (CSU) sodium lidar (41 ° N, 105.1 ° W) from 1990 to 1999 to calculate a temperature climatology. The CSU lidar was upgraded in 1999 from a one beam to a two beam lidar to be able to probe the mesopause during daytime and nighttime (Arnold and She, 2003). Yuan et al. (2008) published the results of the upgraded CSU lidar, giving climatologies for nighttime and daytime between 2002 and 2006. 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 Argall and Sica (2007). Yuan et al. (2008) provides additional years of overlap with our new climatology for validation of our OEM-derived temperatures.

In this paper, we have created a new climatology using the Optimal Estimation Method (OEM) 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) with a full uncertainty budget which goes higher in altitude than the climatology using the method of Hauchecorne and Chanin (henceforth the HC climatology), in addition to including systematic as well as random uncertainties. We then compare the OEM-derived climatology with sodium lidar climatologies to validate the Rayleigh-scatter temperatures. 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. Section 5 is a summary and Section 6 is the conclusion.

2 Procedure for generating the climatology

2.1 Purple Crow Lidar (PCL)

The PCL is a Rayleigh-Raman lidar which was located at the Delaware Observatory (42.52 ° N , 81.23 ° W) near The University of Western Ontario in London, Canada from 1992 to 2010 (Sica et al., 1995, 2000; Argall et al., 2000). In 2012, the PCL was moved to the Environmental Science Western Field Station (43.07 ° N, 81.33 ° W, 275 m altitude). The PCL has been upgraded over time. Currently, the PCL transmitter is a Nd:YAG laser with a power of 1000 mJ per pulse at 532 nm and a repetition rate of 30 Hz. The PCL receiver is a liquid mercury mirror with a diameter of 2.65 m. From 1994 to 1998, the PCL used a single detection channel (the High Level Rayleigh (HLR) channel) over the range of 30 to 110 km (Sica et al., 1995). In 1999, a Low Level Rayleigh (LLR) channel was added, which is nearly linear above 25 km (Sica et al., 2000). This study uses 519 nightly averaged temperature profiles from 1994 to 2013 distributed in time as shown in Tables 1 and 2.

Table 1. Number of nightly mean profiles used to calculate the PCL temperature climatology by month between 1994 and 2013.

Month	Number of profiles
January	9
February	14
March	17
April	19
May	63
June	72
July	109
August	99
September	39
October	37
November	26
December	15
Total	519

Table 2. Number of profiles used to calculate the PCL temperature climatology per year between 1994 and 2013.

Year	Number of profiles
1994	36
1995	40
1996	22
1997	17
1998	78
1999	57
2000	43
2001	2
2002	57
2003	34
2004	5
2005	37
2006	32
2007	34
2012	20
2013	5
Total	519

2.2 Rayleigh Temperature Retrieval Methods

In this section, we briefly review the OEM and HC methods that have been used in our calculations. Each approach has its own benefits and deficiencies. Both of these methods start with a lidar return which is proportional to density and then find temperature using the assumption of hydrostatic equilibrium, the Ideal Gas Law, and the lidar equation.

The lidar equation is a mathematical relation between the number of back-scattered photons detected by lidar and the measurable quantities such as altitude, laser power, scattering cross section, etc. If we consider all atmospheric parameters in the lidar equation to be constant in time, the lidar equation reduces to Eq. (1)

$$N_t(z) = C\psi(z)\frac{n(z)}{z^2} + B(z), \tag{1}$$

where C is a constant standing for a combination of all the constant properties of the lidar, $\psi(z)$ includes height dependent parameters like detector nonlinearities or geometric overlap, $n(z)$ is the atmospheric number density as a function of height and $B(z)$ is the background count due to radiation sources other than the lidar laser, which may or may not be height-dependent.

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 speed of light.

When the pressure gradient of an air parcel in the atmosphere is in balance with its gravitational force, the atmosphere is in hydrostatic equilibrium, and is dynamically and thermally stable. The hydrostatic equilibrium equation can be expressed as

$$\frac{dP}{dz} = -\rho(z)g(z), \tag{2}$$

where P is the atmospheric pressure, ρ is the density and g is the acceleration due to gravity.

The mean molecular mass of air is considered to be constant within the 30 to 80 km altitude range. However, the mean molecular mass varies with altitude above 80 km, and this variation affects the temperature retrieval, both through the change in mean molecular mass and the effect of composition changes on the Rayleigh-scatter crossection.

2.2.1 HC Method Theory

In 1980, Hauchecorne and Chanin presented a robust method to retrieve temperature from Rayleigh lidar measurements (Hauchecorne and Chanin, 1980). The HC method assumes that the atmosphere is comprised of isothermal layers and uses an equation derived from the Ideal Gas Law and the hydrostatic equilibrium assumption to calculate temperature from relative atmospheric density. Using the assumption of hydrostatic equilibrium, the Ideal Gas Law and the lidar equation, they found a relation between the measured lidar signal and temperature at each altitude in the lidar range. This relation can be integrated from $z - \frac{\triangle z}{2}$ to $z + \frac{\triangle z}{2}$ for a layer with thickness $\triangle z$ as follows:

$$\log\left(\frac{P(z_i + \frac{\triangle z}{2})}{P(z_i - \frac{\triangle z}{2})}\right) = -\int_{z_i - \frac{\triangle z}{2}}^{z_i + \frac{\triangle z}{2}} \frac{M}{R}\frac{g(z)}{T(z)}dz. \tag{3}$$

We can then use the assumption of hydrostatic equilibrium to express the pressure for each layer upon downward integration and derive the following relation for the temperature (Gross et al., 1997):

$$T_i = P_0 \frac{M}{R\rho_i} + \frac{M}{R} \int_{z_i}^{z_0} \frac{\rho(z)g(z)}{\rho(z_i)} dz. \tag{4}$$

In order to integrate the pressure relation from top to bottom, a pressure obtained from a model is used to "seed" or to "tie-on" the pressure at the highest altitude of lidar measurements. Due to the high uncertainties caused by the pressure estimation from a model, the top 10 to 15 km are required to be eliminated from the top of each temperature profile to have accurate results (e.g. Khanna et al. (2012)). The HC method gives a statistical uncertainty of the calculated temperature which assumes the measurements follow Poisson statistics. 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 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 the lack of a regularization term in the grid search forward model. Also, the grid search method uses the Monte Carlo technique to calculate the statistical and seed pressure uncertainties. 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.

2.2.2 Optimal Estimation Method Theory

Sica and Haefele (2015) used a first-principle OEM to retrieve temperature from Rayleigh-scatter lidar measurements. Here first principle means the forward model is the lidar equation and the measurements to the forward model are the raw (level 0) measurements. The OEM (Rodgers, 2011) solves an inverse problem and uses a forward model to estimate the lidar measurements using a set of input parameters usually referred to as state and model parameters. The inversion of the forward model yields the state vector while the model parameters are known and the measurement is given.

The forward model (**F**) can be written as:

$$\mathbf{y} = \mathbf{F}(\mathbf{x}, \mathbf{b}) + \epsilon, \tag{5}$$

where **y** is the measurement vector, **x** is the state vector, **b** is the model parameter vector, and ϵ is the measurement noise. The state vector is retrieved and contains the temperature profile and some instrument parameters like detector dead time and background. The model parameter vector contains all other parameters needed to represent the measurements. The forward model is the lidar equation (Eq. 1). The measurement noise in lidar measurements implies that the measurements have uncertainties that have a Gaussian distribution of possible values represented by ϵ. The solution of the inverse problem is constrained around

an *a priori* which can be found in atmospheric models, such as the CIRA-86 model or the US Standard model. CIRA-86 can provide the monthly temperatures for use as the *a priori*. In its most likely state $\hat{\boldsymbol{x}}$, the solution is a minimum of a cost function:

$$\text{cost} = [\boldsymbol{y} - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})]^T {\boldsymbol{S_\epsilon}}^{-1} [\boldsymbol{y} - \mathbf{F}(\hat{\mathbf{x}}, \mathbf{b})] + [\hat{\boldsymbol{x}} - \boldsymbol{x_a}]^T {\boldsymbol{S_a}}^{-1} [\hat{\boldsymbol{x}} - \boldsymbol{x_a}]. \quad (6)$$

where $\boldsymbol{S_\epsilon}$ is the covariance of the system's state, $\boldsymbol{x_a}$ is the *a priori* vector, and $\boldsymbol{S_a}$ is the *a priori* covariance.

