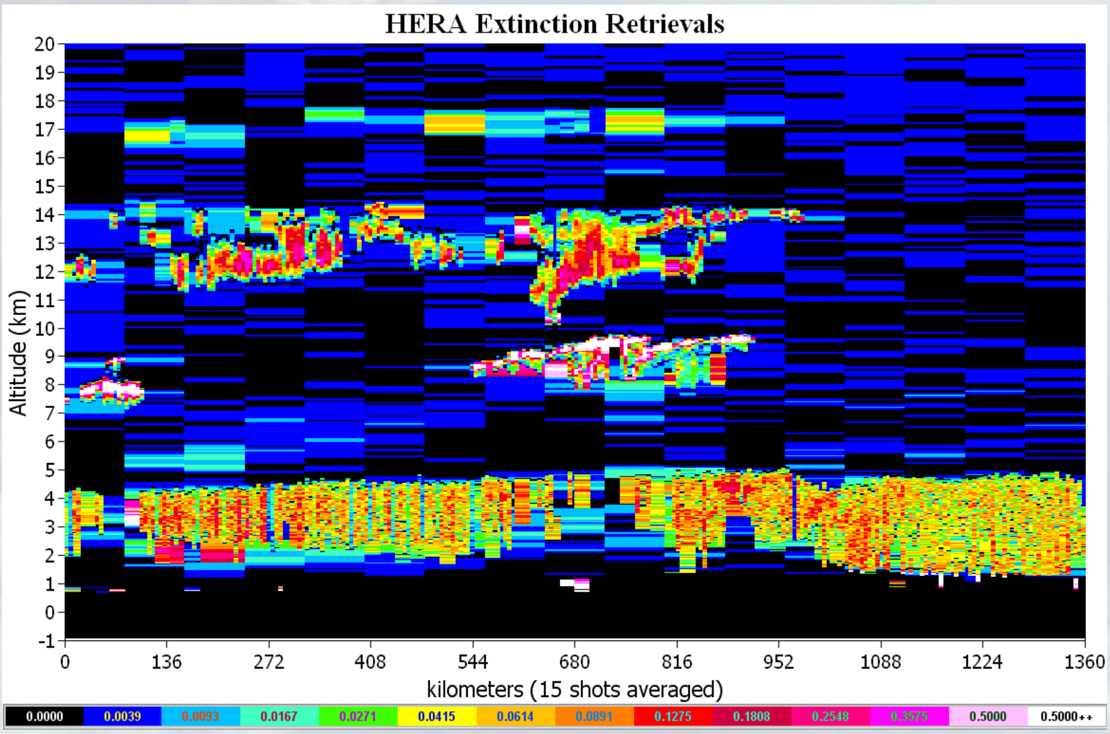
# The Hybrid Extinction Retrieval Algorithms (HERA) for the analysis of lidar data from space.

Stuart Young




CSIRO
Atmospheric Research
CSIRO Atmospheric Research Technical Paper No 54.

**National Library of Australia Cataloguing–in-Publication Entry**

Young, Stuart A.
The Hybrid Extinction Retrieval Algorithms (HERA) for the analysis of lidar data from space.

ISBN 0 643 06650 0.

1. Atmosphere - Remote sensing. 2. Optical radar.
I. CSIRO Atmospheric Research. II. Title.

551.510287

This work is supported by NASA Langley Research Center under contracts NAS1-19570 and NAS1-02058.

CSIRO Atmospheric Research
PB 1, Aspendale, Victoria 3195, Australia
Ph: (+61 3) 9239 4400; Fax: (+61 3) 9239 4444
E-mail: chief@dar.csiro.au

CSIRO Atmospheric Research Technical Papers may be issued out of sequence. From July 2000, all new Technical Papers will appear on the Web site of CSIRO Atmospheric Research. Some Technical Papers will also appear in paper form.

A complete list of CSIRO Atmospheric Research Technical Papers can be found at http://www.dar.csiro.au/info/TP.htm

# The Hybrid Extinction Retrieval Algorithms (HERA) for the analysis of lidar data from space.

Stuart A. Young,

CSIRO Atmospheric Research,
PMB 1 Aspendale VIC 3195,
Australia.

**Abstract**

The particular difficulties encountered in the automatic analysis of lidar data from space are discussed. Adaptive methods, which select different analysis parameters autonomously according to the nature of the atmospheric target and quality of the lidar signal, are proposed. These methods have been developed by the author and incorporated into software modules for testing various aspects of their performance. The application of the resulting Hybrid Extinction Retrieval Algorithms (HERA) to the analysis of lidar data from NASA's planned CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) mission is illustrated with suitably modified LITE (Lidar In-space Technology Experiment) lidar data.

## 1. Introduction

The analysis of data obtained from satellite-borne lidar faces a number of difficulties not faced by ground-based lidar. These are the result of the combination of the large range of the target from the lidar, typically 500 km to 700 km, the high speed at which the lidar sweeps across the target space (typically 7 km per second), the variation in the nature of the target along the satellite ground track, and limitations on the data acquisition and transfer rate caused by the limited available power. The large distance means that the signal-to-noise ratio (SNR) of the data will be low. The low firing rate combined with the high satellite ground speed and variations in the target along this track mean that traditional methods of improving the SNR by averaging a sufficiently large number of profiles must be modified.

Because of the non-linear way in which backscatter and extinction are related in a lidar signal, averaging of profiles that are significantly different in structure can produce a signal profile in which the relationship between backscatter and extinction is quite different from that in a single profile, and analysis of such a profile can produce results that are unrepresentative of the actual situation in the atmosphere. For this reason, considerable effort is made to average only that number of profiles that will produce a SNR that is sufficient for successful analysis. Because both the along-track target homogeneity and the strength of the signals varies depending on the type of target, different amounts of averaging are required for different targets. The strong signals from water clouds require relatively few profiles to be averaged in order to produce a signal with a SNR high enough for successful analysis. However, a weak aerosol layer in the free troposphere may require very many profiles to be averaged. For the CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) data set, initial horizontal-averaging intervals of 5 km, 20 km and 80 km have been established for different atmospheric targets. Details of this averaging technique and of the way in which specific targets are detected using the Selective Iterated BoundarY Locator (SIBYL) are described by Vaughan (2002).

Different atmospheric targets are identified on averaging different numbers of profiles, so the resultant, averaged profiles have different horizontal resolutions. The features contained in these signals are only identified over particular signal ranges or height intervals). So a dense water cloud may be identified in a narrow height band in a 5-km average profile, while a tenuous cirrus cloud may only be detected in an 80-km average profile. The different features identified in a profile will usually have different optical properties, the extinction-to-backscatter (or lidar) ratio, for example. In addition to these issues, the signals from different atmospheric features are amenable to different analysis methods. While it is common to process lidar signals using the so-called backward analysis direction (e.g. Klett, 1981, 1985) in order to ensure stability of the solution, this is not always possible. Because of the different horizontal resolutions, and different requirements for the analysis of different features within a given profile, traditional lidar analysis techniques are not applicable to the analysis of CALIPSO lidar data. Instead, a set of Hybrid Extinction Retrieval Algorithms (HERA) is required that will be flexible enough to allow the analysis of separate sections of profiles using quite different analysis parameters. Once these separate sections of profiles, representing different height intervals and horizontal resolutions, have been analyzed, the results are combined to produce a composite output using the extinction Over-Painting Technique (eOPT).

For testing the performance of the analysis algorithms, the process of analyzing CALIPSO lidar data has been divided between separate software modules. These modules are responsible for feature boundary location, scene classification, and extinction profile calculation. The first module contains the Selective Iterated BoundarY Locator (SIBYL) and Scene Classifier Algorithm (SCA) routines. CALIPSO lidar data will be analysed in 80-km blocks. The (SIBYL) first averages CALIPSO lidar profiles to 5-km, 20-km or 80-km horizontal resolution (corresponding to averages of 15, 60, and 240 individual profiles respectively). It then detects features corresponding to cloud or aerosol layers at these resolutions. Finally, the feature optical thickness is estimated where possible using the transmittance method (Young, 1995). Next, the Scene Classifier Algorithm (SCA) determines initial values of lidar ratios and multiple scattering factors for the features. This information is passed to the HERA as a block of data consisting of one 80-km averaged set and as many as four 20-km sets and sixteen 5-km sets, depending on the targets in the scene, each set containing both feature and profile data. A Meteorological Data Manager supplies profiles of molecular backscatter, transmittance and ozone transmittance. Details of the second module, containing the HERA and eOPT routines, are covered in this paper.

## 2. An overview of the functions of the HERA and eOPT routines

The HERA has several preferred, or default, processing decision paths. Wherever possible, a backward solution (far-point calibration) is used to provide a stable retrieval. The accuracy of the retrieval is also maximized by constraining the retrieved optical thickness to agree with that measured by the SIBYL. This agreement is achieved by automatic, iterative adjustment of the lidar ratio in the case of features, or of the normalization factor (in the calibrated data here, the product of the transmittances of the overlying layers) in the case of "feature-free" or apparently clear regions. There are situations where these preferred processing paths are not possible, and the algorithms select the analysis settings that will give the best result for that particular situation.

The processing of the profiles of attenuated backscatter produced by the SIBYL is controlled in the prototype analysis software by a shell routine in which the analysis parameters are selected automatically and the extinction retrieval algorithm called. An overview of the shell routine (SUBROUTINE SOLVER) is presented in Figure 2. In order that the performance of several different algorithm configurations can be tested, this software is supplied with a settings file that allows the user to select different processing options. This allows, for example, many of the default options in the HERA to be overridden. This flexibility will assist in the specification of the final analysis code, which will run autonomously with little user intervention.

The function of the HERA and eOPT routines is to extract particulate (cloud or aerosol) backscatter and extinction profiles, and their uncertainties, at both 532 nm and 1064 nm, from the blocks of data supplied by the SIBYL / SCA module. First the 80-km profile, as processed by the SIBYL, is analyzed in the regions where the SIBYL has identified no features in order to produce a profile of the background (or wash coat – see eOPT section below) aerosol for the whole 80-km block. This may require analysis of several, separate sections of profile both above and below partially transmitting clouds. Then the various features found during the 80-km scan are analyzed using the appropriate parameters. Next, the four 20-km average profiles are analyzed to produce extinction and backscatter data in the height regions in which the SIBYL detected features. The process is then repeated for the detected features in the sixteen 5-km averages. Note that the background aerosol is not retrieved at 20-km or 5-km resolution, or in any of the 1064-nm data, as the SNR is too low. Finally the retrieved profiles are recombined using the extinction over-painting technique (eOPT) to produce a composite picture of the extinction over the 80-km block.

