

Extinction and Optical Depth Retrievals for CALIPSO's Version-4 Data Release: Supplementary Material

Stuart A. Young¹, Mark A. Vaughan², Anne Garnier¹, Jason L. Tackett¹, James B. Lambeth¹, Kathleen A. Powell²

5 ¹Science Systems and Applications, Inc. (SSAI), Hampton, VA 23666, USA

²NASA Langley Research Center, Hampton, VA 23681, USA

Correspondence to: Stuart A. Young (stuart.young01@gmail.com)

S1 Testing for the existence of backscatter and backscatter uncertainty solutions

The symbols in this supplement are the same as in Young and Vaughan (2009) and Young et al. (2013) where $\delta r(r)$, $\beta'(r)$,
10 $\beta(r)$ and $\sigma(r)$ are, respectively, the range increment, attenuated backscatter, backscatter and extinction coefficients at range r ,
 $\chi(r_N, r)$ and $\tau(r_N, r)$ are, respectively, the integrals of backscatter and extinction over the range interval from the normalization
range r_N to r , and $T^2(r_N, r)$ is the two-way transmittance over the same interval. S is the lidar (extinction-to-backscatter) ratio
and the subscripts M , P and T designate the molecular, particulate and total components. The subscript N indicates that the
attenuated backscatter has been rescaled to correct for overlying molecular and retrieved particulate attenuation between the
15 lidar and the normalization range. The prefix Δ indicates the absolute uncertainty in the quantity.

S1.1 Backscatter

Following the development given in Young and Vaughan (2009), the particulate backscatter coefficient at each range, r ,
within a feature is found by solving equation

20
$$\beta_p(r) = \frac{\beta'_N(r)}{T_M^2(r_N, r)T_P^2(r_N, r)} - \beta_M(r), \quad (\text{S1})$$

where

$$T_p^2(r_N, r) = \exp\left[-2\eta S_p \int_{r_N}^r \beta_p(z) dz\right] = \exp[-2\eta \tau_p(r_N, r)]. \quad (\text{S2})$$

We choose to factor the transmittance in the following way.

$$\begin{aligned}
T_P^2(r_N, r) &= \exp \left\{ - \left[2\eta S_P \int_{r_N}^{r-\delta r} \beta_P(z) dz \right] - \eta S_P \delta r [\beta_P(r-\delta r) + \beta_P(r)] \right\} \\
&= T_P^{2*}(r_N, r) \exp[-\eta S_P \delta r \beta_P(r)],
\end{aligned} \tag{S3}$$

which permits Eq. (S1) to be rewritten as

$$\beta_P(r) = \frac{\beta'_N(r) \exp[\eta S_P \delta r \beta_P(r)]}{T_M^2(r_N, r) T_P^{2*}(r_N, r)} - \beta_M(r). \tag{S4}$$

This is of the form

$$x = a \exp(bx) - c, \tag{S5}$$

where $x = \beta_P(r)$ and a , b and c are constants at range r :

$$\begin{aligned}
a &= \frac{\beta'_N(r)}{T_M^2(r_N, r) T_P^{2*}(r_N, r)}, \\
b &= \eta S_P \delta r, \\
c &= \beta_M(r).
\end{aligned} \tag{S6}$$

We rearrange Eq. (S5) and use Newton's Method (Newton-Raphson) iteration to solve the equation

$$f(x) = a \exp(bx) - c - x = 0. \tag{S7}$$

10 (To ensure that the initial value in the iteration is near the final value, it is selected as described in Sect. S2.)

For solutions to Eq. (S7) to exist, $f(x)$ must have a minimum less than or equal to zero (for $a > 0$). The minimum occurs where

$$f'(x) = ab \exp(bx) - 1 = 0, \tag{S8}$$

i.e. at

$$15 \quad x_{\min} = \ln(1/ab) / b = -\ln(ab) / b, \tag{S9}$$

giving

$$f_{\min} = f(x_{\min}) = [1 + \ln(ab)] / b - c. \tag{S10}$$

For solutions to exist for Eq. (S7) then, $f_{\min} \leq 0$ and the following condition must hold:

$$\ln(ab) \leq cb - 1, \tag{S11}$$

20 or, writing out using the substitutions (S6),

$$\ln \left[\frac{\beta'_N(r)}{T_M^2(r_N, r) T_P^{2*}(r_N, r)} \eta S_P \delta r \right] \leq \eta S_P \delta r \beta_M(r) - 1. \quad (\text{S12})$$

It is apparent, then, that solutions may not exist for large values of S_P combined with large values of normalized backscatter signal, $\beta'_N(r)$.