Unlike the HC method, the OEM produces a complete uncertainty budget for all parameters in the temperature retrieval process on a profile by profile basis. The uncertainty budget includes the uncertainty due to the seed pressure and the other model parameters and measurement noise as well as smoothing. A diagnostic variable of the OEM is the averaging kernel, **A**, which describes how the retrieval reacts to a given change in the real atmosphere. A perfect retrieval means the retrieved temperature changes in the same way as the real atmosphere and **A** is equal to the identity matrix (Rodgers, 2011). However, if the contribution of the *a priori* increases in the temperature retrieval, **A** drops to < 1 at each point in the altitudes where the *a priori* has more influence. If **u** is a vector with unit elements, **Au** is the sum along the rows of the averaging kernel and it can be used as a representation of the amount of information coming from the lidar measurements and how much is as a result of the *a priori*. Therefore, **Au** was used as the cutoff height reference in the OEM instead of removing 1 or 2 scale heights from top of each profile as in the traditional method. Values of **Au** equal to 0.9 and 0.8 are considered as a cutoff height. These values represent the fractional contribution of the measurements as compared to the *a priori* in the temperature retrieval and are generally recognized in the OEM community as levels above which the effect of the *a priori* is minimal.

2.3 Methodology to calculate temperature climatology

2.3.1 OEM Methodology

The OEM uses the forward model and non-paralyzable dead time correction equation (Sica and Haefele, 2015) (henceforth SH2015) to retrieve the nightly average temperature profiles from the LLR and HLR channels simultaneously. In SH2015 the dead time, background and temperature were retrieved. They considered the lidar constant as a forward model parameter, but in this study, the lidar constants for LLR and HLR channels were retrieved rather than specified . The OEM uses an estimation of the covariances of the measurements, retrieval, and forward model parameters. The model parameter covariance matrices used in this study are based on SH2015, where the summary of the values and related uncertainties of the measurements and the retrieval and forward model parameters are presented in Table 3. The data grid is 264 m, and the retrieval grid is 1056 m. Due to the PCL measurements between 1994 and 1998 having only the HLR channel measurements, temperature and background were retrieved but not dead time. Instead, the systematic uncertainty due to the saturation was calculated. The PCL measurements from 1999 to 2011 used the LLR digital channel to get more temperature information and the dead time of the HLR channel was retrieved using an *a priori* value of 10 ns (Table 3). The LLR dead time was treated as a model parameter

Table 3. Values and associated uncertainties of the measurements and the *a priori*, retrieval and forward model parameters

Parameter	Value	Standard deviation
Measurement		
HLR (1994-2013)	Measured	Poisson statistics
LLR (1999-2013)	Measured	Poisson statistics
Retrieval parameters (a priori)		
Temperature profile	taken from CIRA-86	35 K
Background for LLR	Average of photocounts above 90 km	std. above 90 km
Background for HLR	Average of photocounts above 115 km	std. above 115 km
Deadtime for LLR and HLR (1999-2011)	10 ns	5.7 and 11.19% respectively
Deadtime for HLR (2012-2013)	4 ns	0.5%
Lidar constant for HLR	Estimated using forward model (55-60 km)	10%
Lidar constant for LLR	Estimated using forward model (45-50 km)	10%
Forward model parameters		
Pressure profile	Fleming et al. (1988)	5%
Ozone density	McPeters et al. (2007)	4%
Ozone cross section	Griggs (1968)	2%
Acceleration due to gravity	Mulaire (2000)	0.001%
Rayleigh scattering cross section	Nicolet (1984)	0.2%
Air number density	CIRA-86	0.2%5%

and a standard deviation of 5.7% was considered.The CIRA-86 model atmosphere was chosen as the temperature *a priori* with a variance of $(35\,\mathrm{K})^2$ at all altitudes (Fleming et al., 1988).

2.3.2 HC Methodology

The climatology formed using the methodology of Argall and Sica (2007) (henceforth AS2007), who used the HC method, was formed as follows. First, the quality of each one-minute scan profile of measurements was checked. Then, nightly averaged temperature profiles were calculated. The quality of the nightly averaged measurements was assessed based on the measurement signal-to-noise ratio. Measurements were accepted if the signal-to-noise ratio was greater than 2 at the highest altitude for the initialization of downward integration. 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.AS2007 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 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. The raw photon count profiles have been co-added to produce height bins of 1008 m and a 3's and 5's filter (Hamming, 1989) was applied to the calculated temperature profiles to smooth them in the climatology. The co-added height value of 240 m was

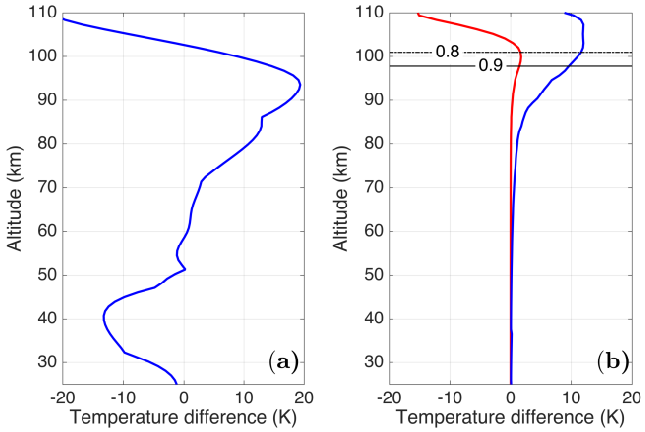

Figure 1. a) Temperature difference between the *a priori* temperature profiles, US Standard Atmosphere and CIRA-86 (blue line). b) Temperature difference between the OEM retrieved temperature profiles using the *a priori* profile shown used in Fig 1a, for 24 May 2012 (red line) and the calculated OEM statistical uncertainty (blue line). The solid black and solid-dashed black lines are the height above below which the temperature profile is more than 90% (0.9) and 80% (0.8) due to the measurements, respectively.

chosen as a data grid for the OEM to be consistent with the vertical resolution of the HC. The vertical resolution definition and calculation is based on Leblanc et al. (2016a).

The following steps were followed to make a composite year temperature climatology after calculating all lidar temperature profiles. Only the profiles were used that, after removing the top 10 km, extended up to 80 km. Each temperature profile was then interpolated to an altitude grid with 1 km interval between 35 and 110 km. For dates with multiple measurements over the years (e.g. 25 July 1994, 2003 and 2003), a weighted average profile was calculated using each profile's statistical uncertainty as weights. Then, linear interpolation was used to fill the gaps where no measurements existed and a 33-day triangular filter was applied to smooth the composite temperature climatology.

2.4 Effect of *a priori* on the retrieval temperature profiles in the OEM

A retrieved temperature profile using the CIRA and the US Standard Atmosphere as the *a priori* for a sample PCL night (24 May 2012) was plotted in order to demonstrate the contribution of the *a priori* temperature profiles in the retrieval results, as well as the temperature difference between the *a priori* temperature profiles (Fig. 1). The temperature difference between the *a priori* profiles is shown in Fig. 1a. The temperature difference around 94 km which is below the OEM cut-off heights is about 20 K. In Fig. 1b, the red profile is the temperature difference due to the *a priori* and the blue profile is the statistical uncertainty calculated by the OEM. The 0.9 and 0.8 value lines in the **Au** are the cut-off heights for the OEM and are shown with solid and dashed lines, respectively. It can be seen that the choice of *a priori* has little effect (1.5 K below the 0.9 line and less than 2 K below the 0.8 line) on the retrieved temperature and that the difference between the retrieved temperatures from each choice is much less than the statistical uncertainty (10 K below 0.9 line and 12 K below 0.8 line) at the top of the profiles.

3 Results

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. 2).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. 2). The 0.9 and 0.8 values of **Au** are superimposed in Fig. 2 with white lines. To estimate the annual temperature variability, temperature difference between PCL temperature climatology using the OEM and the calculated climatology from monthly CIRA-86 temperature profiles are plotted in Fig. 3 for each month. There is a temperature difference on the order of 5 K below 52 km. There is a bias smaller than 3 K between the CIRA profiles and the PCL monthly mean temperatures between 55 and 65 km except in the winter. Above 65 km the CIRA is warmer, on average around 8 K, than the PCL up to 90 km, but much colder (on the order of 20 K) above 90 km. CIRA temperature profiles have a smaller difference (less than 10 K) as compared to the PCL in summertime rather than wintertime up to around 90 km.