The 80-km, attenuated backscatter profile is inspected by the HERA prior to analysis in order to identify no-signal regions where features have been identified by the SIBYL at finer horizontal resolutions and removed. The wash coat regions are thus subdivided and analyzed separately with the cumulative transmittance being tracked as the solution proceeds from the top altitude to the last valid range. By default, the wash coat analysis is refined by testing the overall particulate optical thickness against that supplied in a reference profile and the normalization factor (the product of all the transmittances for all the overlying regions) for that region adjusted in order to bring agreement. The lidar ratios required for the analysis of the wash coat profile in the stratosphere and the free troposphere are supplied by the SCA. The wash coat profile is only calculated for the 532-nm signal.

The analysis of profiles in those height regions where there are identified cloud or aerosol layer features is initiated using the lidar ratio supplied by the SCA. If iterative improvement of the solution is not possible, this lidar ratio is used for the final value, unless the solution diverges and the lidar ratio is reduced by the HERA routines. For features with a measurable transmittance this lidar ratio is adjusted to bring agreement between the feature optical thickness calculated in the extinction routine and the value measured by the SIBYL. The situations where this iterative improvement is not performed are where the SIBYL is unable to measure the feature optical thickness, either because it is too low to measure reliably, or because it is too high and the lidar signal is totally attenuated. Included in the first situation are profiles measured at 1064 nm, as the transmittance method for determining the optical thickness of optically thin clouds requires a detectable signal from the clear atmosphere on both sides of the feature, and these are not measurable at 1064 nm. As the optical thickness cannot be determined directly for aerosol layers in contact with the surface, the planetary boundary layer for example, iteration is not performed for these situations either.

Two extinction algorithms are available for testing the prototype software, a modified Fernald (1984) algorithm and a linear, iterative-convergent algorithm (Elterman, 1966; Gambling and Bartusek, 1972). Either forward or backward solution directions are possible with both algorithms as required by the circumstances. Note that for the analysis of 1064-nm signals, it is not possible to measure the feature optical thickness or the signal in the clear region below the feature, and solutions are only carried out in the forward direction. As in the 532-nm case, the cumulative optical thickness is calculated and used in the normalization of subsequent features.

The HERA and eOPT modules produce an output file that contains the retrieved profiles of aerosol backscatter and extinction coefficients and their uncertainties. Additionally, the updated values for the normalization factors, lidar ratios and layer optical thicknesses, and the layer-averaged, weighted mean color ratio are calculated.

## 3. Hybrid Extinction Retrieval Algorithms – mathematical basis

The lidar signal *P(r),* from a range *r* from the lidar, and for an atmosphere containing molecules, aerosols, clouds and ozone can be written:

$$P(r) = C[\beta_M(r) + \beta_A(r) + \beta_C(r)]T_M^2(0,r)T_{O_3}^2(0,r)T_A^2(0,r)T_C^2(0,r)/r^2 + P_0\,, \qquad (1)$$

where C is the lidar system or calibration constant, and $\beta(r)$ and $T^2(0,r)$ are, respectively, the backscatter coefficient and transmittance. The subscripts $M$, $O_3$, $A$ and $C$ denote, respectively, molecular, ozone, aerosol and cloud. The quantity $P_0$ represents a background or offset signal.

During Level 1 processing of the data, the signal is calibrated and the offset removed. In the SIBYL the signal is corrected for ozone transmittance using model values, various numbers of profiles are averaged, and the resulting signal processed to extract information on any detected features. The profile supplied to the HERA by the SIBYL is of the attenuated backscatter signal:

$$\beta'(r) = [\beta_M(r) + \beta_A(r) + \beta_C(r)]T_M^2(0,r)T_A^2(0,r)T_C^2(0,r)\,. \qquad (2)$$

The uncertainty in this quantity is also supplied:

$$\Delta\beta'(r) = \left(\left(\frac{\Delta P(r)}{P(r)}\right)^2 + \left(\frac{\Delta C}{C}\right)^2 + \left(\frac{\Delta T_{O_3}^2(0,r)}{T_{O_3}^2(0,r)}\right)^2\right)^{1/2} \beta'(r)\,, \qquad (3)$$

where the $\Delta$ symbol represents the uncertainty in the quantity following the symbol. The assumptions made in this process are:

1. The $C$ for the group of profiles averaged is the same as that for which the value of $C$ was derived.
2. The value of $T_{O3}^2(0,r)$ for the group of profiles averaged is the same as that for the model used.

The SCA will also supply profiles of the multiple scatter function $\eta(r)$, and its uncertainty. The molecular backscatter and transmittance profiles and their uncertainties are provided by the Meteorological Manager Module (MMM). To account for a low-magnitude, background aerosol that is below the detection threshold of the SIBYL, provision is made for a background reference backscatter profile $\beta_{Aref}(r)$ and transmittance profile $T^2_{Aref}(0,r)$ to be supplied by the MMM. (In the current software these quantities are set to zero and unity respectively at all altitudes, as no aerosol reference profile is available. This reference profile should also be used by the SIBYL in determining the optical thickness and lidar ratios of the features, otherwise errors in the quantities occur. *It is imperative that both the SIBYL and the HERA use the same reference profile.*) Statistics of each layer detected by the SIBYL and SCA, including the altitude of the base and top, layer-averaged multiple scattering factor, optical thickness and lidar ratio, and their uncertainties are also supplied.

### 3.1 Analysis of the top "clear" region of the "wash coat" profile.

Consider now the steps followed by HERA in processing the profile depicted in Figure 1, which corresponds to an 80-km average profile. It contains two features (a cloud layer and an aerosol layer at the surface), and signal in the apparently clear regions from which the background or "wash coat" aerosol extinction will be calculated. Assume that the SIBYL has detected a cloud between $r_{b1}$ and $r_{t1}$ and a surface aerosol layer with a top at altitude $r_{t2}$. The analysis proceeds from the top of the profile down to the lowest valid altitude over the range interval $r_{min}$ to $r_{max}$ . In the 80-km wash coat profiles, the region from $r_{min}$ to $r_{t1}$ is analysed first. Although the backward solution direction is preferred in the HERA, this may be overridden in the current software in order to test the performance of a forward solution.

A normalization factor is chosen at a normalization (or calibration) range $r_c = r_{min}$ , for the forward direction, or $r_c = r_{t1} - \delta r$, (where $\delta r$ is the range increment) for the backward direction,

$$C_N(r_c) = T^2_M(0,r_c)T^2_{Aref}(0,r_c)\,, \tag{4}$$

with uncertainty

$$\Delta C_N(r_c) = \left( \left( \frac{\Delta T^2_M(0,r_c)}{T^2_M(0,r_c)} \right)^2 + \left( \frac{\Delta T^2_{Aref}(0,r_c)}{T^2_{Aref}(0,r_c)} \right)^2 \right)^{1/2} C_N(r_c)\,. \tag{5}$$

The assumption made in this step is that the aerosol transmittance from the lidar to the normalization range in the profile under analysis is the same as that in the reference profile. Except in conditions of enhanced volcanic aerosol loading, when the SIBYL should detect such layers anyway, this assumption is likely to be reasonable.

Once the normalization and minimum and maximum ranges are selected by the HERA, the section of profile and accompanying data are sent to the selected extinction analysis subroutine. In this subroutine, the first step is to normalize the profiles of attenuated backscatter and its uncertainty:

$$\beta'_N(r) = \beta'(r)/C_N(r_c)\,, \tag{6}$$

$$\Delta\beta'_N(r) = \left( \left( \frac{\Delta C_N(r_c)}{C_N(r_c)} \right)^2 + \left( \frac{\Delta\beta'(r)}{\beta'(r)} \right)^2 \right)^{1/2} \beta'_N(r)\,. \tag{7}$$

Strictly, the quantity $\beta'_N(r)$ represents a family of curves and should be written as $\beta'_N(r_c, r)$ to reflect the dependence on the normalization range $r_c$ , but to simplify the notation, the inclusion of the normalization factor in the normalized attenuated backscatter is signified by the subscript *N*.

In those situations where the *forward analysis* direction is used, the analysis, using the example of the linear, iterative-convergent technique, proceeds as follows. In the forward direction, $r_c < r$, so all the transmittance factors can be factorized as shown:

$$T^2(0,r) = T^2(0,r_c)T^2(r_c,r)\,. \tag{8}$$

The normalized attenuated backscatter coefficient, then becomes

$$\beta'_N(r) = [\beta_M(r) + \beta_A(r)]T_M^2(r_c,r)T_A^2(r_c,r)\,. \tag{9}$$

The first estimate for the aerosol backscatter at range r is

$$\beta_A(r) = \beta'_N(r)/(T_M^2(r_c,r)T_A^2(r_c,r)) - \beta_M(r)\,. \tag{10}$$

Note that as

$$T_A(r_c,r) = \exp[-\eta(r)S_A \int_{r_c}^{r} \beta_A(z)dz] = \exp[-\eta(r)\tau_A(r_c,r)]\,, \tag{11}$$

the aerosol transmittance factor in (10) includes the unknown, $\beta_A(r)$, being sought. ($S_A$ is the lidar ratio for the aerosols in the region being analyzed, and $\tau_A$ is the aerosol optical thickness.) This apparent impasse is overcome by assuming initially that the contribution of $\beta_A(r)$ in (11) is zero, then using the new value calculated using (10) to refine the transmittance in (11), and continuing the iteration between (10) and (11) until relative changes in consecutive values of $\beta_A(r)$ are less than some specified tolerance. In the numerical implementation of these equations in the software, the integral in (11) is evaluated using the trapezoidal rule and the quantity being evaluated in (10) is the aerosol backscatter at the current range step $r_j$. (Elterman, 1966; Gambling and Bartusek, 1972). In order to accelerate the convergence, at ranges other than the normalization range, the initial value of $\beta_A(r)$ is chosen as the value calculated at the previous range. Once the aerosol backscatter has been retrieved, the aerosol extinction profile is obtained by scaling by the appropriate lidar ratio:

$$\sigma_A(r) = S_A\beta_A(r)\,. \tag{12}$$

Where the *backward analysis direction* is used, a similar procedure is followed. Now however, $r_c > r$, so all the transmittance factors can be expressed as

$$T^2(0,r_c) = T^2(0,r)T^2(r,r_c)\,. \tag{13}$$

The normalized, attenuated backscatter for the backward direction becomes

$$\beta'_N(r) = [\beta_M(r) + \beta_A(r)]/(T_M^2(r,r_c)T_A^2(r,r_c))\,. \tag{14}$$

Equations (10) and (11) become, for the backward direction,

$$\beta_A(r) = \beta'_N(r)(T_M^2(r,r_c)T_A^2(r,r_c)) - \beta_M(r)\,, \tag{15}$$

and

$$T_A(r,r_c) = \exp[-\eta(r)S_A\int_r^{r_c}\beta_A(z)dz]\,. \tag{16}$$

Note that, in the above equations, the assumption is *not* made that the aerosol profile is the same at all ranges as in the aerosol reference profile, that is that $\beta_A(r) = \beta_{Aref}(r)$ for all ranges $r$, nor that $T_A^2(0,r) = T_{Aref}^2(0,r)$, so the retrieved aerosol profile need not be the same as the aerosol reference profile. The assumption that $T_A^2(0,r_c) = T_{Aref}^2(0,r_c)$ was made only to allow the normalization factor to be calculated in (4).