During each attempt at retrieving $\beta(r)$, the magnitude of f_{\min} is evaluated. A value greater than zero indicates that no solution exists. This is interpreted as being caused by an overestimate of the lidar ratio, which is then reduced and the retrieval of the backscatter profile restarted from the top of the feature. The main caveat here is that incorrect rescaling of the normalized attenuated backscatter as a result of the use of a lidar ratio that is too large, but not large enough to cause failure, in the retrieval of the optical depth of an overlying layer may cause the retrieval in the lower layer to fail even if the lidar ratio in the lower layer is correct. Conversely, the use of a lidar ratio that is too small in the analysis of an overlying layer may allow the retrieval of backscatter profile in a lower layer even though the lidar ratio used in the lower layer is considerably larger than the correct value. Data users are advised to use caution when using retrievals in layers that underlie several other layers as errors in the retrieved optical depths in the overlying layers can cause errors to accumulate in the renormalization of the integrated attenuated backscatter in the lower layer. An indication of the potential severity of these renormalization errors can be gained by summing the uncertainties in the retrieved optical depths of the overlying layers. If the renormalization error is assumed to be a result solely of the error in the optical depth, which we approximate by its uncertainty, $\Delta\tau_P$, then the renormalization error (effectively the relative error in the overlying effective two-way transmittance) can be estimated as $\Delta T_P^2 / T_P^2 = \Delta\tau_P / (2\eta)$.

As stated in the CALIPSO layer properties user's guide (https://www-calipso.larc.nasa.gov/resources/calipso_users_guide/data_summaries/layer/index.php#layer_iab_qa_factor), the layer integrated attenuated backscatter QA factor (layer IAB QA factor) should also be considered. Values of this quantity that are close to unity are associated with little overlying attenuated backscatter, whereas lower values indicate the presence of greater quantities of scattering cloud and aerosol material and the increased likelihood of renormalization and, hence, retrieval errors. (See also the discussion in the user's guide under the headings Integrated Attenuated Backscatter 532 and Integrated Attenuated Backscatter Uncertainty 532.)

25 **S1.2 Backscatter uncertainty**

The existence of a solution for the uncertainty in the particulate backscatter can also indicate whether the lidar ratio used in the retrieval is too large or not. In the following uncertainty analysis, which was used in the creation of the version-4 data products, it will be assumed that uncertainties in the different quantities are random and uncorrelated. While this is not always strictly true, it is expected that most errors incurred in this approximation will be smaller than the main contributors to the overall uncertainties. The results of this simplified analysis are expected to give acceptable estimates of the

uncertainties for low to moderate optical depths and small values of uncertainties in the analysis parameters. An equation for the uncertainty in the retrieved particulate backscatter coefficient can be developed from Eq. (S1) as follows.

$$\begin{aligned}
(\Delta\beta_p(r))^2 &= (\Delta\beta_M(r))^2 + \beta_T^2(r) \left[\left(\frac{\Delta\beta'_N(r)}{\beta'_N(r)} \right)^2 + \left(\frac{\Delta T_M^2(r_N, r)}{T_M^2(r_N, r)} \right)^2 \right] + \beta_T^2(r) \left(\frac{\Delta T_P^2(r_N, r)}{T_P^2(r_N, r)} \right)^2, \\
&= A + \beta_T^2(r) [2\eta\tau_p(r_N, r)]^2 [\Delta\eta/\eta]^2 + \beta_T^2(r) [2\eta\tau_p(r_N, r)]^2 [\Delta\tau_p(r_N, r)/\tau_p(r_N, r)]^2, \\
&= A + \beta_T^2(r) [2\eta\tau_p(r_N, r)]^2 [\Delta\eta/\eta]^2 + \beta_T^2(r) [2\eta\tau_p(r_N, r)]^2 \left[\left(\frac{\Delta S_P}{S_P} \right)^2 + \left(\frac{\Delta\gamma_P(r_N, r)}{\gamma_P(r_N, r)} \right)^2 \right], \\
&= A + B + \beta_T^2(r) [2\eta\tau_p(r_N, r)]^2 \left(\frac{\Delta\gamma_P(r_N, r)}{\gamma_P(r_N, r)} \right)^2, \\
&= A + B + \beta_T^2(r) [\eta S_P]^2 \left\{ [\delta r(r_N) \Delta\beta_P(r_N)]^2 + \left[\sum_{i=jN+1}^{j-1} (\delta r_i + \delta r_{i+1})^2 [\Delta\beta_P(r_i)]^2 \right] \right\} \\
&\quad + \beta_T^2(r) [\eta S_P]^2 [\delta r(r) \Delta\beta_P(r)]^2.
\end{aligned} \tag{S13}$$

