

Micropulse Lidar Cloud Mask (MPLCMASK) Value-Added Product for the Fast-Switching Polarized Micropulse Lidar Technical Report

D Flynn
J Comstock

C Sivaraman
D Zhang

December 2020



DISCLAIMER

This report was prepared as an account of work sponsored by the U.S. Government. Neither the United States nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

Micropulse Lidar Cloud Mask (MPLCMASK) Value-Added Product for the Fast-Switching Polarized Micropulse Lidar Technical Report

D Flynn
C Sivaraman
J Comstock
D Zhang
All at Pacific Northwest National Laboratory

December 2020

Work supported by the U.S. Department of Energy,
Office of Science, Office of Biological and Environmental Research

Acronyms and Abbreviations

ADI	application data interface
ARM	Atmospheric Radiation Measurement
ARSCL	Active Remote Sensing of Clouds
LDR	linear depolarization ratio
ML	machine learning
MPL	micropulse lidar
MPLCMASK	micropulse lidar cloud mask
MPLPOLFSSHIPCOR	MPL Corrected for Ship Motion
NRB	normalized relative backscatter
NSA	North Slope of Alaska
QC	quality control
SGP	Southern Great Plains
SNR	signal-to-noise ratio
SONDE	balloon-borne sounding system
TWP	Tropical Western Pacific
VAP	value-added product

Contents

Acronyms and Abbreviations	iii
1.0 Introduction	1
2.0 Cloud Detection Algorithm	1
2.1 Algorithm Basic Steps.....	2
2.2 Algorithm Additional Steps	2
3.0 Post-Cloud Detection Processing	3
3.1 Creating Cloud Mask and Screening for Noise Clutter.....	3
3.2 Signal Attenuation.....	4
4.0 MPL Processing.....	4
4.1 The MPL	4
4.2 MPL Normalized Relative Backscatter and LDR	4
4.3 SNR.....	7
4.4 MPL Instrument Correction Options Default Setting	9
5.0 The Input Data.....	10
6.0 Flow Chart.....	10
7.0 Output Data	12
8.0 Summary and Future Work	14
9.0 References	15
Appendix A – Release History and Updates.....	A.1
Appendix B – Version Comparison.....	B.1
Appendix C – Applying Afterpulse Correction	C.1

Figures

1 MPLCMASK flowchart	12
2 Example plots of NRB, LDR and the cloud mask for MPLCMASK for the ARM Southern Great Plains (SGP) observatory, October 18, 2018.....	14

Tables

1 Release history and updates.....	A.1
------------------------------------	-----

1.0 Introduction

Lidar backscattered signal is routinely used for identifying vertical cloud structure in the atmosphere. Cloud boundaries derived from lidar signals are a necessary input for popular U.S. Department of Energy Atmospheric Radiation Measurement (ARM) user facility data products, such as the Active Remote Sensing of Clouds (ARSCL) Value-Added Product (VAP). An operational cloud boundary algorithm (Wang and Sassen 2001) has been implemented for use with the ARM micropulse lidar (MPL) systems. In addition to retrieving cloud boundaries above 0.5 km, the Micropulse Lidar Cloud Mask (MPLCMASK) VAP applies lidar-specific corrections to the measured backscattered lidar and calculates the linear depolarization ratio (LDR) using the methodology developed by Flynn et al., (2007). The current version of MPLCMASK applies the instrument corrections as described in Campbell et al., (2002) with modifications specific to a polarized fast-switch MPL system. The cloud boundaries output from MPLCMASK will be the primary lidar cloud mask for input to the ARSCL product.

Complete and accurate identification of cloud boundaries often requires the combination of multiple instrument measurements. Understanding the strengths and weaknesses of any one system is important when considering what type of clouds are being profiled. In the case of the lidar, the short wavelength and eye-safe low power requirement of most operational ground-based systems can limit the complete detection of optically thick clouds. While radar systems are able to penetrate these clouds, the lidar signal may be attenuated prior to the true cloud top and may also be unable to detect any cloud layers above an optically thick cloud. Lidar measurements are, however, very effective at determining cloud structure of optically thin clouds and are more sensitive to smaller cloud particles than other ground-based profilers, such as radar. Strengths of this lidar sensitivity include the identification of clouds in the presence of large rain droplets, supercooled water clouds with small water droplets, and cold ice or mixed-phase clouds with low concentrations of small ice particles. The ARM ARSCL data product combines lidar and radar as an effective synergy to provide a more complete cloud product than any one system by itself. In general, though, lidar measurements can be used to detect most mid- to high-level clouds as well as optically thin low clouds.

2.0 Cloud Detection Algorithm

The primary derived quantity determined in this VAP is the vertical location of cloud layers as clouds pass over the lidar system. The aim of the cloud detection algorithm in the VAP is to identify the number of cloud layers as well as each layer's cloud top and cloud base in a given MPL instrument corrected lidar profile. A detailed description of the processing of the MPL signal to provide corrected lidar profiles is provided in section 4.0 of this report. Lidar profiles for MPLCMASK consist of 30-second time-averaged and 30-meter vertical height resolution-averaged data. A cloud mask for a day of data is then compiled at this temporal and spatial resolution.