### *3.1.1 Refinements to the extinction algorithm – adjustments to S to prevent retrieved extinction too large or too small.*

An inspection of (10) and (11) reveals a potential instability in the analysis in the forward direction. If the normalization constant calculated using (4) is too small or the lidar ratio, $S_A$, in (11) too large, then the attenuated backscatter in (9) and retrieved backscatter in (10) will be too large. As the retrieved backscatter is used in calculating the aerosol transmittance correction through (10), this latter value will be too small and leads to an even larger value of $\beta_A(r)$. In extreme cases divergence can occur with the solution eventually becoming undefined. Divergence is determined by testing for an increase in the absolute value of the difference between successive estimates of $\beta_A(r)$. The situation is corrected by decreasing the value of the lidar ratio and restarting the retrieval from the calibration range.

Another form of error can occur where the lidar ratio is too small. This can lead to the retrieved aerosol backscatter profile going negative. This condition is determined by testing if the retrieved values of $\beta_A(r)$ at a number of consecutive ranges are negative when the corresponding values of the attenuated backscatter are positive. In these cases the lidar ratio is increased.

In situations of moderate to high optical thickness, it is possible that an adjustment of the lidar ratio in one direction in order to overcome one problem may well cause the other problem. For example, increasing the lidar ratio to prevent negative retrievals may well cause the solution to diverge. In order to overcome this difficulty the following scheme is adopted, in which the adjustments of the lidar ratio up and down are linked. If a change in one direction (a decrease, for example) is required and if there has been no previous adjustment in that or in the opposite direction (an increase), then a one percent adjustment (a decrease in this example) is made to the lidar ratio. Once there has also been an adjustment in the opposite direction, the new value of the lidar ratio is chosen as the mean of the last increased value and the previous decreased value. The result is a lidar ratio that is neither too large nor too small. The scheme is illustrated in Figure 3.

### *3.1.2 Error analyses*

The error analyses for both the forward and backward directions are virtually identical, as can be deduced by comparing equations (10) and (11) with (15) and (16). While the analyses are identical for both forward and backward solutions, because of the quite different magnitudes of the various quantities at near-point and far-point normalization ranges, the growth of the uncertainties with range is quite different. The analysis that follows will use the forward direction. In the following error analysis, it will be assumed that errors in the different quantities are random and uncorrelated. While this may not always be strictly true, it is expected that the errors incurred in this approximation will be much smaller than the main contributors to the overall uncertainties.

From (10), the uncertainty in the retrieved aerosol backscatter coefficient can be written:

$$(\Delta\beta_A(r))^2 = \beta_T^2(r)\left[\left(\frac{\Delta\beta'_N(r)}{\beta'_N(r)}\right)^2 + \left(\frac{\Delta T_M^2(r_c,r)}{T_M^2(r_c,r)}\right)^2 + \left(\frac{\Delta T_A^2(r_c,r)}{T_A^2(r_c,r)}\right)^2\right] + \left(\Delta\beta_M(r)\right)^2. \tag{17}$$

Here, $\beta_T(r)$ is the total backscatter coefficient at range $r$. The uncertainty in the aerosol transmittance term in (17) is derived from (11):

$$\left(\Delta T_A^2(r_C,r)\right)^2 = \left(\frac{\partial T_A^2(r_C,r)}{\partial\eta(r)}\right)^2 (\Delta\eta(r))^2 + \left(\frac{\partial T_A^2(r_C,r)}{\partial\tau_A(r_C,r)}\right)^2 (\Delta\tau_A(r_C,r))^2, \tag{18}$$

which leads to

$$\left(\frac{\Delta T_A^2(r_C,r)}{T_A^2(r_C,r)}\right)^2 = (2\tau_A(r_C,r)\Delta\eta(r))^2 + (\eta(r))^2(2\Delta\tau_A(r_C,r))^2. \tag{19}$$

The optical thickness can be expressed as the product of the lidar ratio and the integrated backscatter:

$$\tau_A(r_C,r) = S_A\gamma_A(r_C,r), \tag{20}$$

which leads to an expression for the uncertainty:

$$(\Delta\tau_A(r_c,r))^2 = (\Delta S_A)^2\gamma_A^2(r_c,r) + (\Delta\gamma_A(r_c,r))^2 S_A^2. \tag{21}$$

In the software implementation of the extinction retrieval algorithm, the integrated aerosol backscatter coefficient is calculated at each range increment using trapezoidal integration:

$$\gamma_A(r_c,r) = \gamma_{Aj} = 0.5(\delta r_j\beta_{Aj} + 2\delta r_{j-1}\beta_{Aj-1} + ... + 2\delta r_{jcal+1}\beta_{Ajcal+1} + \delta r_{jcal}\beta_{Ajcal}), \tag{22}$$

where $\delta r_j$ is the $j$th range increment (note that these are not all the same). The uncertainty in $\gamma_{Aj}$ can now be calculated as

$$(\Delta\gamma_{Aj})^2 = 0.25((\delta r_j \Delta\beta_{Aj})^2 + (2\delta r_{j-1}\Delta\beta_{Aj-1})^2 + ... + (2\delta r_{jcal+1}\Delta\beta_{Ajcal+1})^2 + (\delta r_{jcal}\Delta\beta_{Ajcal})^2)\,. \qquad (23)$$

However, as the first term in (23) is the unknown in (17), this term is approximated by the value from the previous range. Finally the uncertainty in the aerosol extinction coefficient can be derived from (12):

$$\Delta\sigma_A(r_c, r) = \left(\left(\frac{\Delta S_A}{S_A}\right)^2 + \left(\frac{\Delta\beta_A(r)}{\beta_A(r)}\right)^2\right)^{1/2} \sigma_A(r)\,. \qquad (24)$$

### *3.1.3 Iterative improvement of retrieved profiles*

Improvement of the retrieved profiles is possible by adjusting iteratively the normalization factor (4) until the retrieved aerosol optical thickness over the analysis interval matches the value in the reference profile. Note that because the aerosol optical thickness in these "clear" regions is very low, it is not very sensitive to adjustments in the lidar ratio, and adjustment of the normalization factor is preferred. This adjustment may be required if the normalization error (5) is too large or if the model ozone transmittance profile over the analysis interval is significantly different from the actual profile. (In the current software implementation, the user has the option of switching off the iteration in the wash coat regions.)

## 3.2 Analysis of lower "clear" regions of the "wash coat" profile.

The analysis of clear regions below detected features, for example between $r_{b1}$ and $r_{t2}$ in Figure 1, proceeds in the same way as that for the top clear region described in the previous section, and the equations are the same with the following exception. The normalization factor now needs to include the transmittance losses of all the preceding (overlying) regions. As the analysis of a profile proceeds from the top down, these transmittance losses, for both the clear regions and the features, can be tracked and included in the normalization. The normalization factor now becomes

$$C_N(r_c) = T_M^2(0, r_c)T_{Aref}^2(0, r_c)T_{prev}^2\,, \qquad (25)$$

where $T_{prev}^2$ is the square of the total aerosol and feature transmittance down to the new normalization level:

$$T_{prev}^2 = \prod_{i=1}^{i=if} T_{clear_i}^2 T_{feature_i}^2\,, \qquad (26)$$

and *if* is the number of the previous, overlying feature. The uncertainty in the new normalization factor is

$$\Delta C_N(r_c) = \left(\left(\frac{\Delta T_M^2(0, r_c)}{T_M^2(0, r_c)}\right)^2 + \left(\frac{\Delta T_{Aref}^2(0, r_c)}{T_{Aref}^2(0, r_c)}\right)^2 + \left(\frac{\Delta T_{prev}^2}{T_{prev}^2}\right)^2\right)^{1/2} C_N(r_c)\,. \qquad (27)$$

## 3.3 Analysis of features

The analysis of features detected by the SIBYL is only a little different from that of the intervening clear regions in the wash coat profile. Indeed, the same extinction algorithms are used for both, and equations (8) to (24) are used as before with the exception that the subscript A for aerosols can be replaced by C for clouds where appropriate. Note that, for the 20-km and 5-km averaged profiles, valid signals only exist in the regions where the SIBYL has identified features. For each feature in a profile, the SIBYL supplies estimates of the top and base altitudes, the optical thickness, and integrated attenuated backscatter, and the uncertainties in these quantities. The Scene Classifier calculates (or derives from a model) the lidar ratio and layer-averaged multiple scattering factors at both 532 nm and 1064 nm.

Where possible, the solution is improved iteratively by adjusting the lidar ratio for the feature until a match is achieved, within a specified tolerance, of the retrieved feature optical thickness and the one measured by the SIBYL. The method used for the iteration of the lidar ratio is illustrated in Figure 4. The situations where this iterative improvement is not performed are where the SIBYL was unable to measure the feature optical thickness, either because it was too low to measure, or because it was too high and the lidar signal was totally attenuated. Included in the first situation are profiles measured at 1064 nm, as the transmittance method (Young, 1995) for determining the optical thickness of optically thin clouds requires a detectable signal from the clear atmosphere both sides of the feature, and these are not measurable at 1064 nm. The optical thickness cannot be determined directly for aerosol layers in contact with the surface, the PBL for example, and iteration is not performed for these situations either.

Either the forward or the backward analysis direction is possible for many cases of signal features. These cases are now described in more detail with reference to the feature illustrated between $r_t$ and $r_b$ in Figure 1. The automatic selection of the solution direction, normalization range and other analysis parameters by the HERA is illustrated in Figure 5

*Elevated features in which the transmittance can be measured.*

The analysis interval $[r_{min}, r_{max}]$ is set between the points in the clear air on either side of the feature. For the forward solution, the normalization range is set to $r_{min}$, and the normalization factor and associated uncertainty are calculated via (25) and (27), with the total overlying transmittance including that of the top clear region (see Figure 5, column 3).

For the backward solution, the normalization range is set to $r_{max}$, and the normalization factor and associated uncertainty are calculated via (25) and (27), with the total overlying transmittance including both that of the top clear region and that of the feature (26) as measured by the SIBYL (see Figure 5, column 4).