Equation (S13) can now be rearranged and solved for $(\Delta\beta_p(r))^2$ as

$$5 \quad (\Delta\beta_p(r))^2 = (A + B + C) / \left[1 - [\eta S_P \delta r(r) \beta_T(r)]^2 \right]. \tag{S14}$$

In the above we have made the following substitutions:

$$\begin{aligned}
A &= (\Delta\beta_M(r))^2 + \beta_T^2(r) \left[\left(\frac{\Delta\beta'_N(r)}{\beta'_N(r)} \right)^2 + \left(\frac{\Delta T_M^2(r_N, r)}{T_M^2(r_N, r)} \right)^2 \right], \\
B &= \beta_T^2(r) [2\eta\tau_p(r_N, r)]^2 \left[\left(\frac{\Delta\eta}{\eta} \right)^2 + \left(\frac{\Delta S_P}{S_P} \right)^2 \right], \\
C &= \beta_T^2(r) [\eta S_P]^2 \left\{ [\delta r(r_N) \Delta\beta_P(r_N)]^2 + \left[\sum_{i=jN+1}^{j-1} (\delta r_i + \delta r_{i+1})^2 [\Delta\beta_P(r_i)]^2 \right] \right\}.
\end{aligned} \tag{S15}$$

Finally, the uncertainty in the particulate extinction coefficient can be written as

$$10 \quad \Delta\sigma_p(r) = \left((\beta_p(r) \Delta S_P)^2 + (S_P \Delta\beta_p(r))^2 \right)^{1/2}. \tag{S16}$$

It can be seen from Eq. (S14) that a solution exists for $\Delta\beta_p(r)$ only if the denominator is greater than zero. (A, B and C are all non-negative.) This condition imposes an upper limit on the value of the particulate lidar ratio:

$$S_p < 1/[\eta\delta r(r)\beta_T(r)]. \quad (\text{S17})$$

In the V4 software used to calculate retrieval uncertainties, values of the denominator that are not greater than zero are interpreted as indicating that the particulate lidar ratio is too large. In these situations, the retrieval of profiles of both the particulate backscatter and its uncertainty are restarted from the top of the feature being analyzed using a suitably reduced value of the lidar ratio. Similar caveats apply here as in the preceding discussion on the existence of the backscatter solution. For underlying layers, the retrieved value of the total backscatter, $\beta_T(r)$, can be too large or too small depending on the accuracy of the values of the lidar ratios used in analyzing the overlying layers, and this can influence whether or not the backscatter uncertainty solution in Eq. (S14) exists.

If adjustment of the lidar ratio within the acceptable limits still cannot produce a backscatter solution or a backscatter uncertainty solution at a particular range, then the retrieval is terminated at the previous range, the remainder of the profiles down to the detected base set to a fill value (-333) and the extinction QC bits are set to indicate an error.

S1.3 Comment on the use of “standard” error propagation techniques.

The uncertainty analysis presented above (and in Young et al., 2013) is based on “standard” error propagation techniques. These methods use differential calculus to attempt to find the value of a function, e.g. $f(x,y)$, at some “nearby” location, $f(x+\Delta x, y+\Delta y)$, where Δx and Δy are small changes (interpreted as absolute uncertainties or errors) in the variables x and y . (Here a function of two variables is used as an example.) The calculation can be approached using a Taylors series expansion around the point (x,y) (Bevington and Robinson, 1992):

$$\begin{aligned} \Delta f(x, y) &= f(x + \Delta x, y + \Delta y) - f(x, y) \\ &= \left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y} \right) f(x, y) + \frac{1}{2!} \left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y} \right)^2 f(x, y) \\ &\quad + \frac{1}{3!} \left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y} \right)^3 f(x, y) + \frac{1}{4!} \left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y} \right)^4 f(x, y) \\ &\quad + \dots + \frac{1}{n!} \left(\Delta x \frac{\partial}{\partial x} + \Delta y \frac{\partial}{\partial y} \right)^n f(x, y) + R_n. \end{aligned} \quad (\text{S18})$$

20

The problem is then linearized by assuming that Δx and Δy are small enough that the second and higher-order derivative terms are negligibly small, thereby allowing only the first term in the series to be retained.