The cloud detection algorithm applied in MPLCMASK is based on Wang and Sassen (2001). This algorithm uses the change in the slope of the backscattered lidar signal as a function of height to identify the presence of cloud layers in the atmosphere, and uses several techniques to distinguish aerosol layers, virga, and drizzle from cloud layers. The algorithm is applied between the surface and 20 km above ground level; however, the cloud products currently provided in this VAP are between 0.5 km and 20 km.

The uncertainty of the instrument corrections applied to MPL in the near-range data is the main factor for this truncation. This issue is discussed further in section 4.2.

2.1 Algorithm Basic Steps

The algorithm has five basic steps:

1. Compute Necessary Algorithm Factors.

The slope of the lidar signal and the standard deviation (σ) of the background noise are computed using range-uncorrected data.

2. Find Potential Cloud Layers.

Beginning at the surface, the computed slope of the lidar signal is examined upward to identify and record possible layers. A strong positive slope indicates a potential layer. Layers are identified by marking the height of the base, peak, and top for each layer. The layer base height corresponds to the location where the strong positive signal slope begins; layer top height corresponds to the location where the signal returns to the molecular backscatter level or noise level; the peak height corresponds to the location of the strongest backscattered signal bounded by the layer.

3. Distinguish Cloud from Aerosol or Noise.

The cloud layer is distinguished from an aerosol layer or random noise by examining the physical characteristics of the signal. Aerosol layers are distinguished from clouds primarily because aerosol layers have smaller extinction or backscatter coefficients than clouds. The algorithm uses the ratio of the peak normalized backscattered signal in the layer over the backscattered signal at the layer base to distinguish cloud layers. As cited in Wang and Sassen (2001), a threshold value of 4 is appropriate for cloud layers 5 km and below, while a threshold of 1.5 is used for layers above 5 km.

4. Determine if Cloud top is Actual or Effective.

The algorithm examines the cloud-top signal to determine if the layer top is the actual cloud top or an effective cloud top. If the lidar signal becomes attenuation-limited in optically thick clouds, the cloud top is defined as the “effective” top. The effective top is identified when the signal above cloud top is (1) below the minimum reliable signal and has a zero mean slope, and (2) there is strong attenuation (according to the negative slope) detected within the cloud. To be considered an actual cloud top, the slope within the cloud must be between -0.2 and -1.2.

5. Distinguish Cloud Base from Virga or Drizzle.

Cloud base is distinguished from virga or drizzle below clouds by examining the slope below the layer base. The signal slope in virga is usually much smaller in magnitude than the slope in cloud signal. Dense water clouds, which often have drizzle or virga below, tend to have a large positive slope above the virga layer, followed by a strong negative slope as the signal becomes attenuated in the cloud.

2.2 Algorithm Additional Steps

Within the cloud detection algorithm, additional steps can (1) apply range correction to signal in cases to identify cloud layer in the presence of subcloud precipitation, virga, or aerosol (2) extend the cloud base or top of an identified layer, or (3) merge multiple layers into a single layer.

1. Detecting Cloud Base with Subcloud Interference.

In the presence of subcloud precipitation, virga, or aerosol, the range correction is applied to the lidar measurement to attempt to determine a change in the signal slope that would be due to a cloud-layer base. This is only attempted for potential cloud layers in the first 5 km.

2. Extend Cloud Base and Top.

If a cloud-layer base is greater than 1 km but less than 5 km, the region 300 meters below the identified base is reassessed to consider adding it to the layer. The assessment requires the slope or the peak-to-base ratio of the lidar signal to meet criteria to be considered cloud but not both. This approach is to determine if there is a weaker part of the cloud layer within this region. If the corresponding cloud top of this layer is greater than 5 km, the region 300 meters above the identified top of this layer is also reassessed to consider adding it to the layer. The assessment is again based on an adjusted criteria for the slope and peak-to-base ratio. For the cloud base, the thresholds are a ratio greater than 3 or a slope less than -4. For the cloud top, the thresholds are a ratio greater than 1.4 or a slope less than -3.5.

3. Merging Multiple Layers.

If a weak cloud layer is detected within 60 m above a strong layer and the signal strength is reasonable (not attenuated to near noise levels), the weak layer and strong layer are merged. This step is included to handle clouds with multiple signal peaks. If multiple cloud layers are detected near each other and the distance between one layer's cloud top and the other's cloud base is within 500 meters, the layers are merged together as a single layer.

3.0 Post-Cloud Detection Processing

3.1 Creating Cloud Mask and Screening for Noise Clutter

The cloud detection algorithm outputs the number of cloud layers as well as each layer's cloud top and cloud base. As mentioned previously, this version of the MPLCMASK limits the output to cloud retrievals above 500 m. A cloud mask is produced from this output. Each 30-meter, 30-second resolution bin is assigned a value of either '0' for clear or '1' for cloud.