The lidar ratio is adjusted from the initial value supplied by the SCA, in order to bring agreement between the feature optical thickness measured by the SIBYL and that calculated by the retrieval algorithm (Figure 4). Note that the calculated feature optical thickness is corrected for the contribution of the background, aerosol reference profile, to prevent a biased result.

The same correction for values of lidar ratios that are either too high or too low as described in the wash coat analysis section, and described in Figure 3, is also used in the analysis of features.

*Elevated features in which the optical thickness is too low to be measured.*

Analysis of these features proceeds in exactly the same fashion as in the previous case via (25) to (27) with the exception that the feature transmittance is taken initially to be unity (zero optical thickness). This value is then recalculated during the extinction analysis and, although still small, it is used in the calculation of the transmittance correction of underlying layers. As there is no independently measured constraint, no iteration of the solution is possible.

*Elevated features measured at 1064 nm.*

Analysis of 1064-nm features proceeds in almost the same fashion as for the previous case. However analysis interval [$r_{min}$ , $r_{max}$] is set between the end points of the feature as detected by the SIBYL [$r_t$ , $r_b$ ]. This is necessary as there is no measurable signal in the clear regions outside most features detected at 1064 nm. Also, analysis is only allowable in the forward direction. A backward solution would require a good normalization, usually performed in the clear air below the cloud. However there is no signal at 1064 nm and the required cloud transmittance cannot be measured (see Figure 5, column 1).

*Elevated features in which the transmittance is zero (totally attenuating features).*

For the totally attenuating features, analysis is performed in the forward direction using the relatively good normalization at cloud top and the lidar ratio estimated by the SCA from the integrated attenuated backscatter using equation (7) of Fernald et al. (1972). A backward solution is not performed as, although the product of the lidar ratio and the average multiple scattering factor for the feature can be estimated, the transmittance at some far point cannot be estimated with any certainty. No iteration is possible either, although the extinction algorithms may adjust the lidar ratio if an error condition is detected (see above). In addition, if the retrieved optical thickness calculated during the analysis of a profile exceeds some, user defined, maximum value, then the solution is terminated at that point and data at lower altitudes are not analyzed (Figure 5, column 2). At such high values of optical thickness, the relative uncertainty of the attenuated signal is high and the modeled value of the multiple scattering correction factor is likely also to be unreliable.

*The Planetary Boundary-Layer aerosol and similar features.*

Because it is not possible for the SCA to determine the optical thickness of features in contact with the surface, it is not possible for the solution to be improved iteratively as is done in some of the cases above. Also, because no estimate of the aerosol backscatter or extinction is available at the lower boundary of such features, only forward solutions are possible. The normalization range is chosen to be at the point in the clear air above the feature for 532-nm signals or at the first point in the feature ($r_{t2}$) for 1064 nm.

## 4. Summary of inputs required by the HERA.

The HERA requires various input data. These consist of meteorological information, statistics of features detected by the SIBYL, pre-processed signal profiles, and model values from the SCA. In addition, for the current, prototype software implementation of the HERA, several analysis parameters need to be selected by the user. The various items are listed below. The parameter names listed in column 1 are the names used in the HERA software.

### 4.1 User-selectable analysis parameters

| | ***Parameter*** | ***Meaning*** |
|---|---|---|
| 1. | numblocks | Number of 80-km blocks in input file |
| 2. | ncoarse | Number of coarse (80-km) records per data block |
| 3. | nmedium | Number of medium (20-km) records per data block |
| 4. | nfine | Number of fine (5-km) records per data block |
| 5. | ialg | Selected analysis algorithm: 1 = Fernald, 2=Linear Iterative |
| 6. | idir | Analysis direction for features: 1 = forward, 2 = hybrid |
| 7. | itvar | Iteration selection switch for features: 0 = OFF, 1=ON |
| 8. | Spmin | Minimum acceptable value of Sp from SCA |
| 9. | Spmax | Maximum acceptable value of Sp from SCA |
| 10. | Spdef | Default value when Sp < Spmin or Sp > Spmax |
| 11. | dSpdef | Uncertainty in default value of Sp (sr) |
| 12. | nwcthr | No signal threshold (number of averaged signals) for wash coat |
| 13. | idirwc | Analysis direction for wash coat: 1 = forward, 2 = backward |
| 14. | itwc | Iteration selection switch for wash coat: 0 = OFF, 1=ON |
| 15. | tolnce | Convergence tolerance in iterations |
| 16. | taumax | Optical Thickness threshold for total attenuation |
| 17. | au | Diagnostics switch: If FALSE, output diagnostics to screen |

Note that items 8 - 11 will not be needed in the final product, as the Scene Classifier Algorithms will supply the correct value of the lidar ratio for any feature.

### 4.2 Meteorological data

| | ***Parameter*** | ***Meaning*** |
|---|---|---|
| 1. | t2m0_532 | The molecular transmittance squared from lidar to 1st point |
| 2. | xb532 | The molecular backscatter cross section at 532 nm ($m^{-2}$) |
| 3. | xs532 | The molecular extinction cross section at 532 nm ($m^{-2}$) |
| 4. | dc532 | Uncertainty in calibration factor at 532 nm |
| 5. | t2m0_064 | The molecular transmittance squared from lidar to 1st point |
| 6. | xb064 | The molecular backscatter cross section at 1064 nm ($m^{-2}$) |
| 7. | xs064 | The molecular extinction cross section at 1064 nm ($m^{-2}$) |
| 8. | dc064 | Uncertainty in calibration factor at 1064 nm |
| 9. | h_trop | Height of tropopause (km) |
| 10. | i_trop | Array index of tropopause |
| 11. | z_lid | Altitude of lidar (km) |
| 12. | Sa_532_strat | Stratospheric Aerosol lidar ratio for wash coat (sr) |
| 13. | dSa_532_strat | Uncertainty in Stratospheric Aerosol lidar ratio (sr) |
| 14. | Sa_532_trop | Tropospheric Aerosol lidar ratio for wash coat (sr) |
| 15. | dSa_532_trop | Uncertainty in Tropospheric Aerosol lidar ratio (sr) |

Note: Items 4 and 8 should have been included in the uncertainty in the attenuated backscatter profiles as in Equation (3) above.

### 4.3 Meteorological profile data

| | *Parameter* | *Meaning* |
|---|---|---|
| 1. | nvals | Number of values (height levels) in profiles |
| 2. | altitude(nvals) | Altitude array for lidar and meteorological data (km) |
| 3. | dens(nvals) | Array of number density values ($m^{-3}$) |
| 4. | ddens(nvals) | Array of number density uncertainty values ($m^{-3}$) |

### 4.4 Detected feature data

| | *Parameter* | *Meaning* |
|---|---|---|
| 1. | nlayers | Number of detected layers (features) |
| 2. | h_surf | Altitude of surface (km) |
| 3. | i_surf | Array index of altitude of surface |
| 4. | h_last | Last valid altitude in signal (km) |
| 5. | i_last | Array index of last valid altitude |
| 6. | hbase(j) | Altitude of base of feature number $j \leq$ nlayers (km) |
| 7. | ibase(j) | Array index of base j |
| 8. | htop(j) | Altitude of top of feature j |
| 9. | itop(j) | Array index of top j |
| 10. | lflag(j) | If flag = 0, feature j is a surface signal |
| 11. | t2l532(j) | Square of transmittance at 532 nm for feature j. |
| 12. | dt2l532(j) | Uncertainty in square of transmittance at 532 nm for feature j |
| 13. | sp532(j) | Average lidar ratio at 532 nm for feature j (sr) |
| 14. | dsp532(j) | Uncertainty in average lidar ratio at 532 nm for feature j (sr) |
| 15. | gp532(j) | Integrated attenuated backscatter at 532 nm for feature j ($sr^{-1}$) |
| 16. | dgp532(j) | Uncertainty in item 15 ($sr^{-1}$) |
| 17. | etal532(j) | Average multiple scattering factor at 532 nm for feature j |
| 18. | detal532(j) | Uncertainty in item 17 |
| 19. | t2l064(j) | Square of transmittance at 1064 nm for feature j |
| 20. | dt2l064(j) | Uncertainty in square of transmittance at 1064 nm for feature j |
| 21. | sp064(j) | Average lidar ratio at 1064 nm for feature j (sr) |
| 22. | dsp064(j) | Uncertainty in average lidar ratio at 1064 nm for feature j (sr) |
| 23. | gp064(j) | Integrated attenuated backscatter at 1064 nm for feature j ($sr^{-1}$) |
| 24. | dgp064(j) | Uncertainty in item 23 ($sr^{-1}$) |
| 25. | etal064(j) | Average multiple scattering factor at 1064 nm for feature j |
| 26. | detal064(j) | Uncertainty in item 25 |

### 4.5 Lidar and model profile data

| | ***Parameter*** | ***Meaning*** |
|---|---|---|
| 1. | bt2_532(j) | Profile of attenuated backscatter $X(r)$ at 532 nm ($(km.sr)^{-1}$) |
| 2. | dbt2_532(j) | Uncertainty in item 1 $\Delta X(r)$ at 532 nm ($(km.sr)^{-1}$) |
| 3. | nav532(j) | Number of averaged profiles at this altitude level |
| 4. | eta532(j) | Profile of multiple scattering factor $\eta(r)$ at 532 nm |
| 5. | deta532(j) | Uncertainty in item 4 $\Delta\eta(r)$ at 532 nm |
| 6. | bt2_064(j) | Profile of attenuated backscatter $X(r)$ at 1064 nm ($(km.sr)^{-1}$) |
| 7. | dbt2_064(j) | Uncertainty in item 6 $\Delta X(r)$ at 1064 nm ($(km.sr)^{-1}$) |
| 8. | nav064(j) | Number of averaged profiles at this altitude level |
| 9. | eta064(j) | Profile of multiple scattering factor $\eta(r)$ at 1064 nm |
| 10. | deta064(j) | Profile of uncertainty in item 9 $\eta(r)$ at 1064 nm |
| 11. | betaAref(j) | Reference profile of background aerosol $\beta_{ARef}(r)$ ($(km.sr)^{-1}$) |
| 12. | dbetaAref(j) | Profile of uncertainty in item 11, $\Delta\beta_{ARef}(r)$ ($(km.sr)^{-1}$) |
| 13. | t2Aref(j) | Reference profile of background aerosol transmittance $T^2_{Aref}(r)$ |
| 14. | dt2Aref(j) | Profile of uncertainty in item 13. $\Delta T^2_{Aref}(r)$ |

## 5. Summary of the Outputs Produced by the HERA

The HERA processes input data produced by the SIBYL and SCA modules. In addition to retrieving profiles of backscatter and extinction, the HERA recalculates and updates values of the layer-averaged feature parameters supplied by the SIBYL and SCA. As in the description of the required input data, variable names used here are the same as those used in the HERA software.