The geophysical variability for the OEM PCL temperature climatology (Fig. 4) was calculated based on the difference between the 33-day temperature standard deviation and the variability of the PCL measurements. The geophysical variability shows the wave activity in the time range of 2 to 33 days. We followed the procedure from AS2007 based on Leblanc et al. (1998) to calculate the geophysical variability. Figure 4 shows the temperature variability related to waves from 2 to 33 days. In Fig. 4 the white spots are due to the negative values under the square root of the difference between the variance of the root mean square of the temperature profiles and temperature statistical uncertainty in the geophysical variability calculation. The temperature change from mid-April to the end of September below 70 km is less than 4 K. However, in the same period of time the highest temperature change is between 80 and 90 km due to the wave activity in the mesosphere. There is a peak at 41 km in January which is related to sudden stratospheric warmings during winter (Argall and Sica, 2007). 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 (AS2007). The temperature variability due to mesospheric inversion layers reaches a maximum between 62 and 72 km during December and January. These results are in good agreement with the results presented in Figure 6 of AS2007.

3.1 Uncertainty budget and vertical resolution

The lidar measurements include both systematic and random uncertainty. Systematic uncertainties originate in the forward model from uncertainties due to model parameters. One of the advantages of the OEM is that it provides systematic uncertainties for all retrieved parameters, as well as the random uncertainties. The systematic uncertainties calculated in the PCL OEM technique (Table 3) are based on the following model parameters:

1. knowledge of the HLR dead time (1994-1998 only)
2. determination of the Rayleigh scatter cross section for air
3. Rayleigh cross section variation with composition in the mesosphere and thermosphere

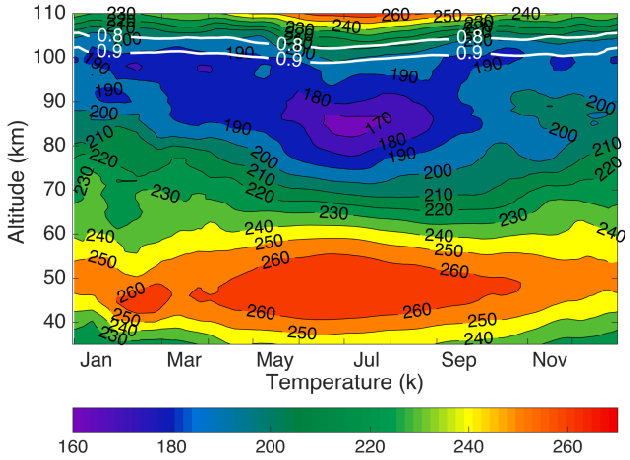

Figure 2. Composite PCL Rayleigh temperature climatology using the OEM. The white lines are the height above below which the temperature climatology is more than 90% (0.9) and 80% (0.8) due to the measurements.

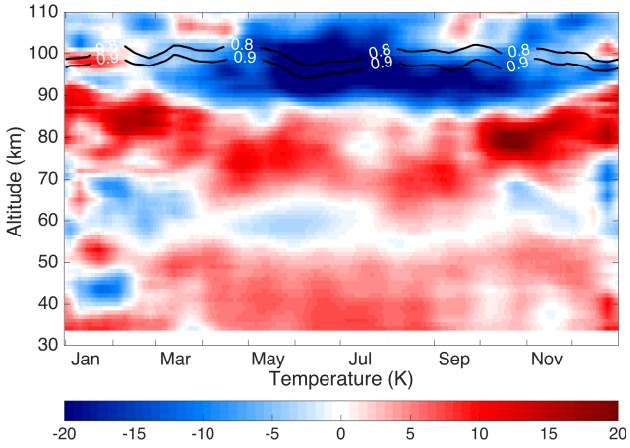

Figure 3. Temperature difference between the calculated climatology from PCL temperature climatology using the OEM and monthly CIRA-86 temperature profiles monthly CIRA-86 temperature profiles and the OEM PCL temperature climatology. The black lines are the height above below which the temperature climatology is more than 90% (0.9) and 80% (0.8) due to the measurements.

4. air number density influence effect on Rayleigh extinction
5. ozone absorption cross section
6. ozone concentration effect on transmission
7. seed (tie-on) pressure
8. acceleration due to gravity
9. mean molecular mass variations with height above 80 km.

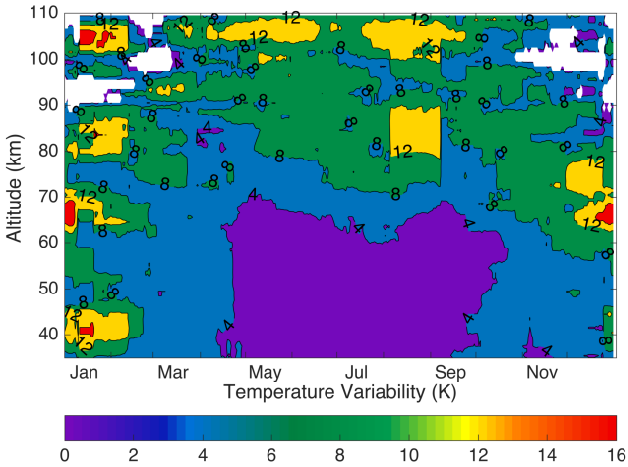

Figure 4. Geophysical variability in temperature for the OEM PCL climatology.

3.1.1 Uncertainty budget for the PCL climatology

A typical case for the temperature statistical and systematic uncertainties for a nightly average retrieval is shown in Fig. 5. The temperature uncertainty due to the seed pressure has the highest contribution among all of the systematic uncertainties at the altitudes above the mesopause. However, temperature uncertainties related to ozone, including the ozone absorption cross section, have the largest effect below 40 km. The uncertainty contribution for the gravity model is almost constant with height and is on the order of 0.002 K.

The nightly OEM statistical uncertainty profiles were used to form the statistical temperature uncertainty of the PCL temperature climatology (Fig. 6) using the procedure described in AS2007. The statistical uncertainty below 75 km is less than 1 K and gradually increases with height until it reaches 0.9 **Au** where it is less than 10 K. The monthly average minimum, maximum and median of temperature uncertainties related to the systematic uncertainties for all months are plotted in Fig. 7.

An improvement of the OEM over the HC method is its ability to yield the vertical resolution at each height (Fig. 8). The vertical resolution is 1056 m below 95 km and is equal to the retrieval grid. It is almost less than 3000 m below the 0.9 cutoff height, however, it increases rapidly to 5000 m around the 0.8 cutoff line. Leblanc et al. (2016b) recommended two standardized definitions for a temperature profile vertical resolution. In order to compare the retrieved temperature profiles using the OEM and the HC method, the two vertical resolution definitions given by Leblanc et al. (2016b) were used to find the best bin size for the HC method so it would have an identical vertical resolution to the OEM retrieval. We found that 264 m co-added bins and a 3's and 5's filter gave a vertical resolution of 1008 m, close to the OEM temperature retrieval grid (1056 m).

3.1.2 Comparison with uncertainty budget of the traditional method

Leblanc et al. (2016b), hereafter NDACC2016, used a Monte Carlo method to calculate the statistical and systematic uncertainties for the temperature retrieval. We have compared our results with his ND:YAG 532 nm lidar results. NDACC2016 and our

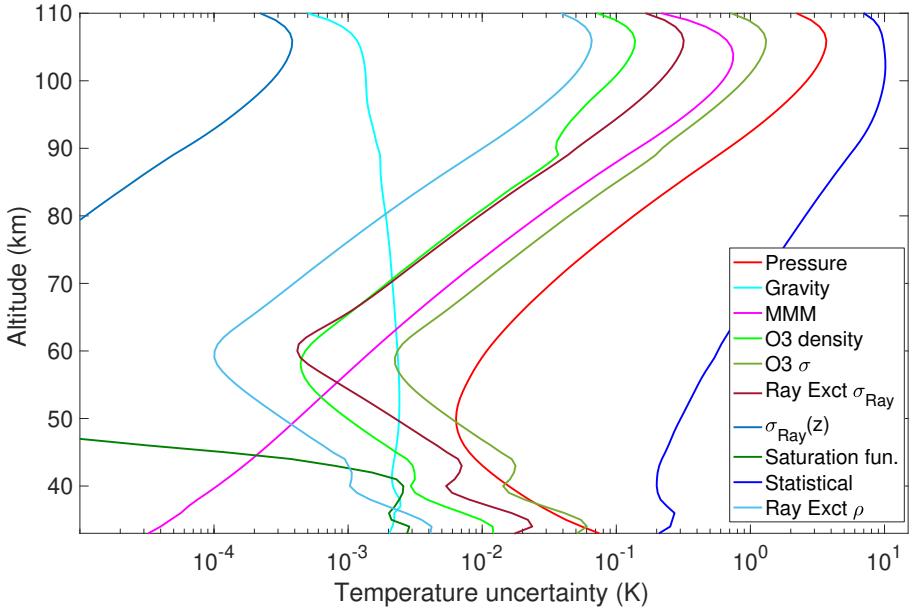

Figure 5. A typical night's systematic and random uncertainties for the OEM temperature retrieval.