As explained in Bevington and Robinson (1992), the size (and sign) of the errors Δx and Δy are usually unknown and they are replaced by estimates of the standard deviations, σ_x and σ_y , derived statistically as the square roots of the corresponding variances. The uncertainty in $f(x,y)$ can then be estimated as the square root of the variance:

$$\Delta f(x, y) \approx \left[\sigma_x^2 \left(\frac{\partial f(x, y)}{\partial x} \right)^2 + \sigma_y^2 \left(\frac{\partial f(x, y)}{\partial y} \right)^2 + \sigma_{xy}^2 \left(\frac{\partial f(x, y)}{\partial x} \right) \left(\frac{\partial f(x, y)}{\partial y} \right) \right]^{1/2}. \quad (\text{S19})$$

5 The covariance scaling the partial derivatives in the third term is zero if x and y are uncorrelated.

If, however, these uncertainties (or standard deviations) are not small, and if the higher-order derivatives are not zero (i.e. the functions are non-linear), then the accuracy of this approximation will generally decrease with increasing values of Δx and Δy . In the analysis of CALIOP data, the uncertainties can sometimes be appreciable. The relative uncertainties in the initial lidar ratios used in unconstrained retrievals, for example, are typically 30% to 40% and the uncertainty in the retrieved
 10 integrated particulate backscatter and, hence, optical depth will grow with increasing penetration into the feature being analyzed. In addition, the calculation of the uncertainty in the particulate transmittance simply by taking the derivative of Eq. (S2) becomes increasingly inaccurate for large values of $\Delta \tau_p$ and needs to be estimated via a differencing analysis as in Young et al. (2013,2016).

In summary, while the above analysis will give satisfactory results for small to moderate uncertainties, the approximations
 15 and neglect of higher-order terms will lead to poorer results at higher values of uncertainties. Because the function relating transmittance to optical depth is non-linear, the absolute magnitude of the calculated overall uncertainty will differ depending on the sign of the input uncertainties. For these situations, data users are advised to refer to the sections in Young et al. (2013,2016) that cover the analysis of the sensitivity of retrievals to various input errors and the biases they produce in the results.

20 **S2 New initialization of backscatter solution**

As explained in the previous section, the particulate backscatter at any range is retrieved using Newton's Method Iteration. In order to ensure that the solution converges to the correct value and in as few iterations as possible, it is important that the initial value be as close to the final value as possible. For small to moderate values of the exponent in Eq. (S4), a good estimate for the initial value can be found by approximating the exponential function by the first three terms of its Taylor
 25 series expansion and solving the resulting quadratic equation. Before the Newton's Method is applied at any range, then, the value of the approximation for the exponential function is compared with the true value. If the difference is less than a specified threshold, then the estimate for $\beta_p(r)$ using the approximation is used as the initial value in the Newton's Method Iteration. If the difference exceeds the threshold, or if there is no solution to the quadratic equation, then the Newton's Method Iteration is initialized using a value of $\beta_p(r)$ derived from a version of Eq. (S4) with the exponential term set to zero.

Currently the threshold is 1%. This corresponds to the requirement $\eta S_p \beta_p = \eta \sigma_p \leq \sim 0.44 / \delta r$. When initialized in this way, the Newton's Method Iteration usually converges on the first iteration for all but the highest backscatter values.

S3 New Initialization of particulate lidar ratio in opaque layers

For opaque layers, it is possible to derive an initial estimate of the particulate lidar ratio, S_p , that is closer to the correct value than the fixed initial values for a particular layer type. Calculation of this estimate follows the approach of Fernald et al. (1972) who derive expressions for the lidar ratio as functions of the integrated lidar signal and the measured two-way layer transmittance. For a two-component atmosphere (one containing both molecules and particles), Fernald et al. (1972) derive the following expression (where we have changed the symbols to those used in this work and in Young and Vaughan (2009) and Young et al. (2013). The ranges, r_N and r_b , refer, respectively, to the normalization range at the top of the layer and the range at the base of the layer.

$$S_p = \frac{\left[1 - T_p^2(r_N, r_b) T_M^{2\eta S_p / S_M}(r_N, r_b) \right]}{2\eta \int_{r_N}^{r_b} \beta'_N(z) T_M^{2(\eta S_p / S_M - 1)}(r_N, z) dz} . \quad (\text{S20})$$

For opaque layers, the particulate transmittance is zero and the numerator becomes unity. As the unknown, S_p , appears on both sides of Eq. (S20), the equation is transcendental and must be solved iteratively. The initial value for this iteration is obtained using a modification of the single-component atmosphere equation (Fernald et al., 1972; Platt, 1973; Vaughan et al., 2005).

$$S_p = 1.0 / 2\eta \int_{r_N}^{r_b} \beta'_N(z) dz . \quad (\text{S21})$$

This equation is not transcendental and can be evaluated directly. Typically, two to three iterations of Eq. (20) are required for convergence of successive iterations to a relative difference of 0.001. The absolute uncertainty, ΔS_p , is then adjusted to maintain the same relative uncertainty as in the fixed initial value.