### 5.1 Updated feature data

| | ***Parameter*** | ***Meaning*** |
|---|---|---|
| 1. | C_n532(n) | Updated value of the 532-nm normalization factor for feature n. |
| 2. | dC_n532(n) | Uncertainty in item 1. |
| 3. | Sf532(n) | Updated (final) value of 532-nm lidar ratio for feature n. (sr) |
| 4. | tauf532(n) | Updated (final) value of 532-nm optical thickness for feature n. |
| 5. | dtauf532(n) | Uncertainty in item 4. |
| 6. | C_n064(n) | Updated value of the 1064-nm normalization factor for feature |
| 7. | dC_n064(n) | Uncertainty in item 6. |
| 8. | Sf064(n) | Updated value of 1064-nm lidar ratio for feature n. (sr) |
| 9. | tauf064(n) | Updated value of 1064-nm optical thickness for feature n. |
| 10. | dtauf064(n) | Uncertainty in item 9. |
| 11. | cr(n) | Weighted mean color ratio through feature n. |
| 12. | dcr(n) | Uncertainty in item 11. |

## 5.2 Profile data

| | ***Parameter*** | ***Meaning*** |
|---|---|---|
| 1. | betap532(j) | Profile of retrieved 532-nm particulate backscatter $(km.sr)^{-1}$ |
| 2. | dbetap532(j) | Uncertainty in item 1. |
| 3. | sigmap532(j) | Profile of retrieved 532-nm particulate extinction $(km)^{-1}$ |
| 4. | dsigmap532(j) | Uncertainty in item 3. |
| 5. | betap064(j) | Profile of retrieved 1064-nm particulate backscatter $(km.sr)^{-1}$ |
| 6. | dbetap064(j) | Uncertainty in item 5. |
| 7. | sigmap064(j) | Profile of retrieved 1064-nm particulate extinction $(km)^{-1}$ |
| 8. | dsigmap064(j) | Uncertainty in item 7. |

## 6. The extinction Over-Painting Technique (eOPT)

As explained above in the description of the HERA, extinction profiles are calculated for features detected over limited height ranges in profiles that have produced by averaging over various horizontal distances. At the lowest level, the background aerosol extinction at 532 nm is calculated from the average of all the profiles in an 80-km block of data. Stronger features are detected in some of the four profiles that are the average over 20 km. The strongest features of all are detected in the 16 profiles produced by 5-km averages. The way in which these pieces of profiles are reassembled into an overall picture of the extinction in the 80-km block of data, consisting of an array with 16 horizontal locations and 500 vertical locations (i.e.16 adjacent columns), is described below.

The approach to creating the overall extinction description can be considered to be similar to the way in which an artist paints a picture. The first information laid down by the artist is the background or wash coat that covers the whole canvas and includes little detail. Then other layers are added over parts of the wash coat. These are smaller regions and are in finer detail. The finest details are added last. Analogously to the application of the wash coat, the retrieved aerosol extinction profile calculated from the 80-km average is assigned initially to all sixteen columns corresponding to the horizontal locations of the 5-km averages. Next, the finer resolution results, analyzed at 20-km horizontal resolution are inserted into the array at all five columns corresponding to each of the four 20-km averages, and at the vertical locations at which the features were detected. Finally, the details of each of the sixteen 5-km averages are added to the array at the appropriate locations. The picture of extinction is now finished. An overview of this method is shown in the flowchart in Figure 6.

An example of the application of this technique to LITE (Lidar In-space Technology Experiment) data, suitably modified in resolution, signal strength and signal-to-noise ratio to represent CALIPSO data, is shown in Figure 7. Figure 7(a) shows the original (modified) LITE data plotted as attenuated backscatter. Figure 7(b) shows the features as detected and identified at the various horizontal resolutions by the SIBYL. Features detected at 5 km are marked by regions colored in brown. Those detected after averaging to 20-km horizontal resolution are marked by the yellow regions, while those detected only after averaging over 80 km are shown in green. Figure 7(c) shows the particulate backscatter coefficient that has been retrieved using the HERA and plotted using the eOPT scheme. The values of the lidar ratio were set manually to realistic values, before processing by the HERA, as the Scene Classifier software is still under development. Similarly, the multiple scattering factors were set to unity at all altitudes, as this work is still under development. The solution direction was selected automatically by the HERA algorithm. Finally, the particulate extinction coefficient outputs, as retrieved by the HERA and combined using the eOPT scheme is shown in Figure 7(d). While the backscatter retrievals are fairly insensitive to variations in the lidar ratio, it can be seen that the extinction solutions are strongly dependent on the selected value of the lidar ratio. This is because, for low to moderate optical thicknesses, the extinction retrieval is approximately equal to the product of the backscatter coefficient and the lidar ratio. In those cases where the solution cannot be constrained by a measurement of the optical thickness of the feature, and iterative improvement of the lidar ratio is therefore not possible, errors in the value of the lidar ratio supplied to the HERA will translate directly to errors in the retrieved extinction coefficient. These results point out the important part that the Scene Classifier Algorithms will play in the accurate retrieval of extinction coefficient from the CALIPSO data.

## Acknowledgments

The author has pleasure in acknowledging the valuable assistance of Mark Vaughan, of Science Applications International Corporation (SAIC), with whom he collaborated closely on this work. Mark Vaughan developed the SIBYL and eOPT software modules that are complementary to the HERA code described here. He also provided the data that had been pre-processed by his SIBYL modules for analysis with the HERA modules. The simulation software that ingests LITE data and modifies it to resemble the data expected from the CALIPSO lidar was written by Kathy Powell of SAIC.

This document was submitted by the author to NASA Langley Research Center as a section in the CALIPSO Algorithm Theoretical Basis Document (ATBD) covering the retrieval of extinction coefficients. During the preparation of this document, the author valued the fruitful interactions with, and helpful suggestions from, Dave Winker of NASA Langley and Principal Investigator on the CALIPSO Mission, and Mark Vaughan of SAIC.

This work is supported by NASA Langley Research Center under contracts NAS1-19570 and NAS1-02058.

## References

Elterman, L. (1966). Aerosol measurements in the troposphere and stratosphere. *Applied Optics,* **5,** 1769-1776.

Fernald, F.G., B. M. Herman and J. A. Reagan (1972). Determination of aerosol height distributions with lidar. *Journal of Applied Meteorology,* **11,** 482-489.

Fernald, F. G. (1984). Analysis of atmospheric lidar observations: some comments. *Applied Optics,* **23,** 652-653.

Gambling, D. and K. Bartusek (1972). Lidar observations of tropospheric aerosols. *Atmospheric Environment*, **6,** 181-190.

Klett, J. D. (1981). Stable analytical inversion for processing lidar returns." *Applied Optics*, **20,** 211-220.

Klett, J. D. (1985). Lidar inversion with variable backscattering/extinction ratios. *Applied Optics*, **24,** 6451-6456.

Vaughan, M.A. (2002). SIBYL: a selective iterated boundary location algorithm for finding cloud and aerosol layers in CALIPSO lidar data, In: Proceedings of the 21st International Laser Radar Conference, Quebec, Canada, 2002.

Young, S. A. (1995). Lidar analysis of lidar backscatter profiles in optically thin clouds. *Applied Optics,* **34,** 7019-7031.

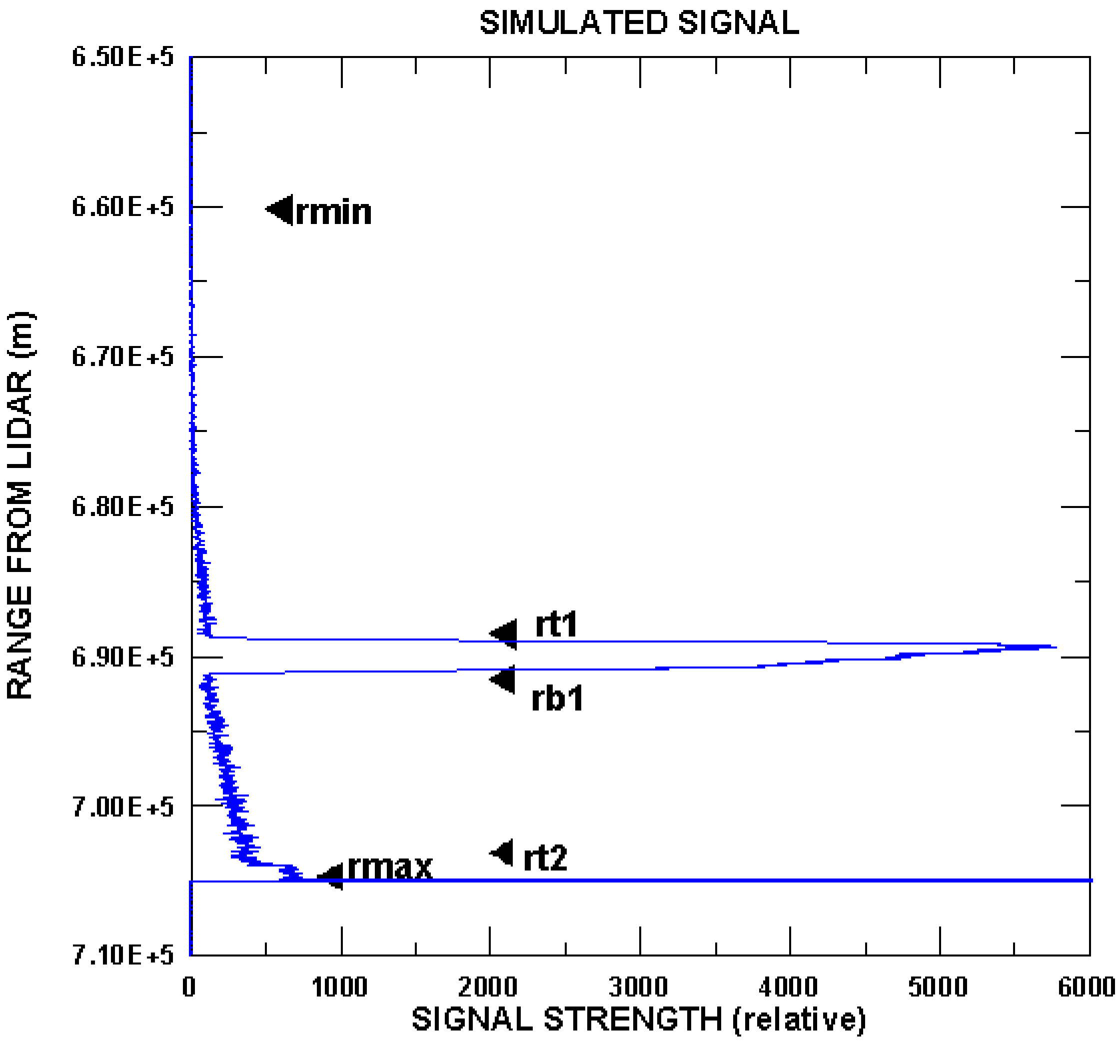



**Figure 1: A simulated space-borne lidar signal with a cloud feature and surface layer aerosol feature. Simulated lidar altitude is 705 km.**