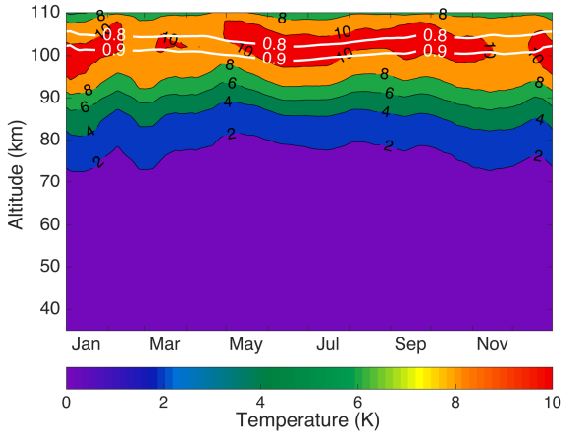

Figure 6. Statistical temperature uncertainty of the temperature climatology. The white lines are the height above below which the temperature climatology is more than 90% (0.9) and 80% (0.8) due to the measurements.

climatology give the temperature uncertainties for several of the same parameters (Table 3), including the statistical uncertainty (detection noise), the Rayleigh cross section, air number density, ozone absorption cross section, ozone number density, and the gravity model. NDACC2016 calculated the temperature uncertainty due to each parameter per 1% uncertainty. In order to compare NDACC2016 results with the PCL uncertainties using the OEM, we need to scale NDACC2016 simulations to the PCL as recommended by Leblanc et al. (2016b). For example, if the temperature uncertainty due to air number density

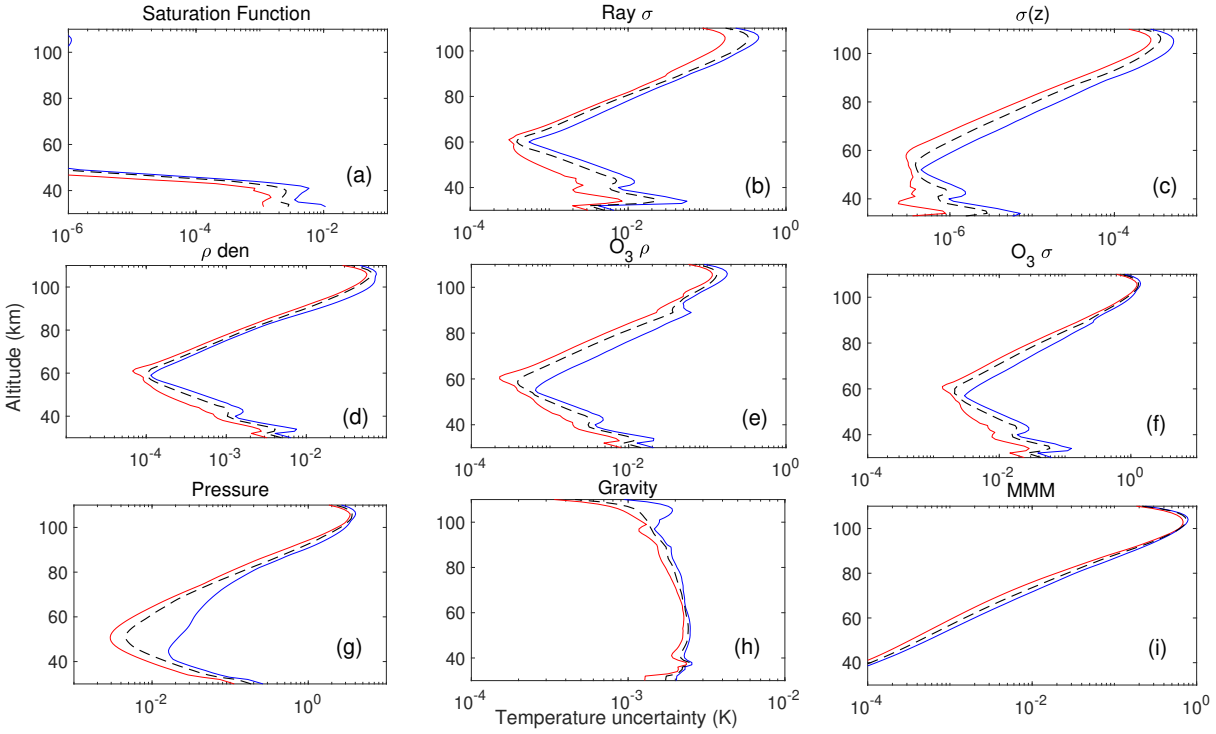

Figure 7. PCL temperature systematic uncertainty due to the a) saturation function (1994 to 1998 only), b) Rayleigh extinction cross section, c) Rayleigh cross section variation with height, d) air density affect on Rayleigh extinction, e) ozone absorption cross section f) ozone concentration, g) seed (tie-on) pressure, h) gravity model, i) Mean molecular mass variation with height. In each figure, red, blue, and black lines are the minimum, maximum and median between all months respectively.

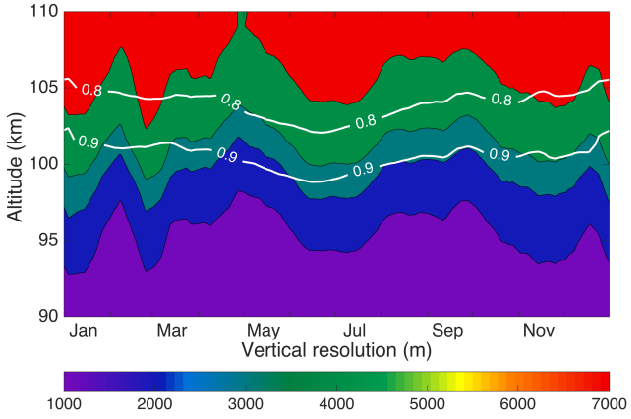

Figure 8. The OEM vertical resolution. The vertical resolution below 80 km is 1056 m, that is it is equal to the retrieval grid spacing (not shown). The white lines are the height above below which the temperature climatology is more than 90% (0.9) and 80% (0.8) due to the measurements.

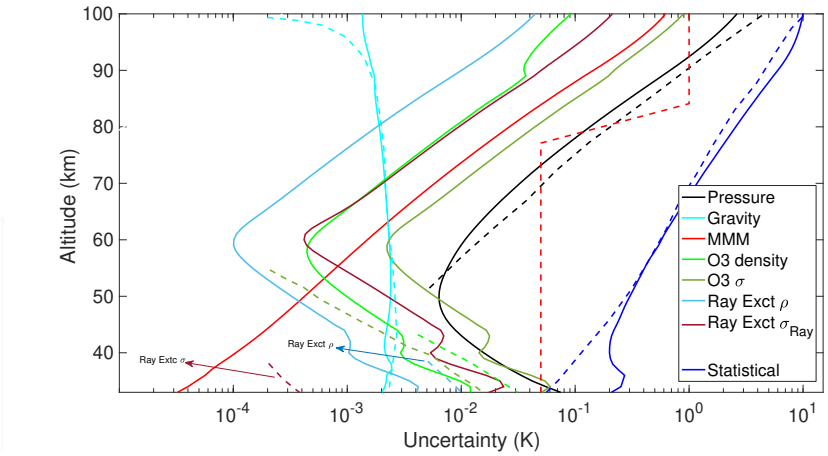

Figure 9. Comparison of the PCL statistical and systematic uncertainties with scaled uncertainties from Leblanc et al. (2016b) as described in the text. The solid lines are the uncertainties due to the PCL and the dashed lines are uncertainties due to NDACC2016.

is per 1% uncertainty in NDACC2016, then we must dividemultiply NDACC2016 uncertainties by a factor of 5 because we assume an air number density uncertainty of 5% (recommended by NDACC2016) in the PCL forward model (Table 3). We have compared our results with the statistical and systematic uncertainties presented in Figures 1 to 9 in NDACC2016 for the case of a 532 nm laser beam with a 1 MHz count rate at 45 km, a height resolution of 300 m, and an integration time of 2 hours (Fig. 9).