20 S4: Changes in extinction quality control (QC) flags from V3 to V4

As a result of changes in the extinction retrieval algorithm and in the lidar ratios and multiple scattering factors used in the retrievals, the behavior and final state of an extinction retrieval may be different in V4 from those in V3. Table S1 maps the changes in extinction QC flags from V3 to V4 for layers that were detected identically (i.e., same time, same place, same detection resolution and same layer number within the column) in both data sets during 2007 - 2010. The QC values for clouds of all types are compared in (a) while QC values for all aerosols are compared in (b). It is immediately apparent that identical QC values were assigned in only 65 % of the identical cloud layers but in 93 % of the aerosol layers. Also apparent is the dramatic increase in the fraction of constrained retrievals (QC = 1) for clouds from 1.8 % in V3 to 10.5 % in V4. The increase in constrained retrievals for aerosol layers, from 0% in V3 to 0.8% in V4, was somewhat smaller. For clouds, the

number of unsuccessful retrievals, grouped under QC = 99, increased from 0.02% in V3 to 3.0 % in V4. For aerosols, the fraction increased from 0.001% to 2.4%. For both clouds and aerosols, the vast majority of these cases are comprised of layers for which the QC flag was either 2 (only QC bit 1 was toggled) or 18 (QC bits 1 and 4 are toggled) in V3 and was changed to QC = 6 in V4 (QC bits 1 and 2 are toggled).

5 There are many reasons for the different QC flag values assigned in V3 and V4. First of all, it should be emphasized that, although identical layers were selected for the comparison, the underlying cloud populations for any QC flag value are likely to be different in V4 from those in V3. In addition to the relaxed conditions that permit an increase in the number of constrained retrievals (see Sect. 2.2.1), improvements to the algorithms determining ice-water phase (Avery et al., 2018) have changed the designated cloud phase assigned to many layers and this leads to the use of different lidar ratios and multiple scattering factors. The default lidar ratios and multiple scattering factors for any cloud type are also different, as explained in Sect. 2.1.2 and, in opaque layers, the lidar ratio is calculated from the data rather than using the default value (Sect. 2.2.3). The new surface detection algorithm (Vaughan et al., 2018) has led to a more reliable detection of the surface and, hence, of cloud opacity. This change yields a reduction in the number of opaque layers (QC = 16 and QC = 18) from 52% in V3 to 41% in V4, but the generally higher effective lidar ratios in V4 have led to many more instances where the initial value has had to be reduced (QC = 18). (Many of the layers that had QC = 16 in V3 now have QC = 18 in V4.)

References

- Avery, M., Ryan, R., Getzewich, B., Vaughan, M., Winker, D., Hu, Y., Trepte, C.: CALIOP V4.1 Cloud Phase Algorithm, in preparation, 2018.
- Bevington, P. R., and Robinson, D. K.: Data Reduction and Error Analysis for the Physical Sciences. 2nd ed. McGraw-Hill Inc., 328 pp., 1992.
- 20 Fernald, F. G., Herman, B. M., and Reagan, J. A.: Determination of aerosol height distributions with lidar. *J. Appl. Meteor.*, 11, 482–489, 1972.
- Platt, C. M. R.: Lidar and radiometer observations of cirrus clouds. *J. Atmos. Sci.*, 30, 1191–1204, 1973.
- Vaughan, Mark A., Winker, David M., and Powell, Kathleen A., 2005: CALIOP Algorithm Theoretical Basis Document. Part 2: Feature Detection and Layers Properties Algorithms. PC-SCI-202 Part 2, Release 1.01, 27 September 2005. https://www-calipso.larc.nasa.gov/resources/pdfs/PC-SCI-202_Part2_v1.0.pdf
- 25 Young, S. A., and Vaughan, M. A.: The retrieval of profiles of particulate extinction from Cloud Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) data: Algorithm description. *J. Atmos. Oceanic Technol.*, 26, 1105–1119, doi:10.1175/2008JTECHA1221.1, 2009.
- 30 Young, S. A., Vaughan, M. A., Keuhn, R. E., and Winker, D. M.: The retrieval of profiles of particulate extinction from Cloud–Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) data: Uncertainty and error sensitivity analyses. *J. Atmos. Oceanic Technol.*, 30, 395–428, doi:10.1175/JTECH-D-12-00046.1, 2013.

Young, S. A., Vaughan, M. A., Keuhn, R. E., and Winker, D. M.: Corrigendum. *J. Atmos. Oceanic Technol.*, 33, 1795–1798, doi: 10.1175/JTECH-D-16-0081.1, 2016.