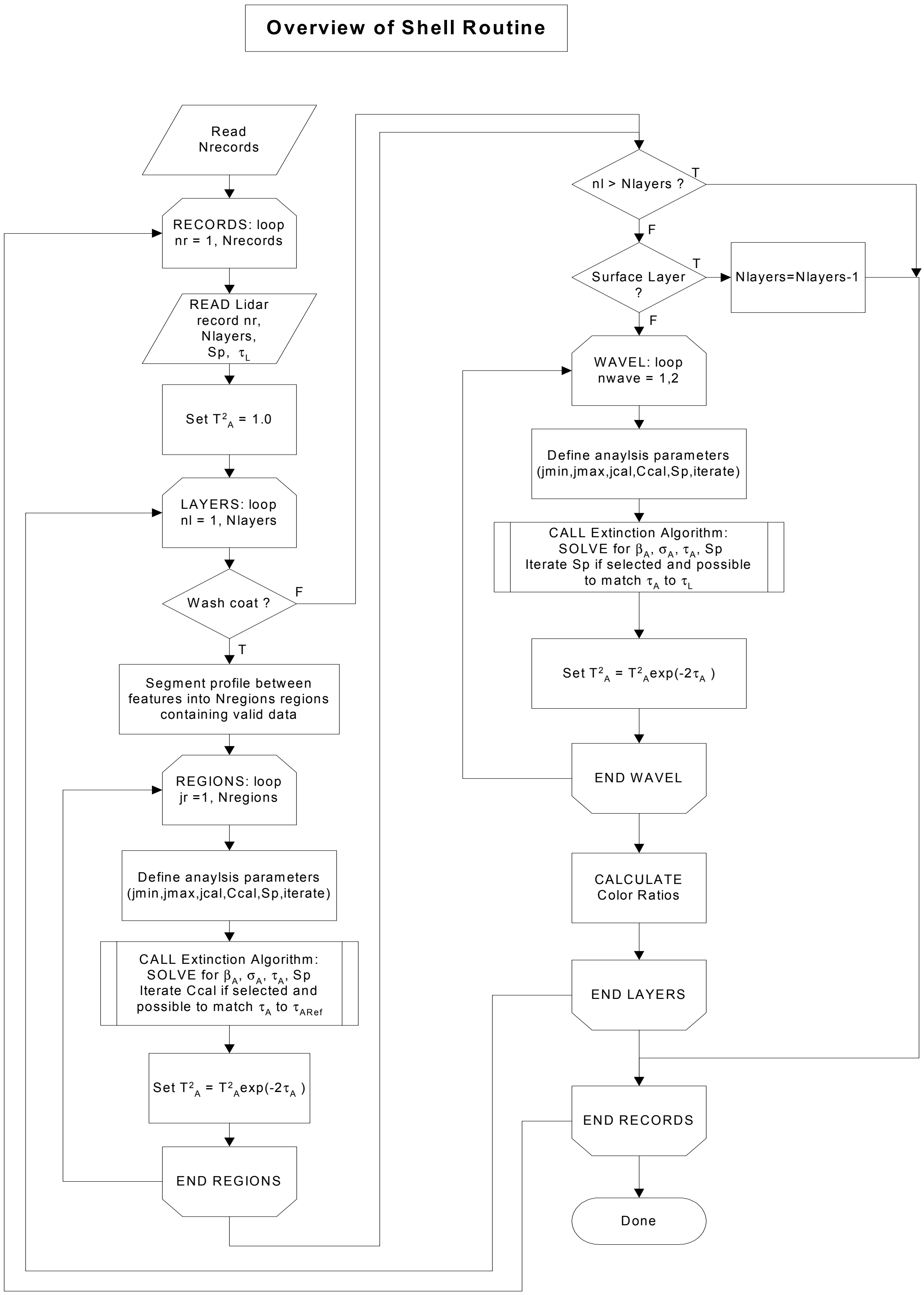



**Figure 2: Overview of shell subroutine (SOLVER) that sets up analysis parameters (jmin, jmax, Ccal, Sp, $\tau_L$ and iterate, the minimum, maximum and normalization (calibration) range indices, the normalization factor, layer optical thickness, lidar ratio and iteration switch) and calls extinction algorithm subroutines to solve for the particulate backscatter, extinction and optical thickness, $\beta_A$, $\sigma_A$.and $\tau_A$ , respectively)**

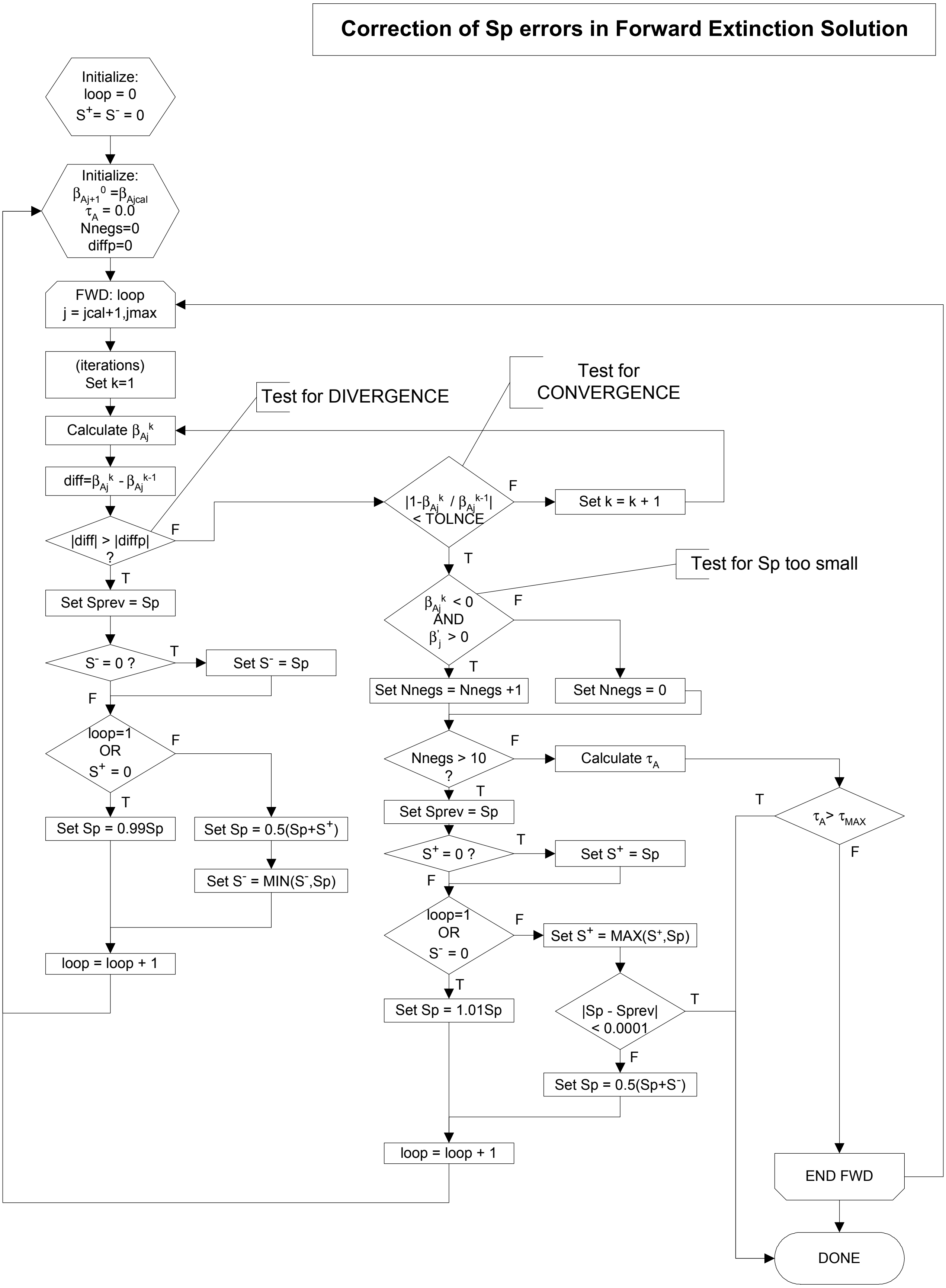



**Figure 3: Correction of errors in the lidar ratio $S_p$ (clouds or aerosols) in the extinction retrieval algorithm routine. $S_{prev}$ represents the value of the lidar ratio at the previous step. $S^+$ and $S^-$ represent the previously increased or decreased value of the lidar ratio.**