The statistical uncertainty comparison between the PCL and NDACC2016 is shown in dark blue in Fig. 9. It can be seen that the NDACC2016 statistical uncertainty almost equals the scaled PCL statistical uncertainty above the stratopause. However, there is a difference at altitudes below 50 km. The statistical uncertainty difference in the lower altitudes is due to using the two Rayleigh channel measurements (HLR and LLR) to calculate the temperature in the lower altitudes. The uncertainties at these altitudes are then a combination of the LLR and HLR uncertainties.

The temperature uncertainty due to the uncertainty in the Rayleigh cross section in NDACC2016 for each 1% at two sample altitudes, 30 and 38 km, is on the order of 0.001 K (NDACC2016, Figure 4). The temperature uncertainty due to the Rayleigh cross section in the OEM is presented per 0.2%, therefore, the scaled cross section uncertainty for NDACC2016 is one order of magnitude smaller than the PCL Rayleigh cross section uncertainty. However, this temperature uncertainty is very small.

The uncertainty due to air number density as an input quantity per 1% is shown in NDACC2016 (their Figure 5 left panel). The NDACC2016 scaled temperature uncertainty due to air number density is almost equal toin same order of magnitude as the OEM-derived uncertainty for the PCL (Fig. 9).

The standard deviation for the ozone cross section in the OEM forward model is 2%. Therefore, the NDACC2016 ozone cross section temperature uncertainties should be doubled to compare them with the PCL. The temperature uncertainty due to the ozone cross-section uncertainty in NDACC2016 (their Figure 6 left) after scaling is about four times smaller. The other temperature uncertainty due to ozone is the ozone number density. The temperature uncertainty due to ozone number density

uncertainty for the NDACC (their Figure 7 left), after scaling by a factor of 0.25 (as the PCL *a priori* assumes 4% uncertainty), is almost twice that of the PCL's. The uncertainties due to ozone number density are so small above 45 km that they have not been listed in the total uncertainty budget in NDACC2016's final results.

The temperature uncertainty due to the choice of pressure at the highest altitude (seed pressure) is called the tie-on uncertainty in NDACC2016. The tie-on uncertainties are in the same range and the small differences between the PCL and NDACC2016 (their Figure 8) are related to the fact that the seed pressure altitude is at 99 km for the NDACC2016 and at 110 km for the PCL.

The gravity temperature uncertainties for both NDACC2016 and the scaled PCL are consistent and are roughly 0.002 K. NDACC2016 states that the temperature uncertainty due to the molecular mass is negligible below 85 km and is on the order of 0.05 K and above 85 km can increase up to 1 K (NDACC2016, Table 3). The OEM shows that the PCL molecular mass temperature uncertainty at 85 km is 0.06 K. The PCL molecular mass temperature uncertainties from 90 to 100 km are between 0.1 and 0.6 K. However, the semi-empirical mean molecular mass variation of the US Standard model is considerably different from the variation assumed by NDACC2016, accounting for the differences in the calculated uncertainties.

4 Comparison of the OEM climatology with other climatologies

In order to evaluate the OEM results, the new OEM PCL temperature climatology was compared with the existing PCL temperature climatology using the HC method, as well as other climatologies including sodium lidar climatologies.

4.1 Comparison between the PCL climatology using the OEM and HC methods

AS2007 used PCL measurements between 1994 and 2004 to calculate a PCL temperature climatology (henceforth, 2004 PCL climatology) using the HC method. The top 10 km of all temperature profiles were removed from the 2004 PCL climatology in order to reduce the effect of seed pressure and the same procedure was followed in the HC calculations for the updated PCL climatology (between 1994 and 2013). The temperature differences between the OEM and updated HC PCL temperature climatologies are shown in Fig. 10. The white space in the upper part of Fig. 10 is due to removing 10 km from the top of each profile for the updated HC PCL climatology. In addition, the lines corresponding to the 10 and 15 km cutoff for the HC method and 0.9 cutoff line for the OEM are superimposed onto Fig. 10.

The OEM temperature climatology is 0.55 ± 0.23 K warmer than the updated HC climatology average from 40 to 60 km. Although the difference is within the statistical uncertainty of the measurements (Fig. 5), there is a warm bias. 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. Dead time correction, particularly in the upper stratosphere is even smaller and not likely contributing significantly. 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

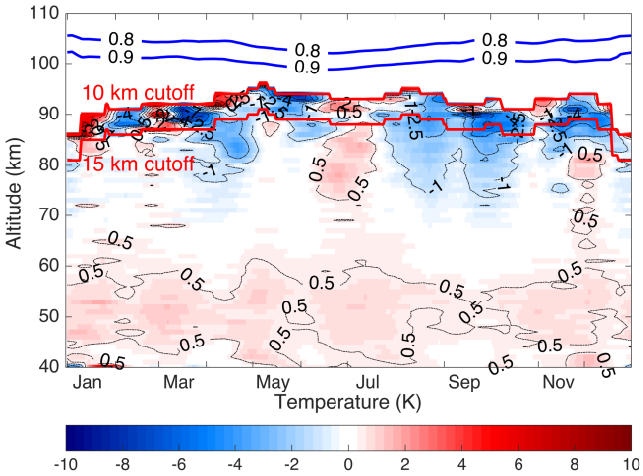

Figure 10. PCL temperature climatology difference between the OEM and HC method using seed pressure. The blue lines show the height above below which the OEM temperature climatology is more than 90% (0.9) and 80% (0.8) due to the measurements. The red lines are the 10 and 15 km cutoff height for the HC method.

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 direct comparison with AS2007.

The OEM temperature above 80 km up to the 10 and 15 km cut-offs is colder than the temperatures obtained using the HC method. The temperature differences above 80 km are mostly due to the sensitivity of the model seed pressure in the HC method. Figure 10 shows that the OEM temperature climatology reaches 5 to 10 km higher in altitude than the HC temperature climatology. The differences between the OEM and HLR are not calculated below 40 km due to the lack of HLR data in the winter months.

Finally, in order to evaluate the effect of the *a priori* on the temperature differences, the same temperature climatologies were calculated using the OEM with the US standard model as the *a priori* temperature profile and the same differences as discussed above were obtained, again demonstrating that the results show little sensitivity to the choice of any reasonable *a priori* profile.

The HC method usually uses a seed pressure value at the highest point of the profile. However, the seed pressure can be substituted by temperature and density and is called the seed temperature (Gardner et al. (1989), equation 86). When a seed temperature is used, the temperature is obtained from the CIRA-86 model, and the measured relative density profile is normalized (typically by a model) to obtain a seed pressure to use in the HC retrieval. The temperature differences between the OEM climatology and the updated HC climatology using the seed temperature (instead of seed pressure) are shown in Fig. 11.

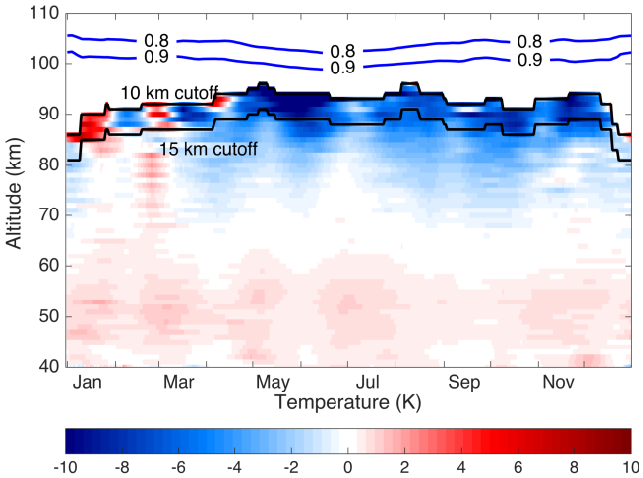

Figure 11. PCL temperature climatology difference using the OEM and HC method using seed temperature. The blue lines show the height above below which the OEM temperature climatology is more than 90% (0.9) and 80% (0.8) due to the measurements. The black lines are the 10 and 15 km cutoff height for the HC method.

Comparing Figs. 10 and 11 reveals that the temperature difference above 80 km between the OEM and the updated HC using seed temperatures is larger than the differences between the OEM and the updated HC using seed pressures. However, the differences below 80 km are identical and the small temperature differences between the OEM and HC method are due to the tie-on temperature or pressure value. The difference between the HC climatologies calculated by these two methods highlights the sensitivity to seed pressure at the greatest heights in this method.

4.2 Comparison with sodium lidar climatologies

The comparison between the PCL Rayleigh temperature climatology using the HC method with sodium lidars was done by AS2007. Their results showed that the average temperature between 83 and 95 km measured by the PCL was between 7 and 7.4 K colder than CSU and URB climatologies, respectively. 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 in the region where the sodium lidars are most sensitive, 90 to 100 km. Sodium lidars directly measure the kinetic temperature without assuming hydrostatic equilibrium or requiring the knowledge of mean molecular mass and molecular cross section variations with height and can be configured to obtain temperatures during both the day and night. She et al. (2000), Yuan et al. (2008), as well as States and Gardner (2000a) have published sodium temperature lidar climatologies in the same latitude range as PCL. Both sites are west of the PCL, but in the case of URB the separation in longitude is less than 8°. 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 AS2007. The upgraded CSU (Yuan et al., 2008) provides additional years of overlap with our new climatology for validation of our OEM-derived temperatures. The nighttime URB and upgraded CSU temperature climatologies were compared with the PCL temperature climatology.

Table 4. Absolute value of the average PCL temperature differences with sodium lidars as well as temperature difference between sodium lidars. The HC method does not provide the temperature above 95 km, therefore the columns with a altitude range greater than 95 km are shown as '-'.

Lidars	Difference (K) 80-85 km	85-90 km	90-95 km	**Au** = 0.9 95-100 km	**Au** = 0.8 100-105 km
PCL(OEM) - URB	11.3	6.0	4.4	3.9	8.3
PCL(HC) - URB	12.8	8.1	6.7	-	-
PCL(OEM) - CSU	-	6.9	5.1	6.6	14.2
PCL(HC) - CSU	-	8.4	6.2	-	-
PCL(OEM) - upgraded CSU	5.6	4.1	3.8	7.8	13.5
PCL(HC) - upgraded CSU	6.7	4.7	3.4	-	-
CSU - URB	-	4.5	3.8	5.1	6.7
CSU - upgraded CSU	-	4.4	4.0	3.2	3.2
URB - upgraded CSU	7.3	4.6	5.7	7.1	5.6

The PCL temperature climatology differences using the OEM compared with the sodium lidars are presented in Figures 12, 13 and 14. The absolute value of the average differences in 5 km height bins between the sodium lidar temperature climatologies and the PCL climatology using the OEM and the HC method are given in Table 4. The absolute value is used to avoid differences cancelling each other. The bottom part of the Table is important, as it gives the differences between the sodium lidars themselves. The differences between the sodium lidars are taken as the level of difference defining agreement between the PCL lidar and the sodium systems. The PCL HC climatology in general does not agree with the sodium lidar climatologies to the same amount to which they agree with each other, while the PCL OEM climatology typically does agree with the sodium lidar climatologies to the level at which they agree with each other. The temperature differences between the PCL OEM and sodium lidar climatologies for the entire range of altitudes (80 to 105 km) are smaller than the temperature difference between the PCL HC climatology and the sodium lidar climatologies in the range of 80 to 95 km for which PCL HC temperatures are available. There is a temperature difference at the winter mesopause between the PCL climatology and CSU climatology, but this difference has decreased in the upgraded nighttime CSU climatology compared to that determined by AS2007. The large temperature differences between the PCL (OEM) and URB temperature climatology during summertime below 85 km existed in AS2007 and may be in part due to the signal-to-noise ratio of the sodium lidar measurements rapidly decreasing below 85 km.

Possible sources of these differences were addressed in AS2007, but they did not have the uncertainty budget now available to assess systematic uncertainties. Possible sources of temperature difference between the PCL OEM and sodium lidar climatologies are as listed below.

1. The assumption of a seed pressure can introduce uncertainty in the PCL temperature retrievals. Using an OEM allows us to calculate the effect of this assumption quantitatively (Fig. 7). In the altitude range of 80 to 95 km, it is less than 1.5 K, increasing to a maximum of 3.5 K at 100 km.

2. The effects of Rayleigh scatter cross section, Rayleigh scatter density and mean molecular mass were mentioned in AS2007 as possible reasons for discrepancies with the sodium lidar temperatures. Figure 7 shows a quantitative determination of the magnitude of these effects. The uncertainties for the Rayleigh scatter cross section and Rayleigh scatter density are much less than the temperature differences between the 2 measurement techniques. Mean molecular mass uncertainty is larger than the other two parameters, but its maximum value is less than 0.7 K at 105 km.

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. 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.

4. Geographic location could be another possible cause. The PCL is about 3 ° north of the sodium lidars and, while relatively close to URB in longitude, the PCL is 24 ° east of CSU. Hence, tides and planetary waves could be the primary cause of the temperature differences between the PCL, URB and CSU lidars. Gravity waves could also contribute, although the effect of gravity waves is minimized by averaging temperature over several hours and using common days at different years to calculate the composite climatology. Sica and Argall (2007) have shown that the seasonal gravity wave activity over London, Canada is large and highly variable, and is possibly related to London's proximity to both Lake Ontario to the west and Lake Erie to the east. The effect of solar tides on the sodium lidar temperature is discussed in States and Gardner (2000b) and Yuan et al. (2006). The upgraded CSU is capable of continuous observation during day and night. Yuan et al. (2008) removed tidal signals from the mean values and calculated diurnal mean monthly temperatures. They show that the amplitude of the diurnal tide is around 5 K at night between 84 and 95 km, increasing to 8 K at 100 km. Hence, we conclude that large-scale waves cause much of the discrepancies between locations.

The comparison with sodium lidars shows that the PCL Rayleigh temperature climatology using the OEM in general agrees as well with the sodium lidar climatologies as the sodium climatologies agree with one another, validating the PCL OEM height-extended climatology.

5 Summary

Here we have confirmed the validity of using the OEM to retrieve Rayleigh-scatter lidar temperatures on a long-term measurement set. The results of our investigation using the OEM on 519 nights of measurements are summarized as follows.

1. Our OEM can estimate a valid cutoff height where the entire temperature profile below that level depends less than a specified level on the choice of the *a priori* temperature profile. Based on best practice in the OEM community, we suggest using measurements whose summed averaging kernels at a retrieval altitude are greater than 0.9.

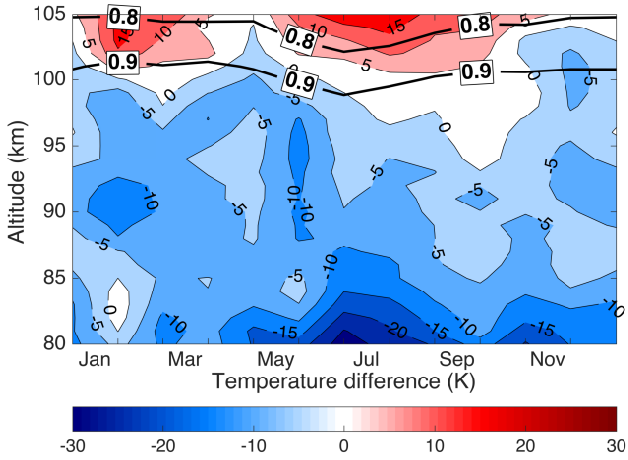

Figure 12. PCL temperature climatology difference from the URB sodium lidar climatology. The horizontal black lines are the height above below which the temperature climatology is more than 90% (0.9) and 80% (0.8) due to the measurements.