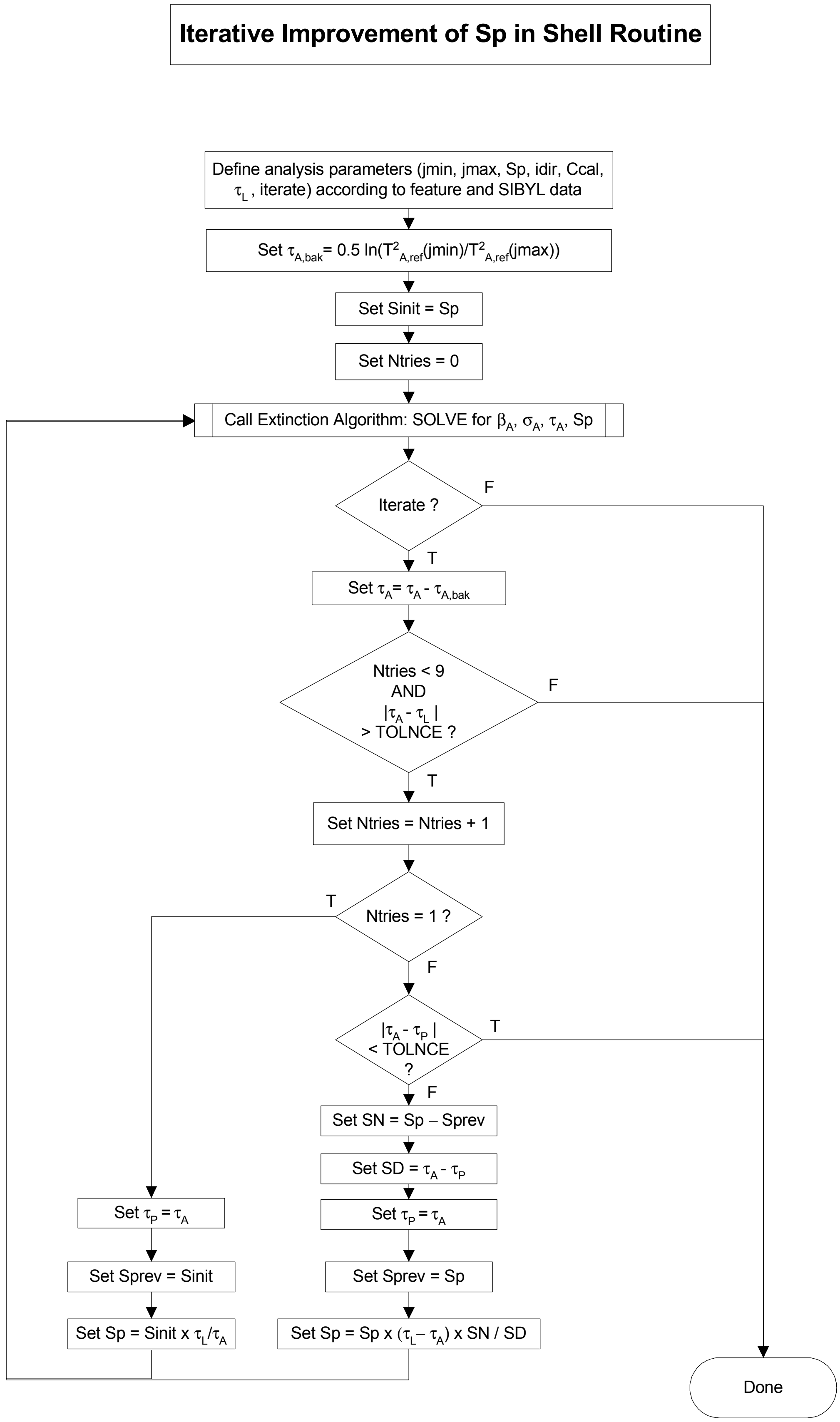



**Figure 4: The iterative improvement of the lidar ratio for features, Sp, in SUBROUTINE SOLVER by comparing the calculated optical thickness, $\tau_A$, with that measured for the layer, $\tau_L$.**