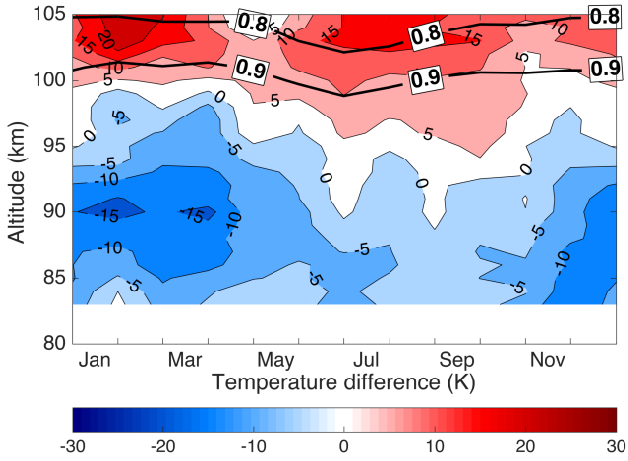

Figure 13. PCL temperature climatology difference from the CSU (1990-1999) sodium lidar climatology. The horizontal black lines are the height above below which the temperature climatology is more than 90% (0.9) and 80% (0.8) due to the measurements.

2. The effect of the temperature *a priori* on the OEM result was evaluated using the CIRA-86 and US standard model. It was shown that the effect of the *a priori* is much smaller than the statistical uncertainty below the OEM cutoff heights for the PCL.

3. We presented a full uncertainty budget for our climatology which includes both random and systematic uncertainties, including the systematic uncertainty for 9 model parameters including mean molecular mass, Rayleigh cross section, Rayleigh cross section variation with composition, seed pressure, air number density (for extinction), ozone absorption cross section, ozone density and acceleration due to gravity. This uncertainty budget is available on a profile-by-profile basis.

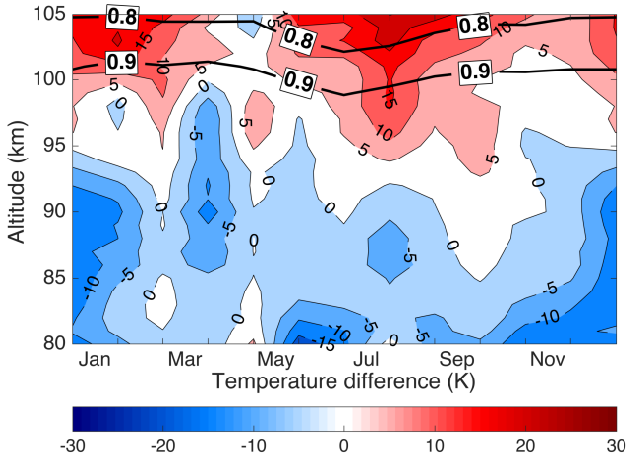

Figure 14. PCL temperature climatology difference from the upgraded CSU (2002-2006) sodium lidar climatology. The horizontal black lines are the height above below which the temperature climatology is more than 90% (0.9) and 80% (0.8) due to the measurements.

4. The PCL uncertainties were compared to the uncertainty budget simulations presented by Leblanc et al. (2016b). The comparison shows in general similar orders of magnitude, except for Rayleigh-scatter cross section which has a larger difference but makes a very small (0.001 K) contribution to the uncertainty budget.

5. Our OEM computes the vertical resolution of each temperature profile. The vertical resolution is equal to the retrieval grid (1056 m) until about 75 km, where it starts to increase and is about 3 km around the 0.9 cutoff height.

6. The PCL temperature climatology is calculated using both the OEM and the HC method. By 15 km below the cutoff height, any differences in the temperature are within the statistical uncertainty at those heights. Our OEM retrieval determines temperature profiles which reach 5 to 10 km higher than the temperature profiles calculated by the HC method, due to the OEM's ability to evaluate the effect of seed pressure on the retrieved temperature.

7. The temperature difference between the OEM PCL temperature climatology with the HC method PCL climatology using seed pressure was smaller than the temperature difference between the OEM PCL temperature climatology with the HC method temperature climatology using seed temperature. Hence, we recommend that when using the HC method it is better to take the seed pressure from the model than a seed temperature.

8. The PCL temperature climatology is compared with three other sodium lidar climatologies. The temperature differences between the PCL climatology using the OEM and the sodium lidar climatologies are smaller by 1 K than the differences between the PCL-OEM and the PCL-HC differences. The temperature differences between the PCL-OEM and the sodium lidars are within the temperature differences between the sodium lidars themselves (Table 4). The OEM provides the PCL temperature profiles to higher altitudes and these profiles show smaller differences with the sodium lidars than the HC method and thus, using the OEM improves the climatology between 80 and 100 km, as validated by the sodium lidar measurements.

9. The statistical uncertainty of the sodium lidar temperatures is lowest in the 95 ± 5 km region of the peak of the sodium layer. Here the precision is about 1 K to 2 K (Papen et al., 1995). The accuracy of the measurement in this region has been studied in detail by Krueger et al. (2015), who obtain an accuracy of 1 to 2.5 K. The statistical uncertainty increases rapidly away from the sodium layer peak. The closest agreement between the PCL temperature climatology and the sodium lidars' climatology is in the range of 85 to 100 km, with larger temperature differences below 85 km and above 100 km where the sodium density is lowest. The URB climatology, which was obtained from a station much closer in longitude to the PCL, shows better agreement than the CSU measurements, although all 3 sodium lidar climatologies have overall good agreement with the PCL OEM climatology. Overall the OEM provides closer temperature results to the sodium lidars than the HC method at all heights, and allows the climatology to extend to a greater altitude.

6 Conclusions

We have shown that using the OEM to retrieve temperature from Rayleigh-scatter lidar measurements has significant advantages over the traditional method, and the advantages shown in our initial study for a small number of nights is practical for a large data set. These advantages include the ability to calculate a full uncertainty budget on a profile-by-profile basis, determination of the vertical resolution, and the availability of averaging kernels. Applying the OEM will help in the standardization of uncertainty budget and vertical resolution calculations for comparisons between lidars, as well as comparisons among other instruments with differing vertical resolutions.

We found that a cutoff height of $\mathbf{Au} = 0.9$ is a good estimate for a cutoff height of the retrieval, based on the comparison with the sodium lidars. It would be recommended to use the 0.9 height cutoff to minimize the effect of the *a priori* on the temperature retrieval while keeping the *a priori* effect on the temperature retrieval less than the statistical uncertainty.

Sodium lidars are well characterized and make the best temperature measurements in the mesosphere and lower thermosphere for validation of the PCL temperature climatology, particularly as the URB and CSU systems are relatively near the PCL. The agreement between the OEM-based PCL climatology and the sodium lidars has improved over the traditional method, and the agreement between the PCL and the sodium lidars is typically as good as the agreement between the sodium lidars themselves. Much of the variability seen in the measurements made at the different locations is likely due to tides and planetary waves.

We hope the results of this study encourage other Rayleigh-lidar groups to process their measurements using our OEM retrieval method.

Author contributions. Ali Jalali was responsible for the data analysis, developing the code to form the climatologies, and manuscript preparation. He also was involved in operating the lidar after July 2012. This work forms part of his doctoral thesis. R. J. Sica was responsible for the supervision of the doctoral thesis, contributions to manuscript preparation, coding the OEM temperature retrieval and, in collaboration with A. Haefele, first applying the OEM to Rayleigh-lidar temperature retrievals. Alexander Haefele helped in manuscript preparation and many OEM and scientific discussions relevant to this work.

Acknowledgements. This project has been funded in part by Discovery Grants and a CREATE Training Program in Arctic Atmospheric Science (PI K. Strong) from the National Science and Engineering Research Council of Canada and awards from the Canadian Foundation for Climate and Atmospheric Science. We would like to thank the Federal Office of Meteorology and Climatology, MeteoSwiss, for its support of this project. Ali Jalali would like to thank Shannon Hicks-Jalali for her comments and suggestions on the manuscript. We would like to thank Stephen Argall for his many contributions to the PCL lidar program and to Tao Yan for providing to us in a digital format the upgraded CSU nighttime sodium temperature climatology.

References

Argall, P. S. and Sica, R. J.: A comparison of Rayleigh and sodium lidar temperature climatologies, Ann. Geophys., 25, 27–35, 2007.

Argall, P. S., Vassiliev, O. N., Sica, R. J., and Mwangi, M. M.: Lidar measurements taken with a large-aperture liquid mirror: 2. The Sodium resonance-fluorescence system, Appl. Optics, 39, 2393–2399, 2000.

Arnold, K. S. and She, C. Y.: Metal fluorescence lidar (light detection and ranging) and the middle atmosphere, Contemp. Phys., 44, 35–49, 2003.

Bills, R. E., Gardner, C. S., and She, C. Y.: Narrow band lidar technique for sodium temperature and Doppler wind observations of the upper atmosphere, Opt. Eng., 30, 1991.

Fleming, E. L., Chandra, S., Shoeberl, M. R., and Barnett, J. J.: Monthly Mean Global Climatology of Temperature, Wind, Geopotential Height and Pressure for 0-120 km, NASA Tech. Memo., NASA TM100697, 85 pp, 1988.

Gardner, C. S., Senft, D. C., Beatty, T. J., Bills, R. E., and Hostetler, C. A.: Rayleigh and sodium lidar techniques for measuring middle atmosphere density, temperature, and wind perturbations and their spectra, In World Ionosphere/ Thermosphere Study Handbook, 2, 141–187, 1989.

Griggs, M.: Absorption coefficients of ozone in the ultraviolet and visible regions, J. Chem. Phys., 49, 857–859, 1968.

Gross, M. R., McGee, T. J., Ferrare, R. A., Singh, U. N., and Kimvilakani, P.: Temperature measurements made with a combined Rayleigh–Mie and Raman lidar, Appl. Opt., 36, 5987–5995, 1997.

Hamming, R. W.: Digital Filters, Prentice Hall, Englewood Cliffs, New Jersey, third edn., 1989.

Hauchecorne, A. and Chanin, M.: Density and temperature profiles obtained by lidar between 35 and 70 km , Geophys. Res. Lett., 7, 565–568, 1980.

Jalali, A.: Extending and Merging the Purple Crow Lidar Temperature Rayleigh and Vibrational Raman Climatologies, Master's thesis, the University of Western Ontario, 2490, 2014.

Khanna, J., Bandoro, J., Sica, R. J., and McElroy, C. T.: New technique for retrieval of atmospheric temperature profiles from Rayleigh-scatter lidar measurements using nonlinear inversion, Appl. Optics, 51, 7945–52, 2012.

Krueger, D. A., She, C.-Y., and Yuan, T.: Retrieving mesopause temperature and line-of-sight wind from full-diurnal-cycle Na lidar observations, Appl. Opt., 54, 9469–9489, 2015.

Leblanc, T., McDermid, I. S., Keckhut, P., Hauchecorne, A., She, C. Y., and Krueger, D. A.: Temperature climatology of the middle atmosphere from long-term lidar measurements at middle and low latitudes, J. Geophys. Res.-Atmos., 103, 17 191–17 204, 1998.

Leblanc, T., Sica, R. J., van Gijsel, J. A. E., Godin-Beekman, S., Haefele, A., Trickl, T., Payen, G., and Gabarrot, F.: Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms – Part 1: Vertical resolution, Atmos. Meas. Tech., 9, 4029–4049, 2016a.

Leblanc, T., Sica, R. J., van Gijsel, J. A. E., Haefele, A., Payen, G., and Liberti, G.: Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms – Part 3: Temperature uncertainty budget, Atmos. Meas. Tech., 9, 4079–4101, 2016b.

McPeters, R. D., Labow, G. J., and Logan, J. A.: Ozone climatological profiles for satellite retrieval algorithms, J. Geophys. Res. Atmos., 112, 2007.

Mulaire, W.: Department of Defense World Geodetic System 1984. Its definition and relationship with local geodetic systems, NIMA TR8350.2, p. 1–175, 2000.

Nicolet, M.: On the molecular scattering in the terrestrial atmosphere: an empirical formula for its calculation in the homosphere, Planet. Space Sci., 32, 1467–1468, 1984.

Papen, G. C., Pfenninger, W. M., and Simonich, D. M.: Sensitivity analysis of Na narrowband wind–temperature lidar systems, Appl. Opt., 34, 480–498, 1995.

Ramaswamy, V., Chanin, M.-L., Angell, J., Barnett, J., Gaffen, D., Gelman, M., Keckhut, P., Koshelkov, Y., Labitzke, K., Lin, J.-J. R., O'Neill, A., Nash, J., Randel, W., Rood, R., Shine, K., Shiotani, M., and Swinbank, R.: Stratospheric temperature trends: Observations and model simulations, Rev. Geophys., 39, 71–122, 2001.

Randel, W., Udelhofen, P., Fleming, E., Geller, M., Gelman, M., Hamilton, K., Karoly, D., Ortland, D., Pawson, S., Swinbank, R., Wu, F., Baldwin, M., Chanin, M.-L., Keckhut, P., Labitzke, K., Remsberg, E., Simmons, A., and Wu, D.: The SPARC Intercomparison of Middle-Atmosphere Climatologies, J. Climate, 17, 986–1003, 2004.

Randel, W. J., Shine, K. P., Austin, J., Barnett, J., Claud, C., Gillett, N. P., Keckhut, P., Langematz, U., Lin, R., Long, C., Mears, C., Miller, A., Nash, J., Seidel, D. J., Thompson, D. W. J., Wu, F., and Yoden, S.: An update of observed stratospheric temperature trends, J. Geophys. Res.-Atmos., 114, d02107, 2009.

Randel, W. J., Smith, A. K., Wu, F., Zou, C.-Z., and Qian, H.: Stratospheric Temperature Trends over 1979–2015 Derived from Combined SSU, MLS, and SABER Satellite Observations, J. Climate, 29, 4843–4859, 2016.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, vol. 2, World Scientific, Hackensack, NJ, USA, 2011.

She, C. Y., Chen, S., Hu, Z., Sherman, J., Vance, J. D., Vasoli, V., White, M. A., Yu, J., and Krueger, D. A.: Eight-year climatology of nocturnal temperature and sodium density in the mesopause region (80 to 105 km) over Fort Collins (41N, 105W), J. Geophys. Res., 27, 3289–3292, 2000.

Sica, R. J. and Argall, P. S.: Seasonal and nightly variations of gravity-wave energy density in the middle atmosphere measured by the Purple Crow Lidar, Ann. Geophys., 25, 2139–2145, 2007.

Sica, R. J. and Haefele, A.: Retrieval of temperature from a multiple-channel Rayleigh-scatter lidar using an optimal estimation method, Appl. Optics, 54, 1872–1889, 2015.

Sica, R. J., Sargoytchev, S., Argall, P. S., Borra, E. F., Girard, L., Sparrow, C. T., and Flatt, S.: Lidar measurements taken with a large aperture liquid mirror. 1. Rayleigh scatter system , Appl. Optics, 34, 6925–6936, 1995.

Sica, R. J., Argall, P. S., Russell, A. T., Bryant, C. R., and Mwangi, M. M.: Dynamics and Composition Measurements in the Lower and Middle Atmosphere with the Purple Crow Lidar, Recent Research Developments in Geophysical Research, 3, 1–16, 2000.

States, R. J. and Gardner, C. S.: Thermal structure of the mesopause region (80-105km) at 40 degrees N latitude. Part I: Seasonal variations, J. Atmos. Sci., 57, 66–77, 2000a.

States, R. J. and Gardner, C. S.: Thermal Structure of the Mesopause Region (80–105 km) at 40°N Latitude. Part II: Diurnal Variations, J. Atmos. Sci., 57, 78–92, 2000b.

Yuan, T., She, C. Y., Hagan, M. E., Williams, B. P., Li, T., Arnold, K., Kawahara, T. D., Acott, P. E., Vance, J. D., Krueger, D., and Roble, R. G.: Seasonal variation of diurnal perturbations in mesopause region temperature, zonal, and meridional winds above Fort Collins, Colorado (40.6°N, 105°W), J. Geophys. Res.-Atmos., 111, d06103, 2006.

Yuan, T., She, C.-Y., Krueger, D. A., Sassi, F., Garcia, R., Roble, R. G., Liu, H.-L., and Schmidt, H.: Climatology of mesopause region temperature, zonal wind, and meridional wind over Fort Collins, Colorado (41N, 105W), and comparison with model simulations, J. Geophys. Res., 113, 2008.

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)
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

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.

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.

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.

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:

$$S_f = \underbrace{G_y}_{\text{Section 1}} \underbrace{K_b \; S_b \; K_b^T}_{\text{Section 2}} \underbrace{G_y^T}_{\text{Section 1}}$$

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 \pm0.05 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.