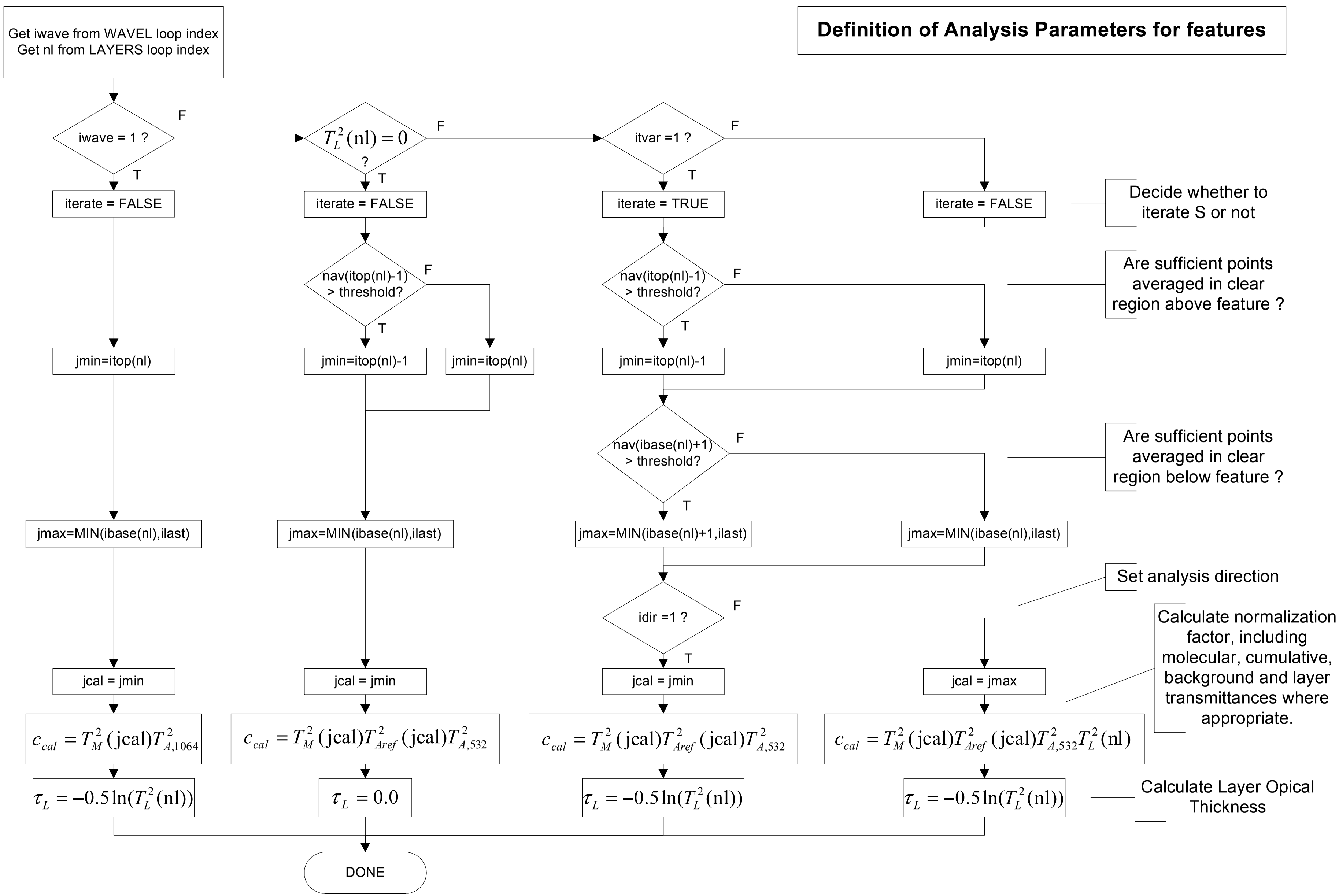



**Figure 5: Definition of analysis parameters (See Figure 2 and text for explanation of symbols)**

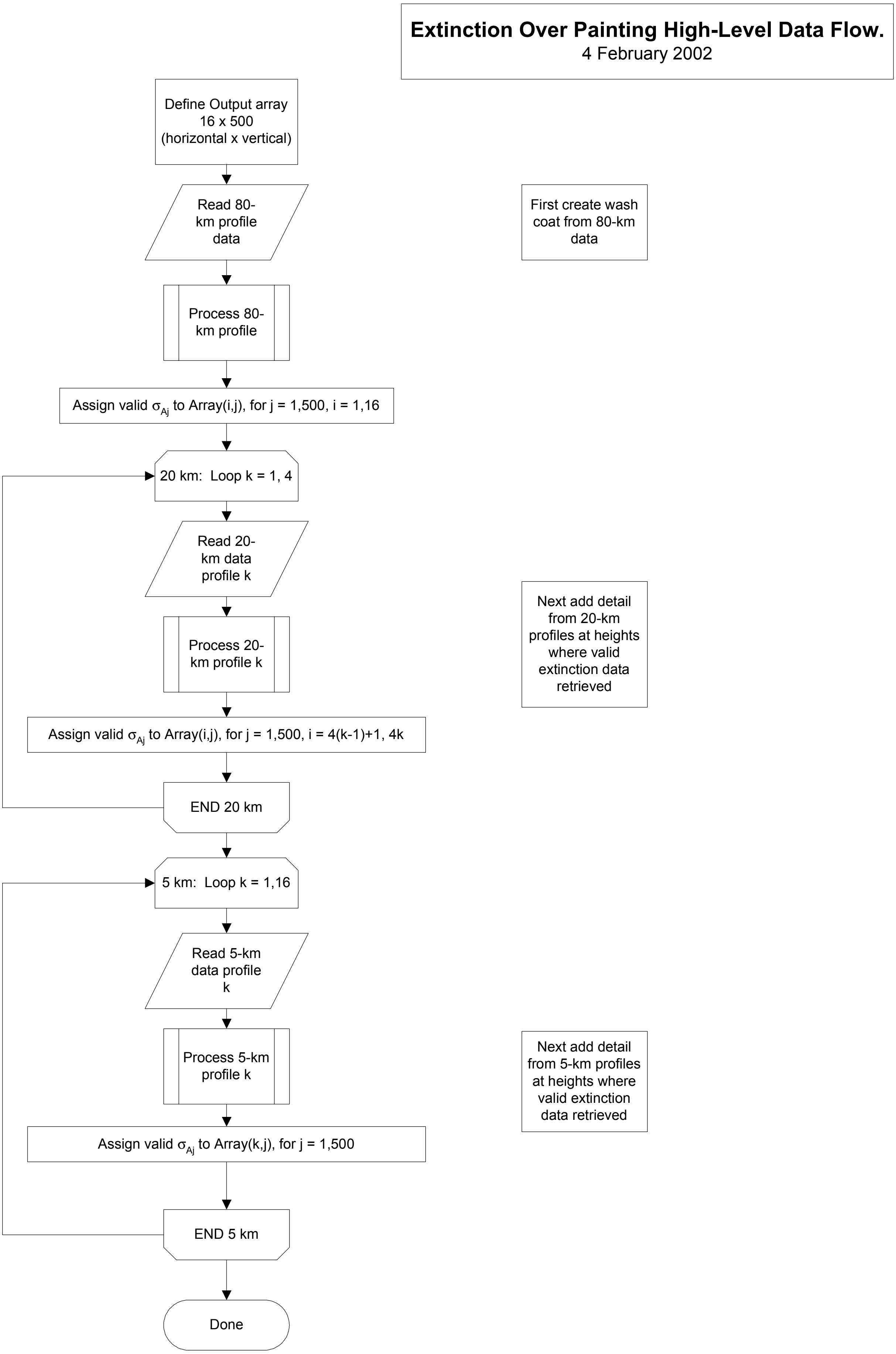



**Figure 6:** Overview of Extinction Over Painting Technique (eOPT).

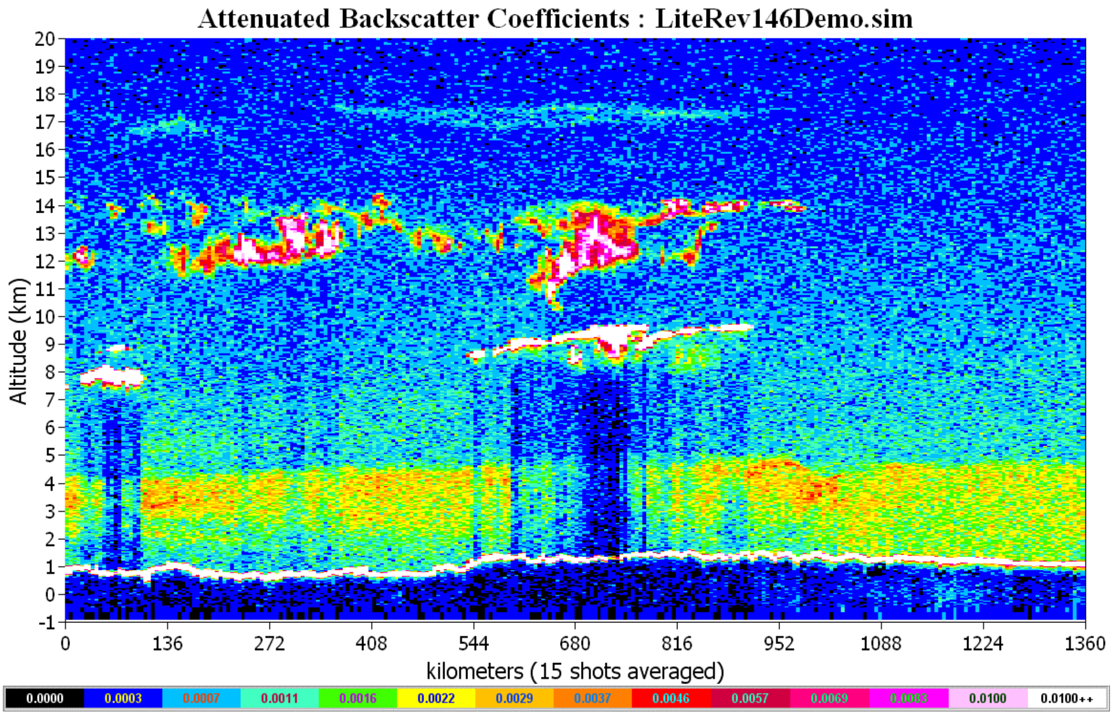



**Figure 7(a): LITE data modified to simulate CALIPSO data**

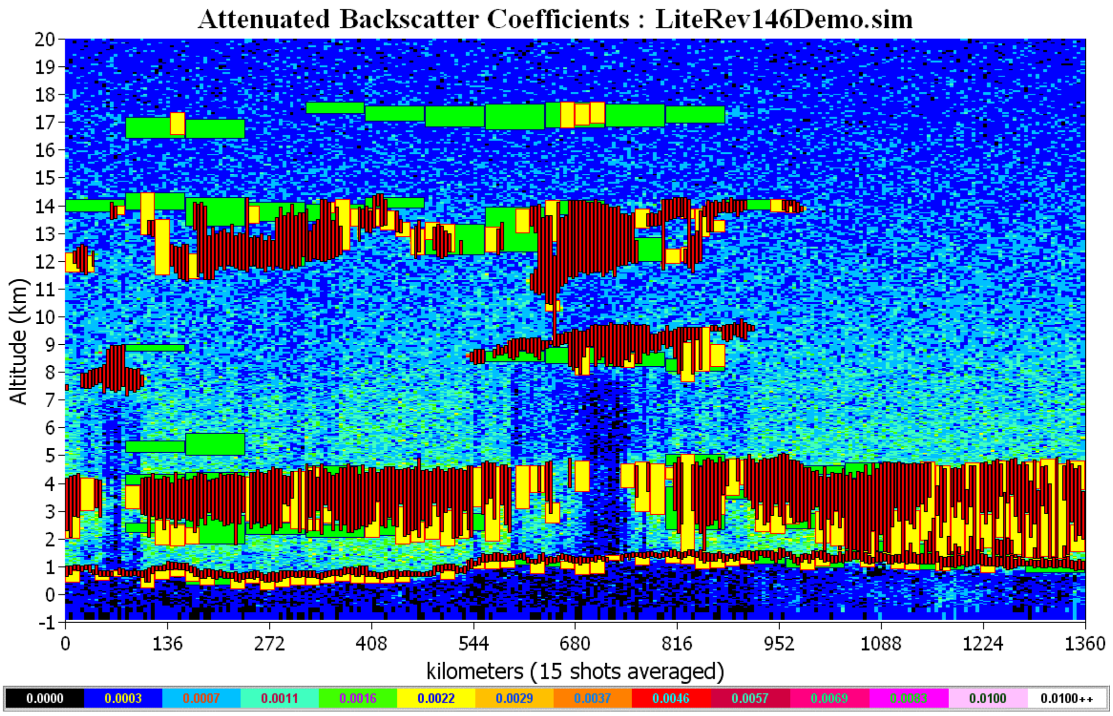



**Figure 7(b): The same scene as in 7(a) with features detected by SIBYL at horizontal resolutions of 80 km, 20 km and 5 km indicated in boxed regions colored in green, yellow and brown respectively.**

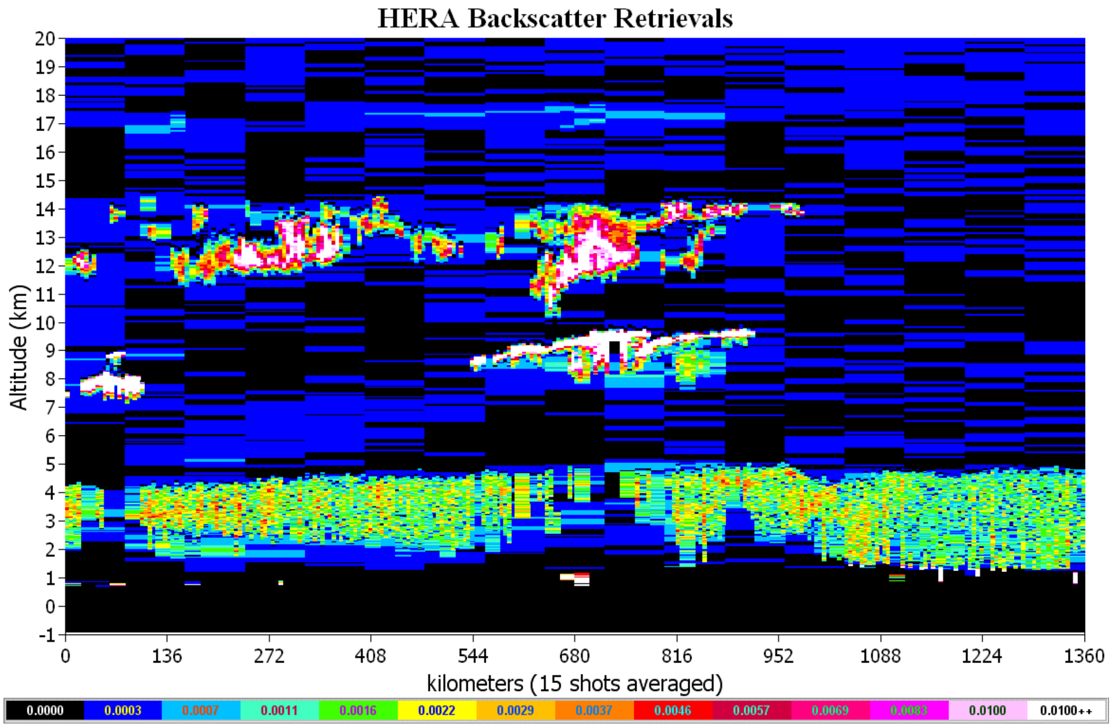



**Figure 7(c): As in 7(a) and 7(b) after analysis by the HERA and eOPT toproduce values of aerosol backscatter (km.sr)$^{-1}$.**

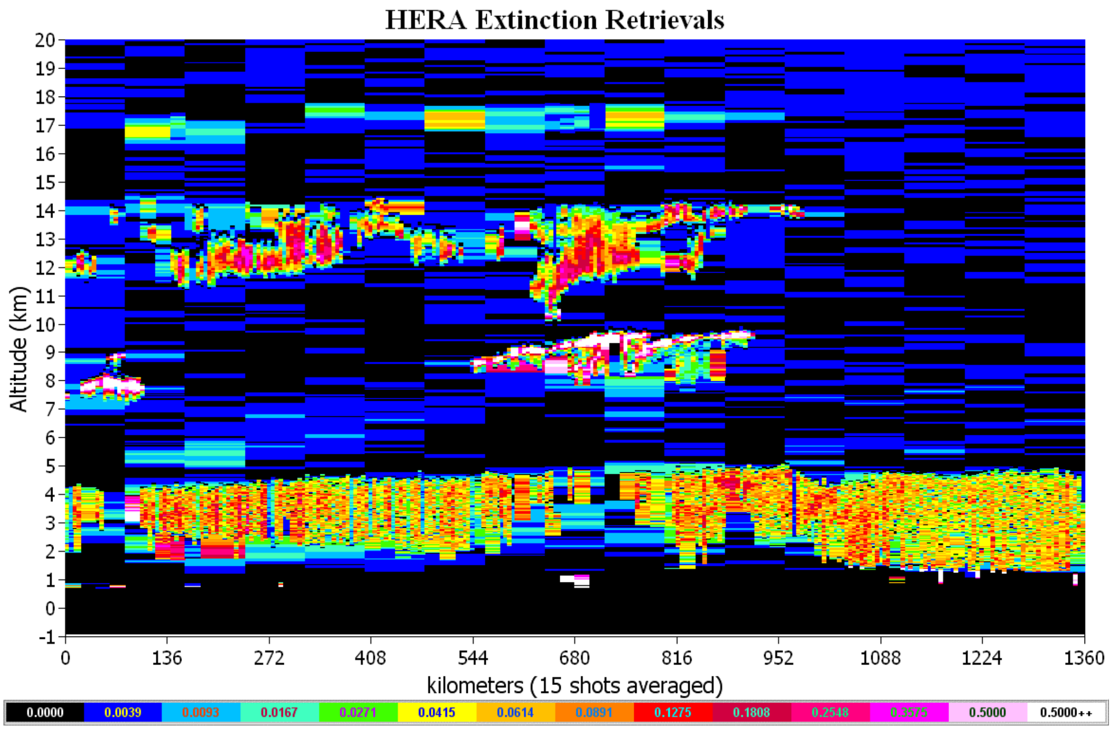



**Figure 7(d): As in 7(a) and 7(b) after analysis by the HERA and eOPT Routines to produce values of aerosol extinction coefficient (km)$^{-1}$**