MPLCMASK then runs a cluster test on all bins identified as cloud above 10 km to detect and attempt to eliminate any small false clouds or clutter. An isolated cloud layer reported by the detection algorithm that is above 10 km and consists of only a few range or time resolutions bins is considered likely a false cloud that erroneously met the cloud criteria in the algorithm. This can occur when the signal-to-noise ratio (SNR) of the lidar signal is poor and near the threshold where cloud and non-cloud can be distinguished.

The cluster test applied uses the principle of point density and inspects a five-by-five grid of bins surrounding any bin identified as cloud that is above 10 km. The central cloud bin within the grid is reassigned '0' for clear if two conditions are met. The conditions required are (1) less than seven of the 24 resolution bins within the grid are cloud and (2) less than four of these seven cloud bins are not at the same time or height (range) as the cloud bin being inspected. After reviewing all cloud bins above 10 km and reassigning any bins identified as clutter to clear, the cloud-layer base and heights are recalculated.

3.2 Signal Attenuation

To determine if the lidar signal above a cloud is likely attenuated, a ratio of the rolling average and rolling standard deviation of the lidar signal is computed. This ratio uses a window of size 13 times steps and is a measure of “scene variability”. If the scene variability is less than a value of 1 above the cloud top, the lidar signal is flagged as attenuated.

4.0 MPL Processing

4.1 The MPL

MPLCMASK uses lidar data from the elastic polarized fast-switch micropulse lidar (MPL) systems deployed by ARM. Polarized MPLs operate by emitting low-energy laser pulses ($\sim 10 \mu\text{J}$) with high pulse rate (2500 Hz) in an alternating linearly (crosspol) and circularly polarized (copol) state. Each return signal that is scattered back is collected using a photon detector and is time-resolved to provide the vertical height of atmospheric features, such as cloud droplets or aerosol, which are responsible for the scattering or attenuation of the emitted pulses. These return signals measured by the photon detector are processed and combined to then provide vertical profiles of the atmospheric structure as attenuated backscatter. Additionally, the linear depolarization ratio (LDR) can be calculated from the linear and circularly return signals to help determine cloud phase (liquid, ice, or mixed) as well as distinguishing aerosol and clear air. A detailed description of the MPL data processing based in part on Campbell et al., (2002) and the calculation of the LDR provided in Flynn et al., (2007) is provided in the next section of this report.

Data collected by the MPL fast-switch polarized lidar are provided in the ARM datastream `mplpols`. Typically the operational ARM MPL polarized fast-switching systems collects lidar data at 10-second and 15-meter time and range resolutions. A notable exception is ship deployment MPL data. It is collected at a time resolution of 1 second; however, it is processed by the MPL Corrected for Ship Motion (MPLPOLFSSHIPCOR) VAP to the typical operational resolution prior to being provided as an input for cloud detection.

In MPLCMASK, the instrument corrections are applied to both polarization measurements at full resolution, after which the measurements are combined and averaged to a reduced 30-second and 30-meter resolution. The cloud detection algorithm accepts this adjusted resolution data and maintains this same resolution for its cloud detection output products. The applied instrument correction that are provided in the VAP output are degraded to the adjusted resolution as well. This form of the instrument corrections is intended for reference purposes only. The MPLCMASK’s degraded MPL resolution allows improved lidar signals through averaging as well as providing the appropriate temporal and spatial resolution ARSCL requires for producing combined lidar and radar cloud masks.

4.2 MPL Normalized Relative Backscatter and LDR

The cloud detection algorithm in MPLCMASK employs both a range-corrected and a range-uncorrected normalized relative backscatter (NRB). Described in this section is the range-corrected NRB.

From the lidar signal, NRB is calculated (Campbell et al. 2002) as

$$NRB = \frac{(n(r)Dtc[n(r)] - n(r)_{ap} - n_b)r^2}{O(r)E}$$

Where

$n(r)$	MPL Raw Counts
$Dtc[n(r)]$	Detector Dead-Time Correction
$n(r)_{ap}$	Afterpulse Correction
n_b	Background Correction
r^2	Range Squared
$O(r)$	Overlap Correction
E	Energy

Additionally, NRB can be expressed as

$$NRB = C B(r)T(r)^2$$

With

C	Calibration
$B(r)$	Backscatter Coefficient
$T(r)^2$	Atmospheric Round Trip Transmittance

And

$$T(r)^2 = \exp \left[-2 \int \sigma(r) dr \right]$$

With

$\sigma(r)$	Extinction Coefficient
-------------	------------------------

Note that NRB is a “backscattered” lidar signal (also called attenuated backscatter) but as shown is different from lidar backscatter coefficient.

Because the input MPL datastream for this version of MPLCMASK is a combination of the linear and circular polarization measurements, additional steps and modifications to the lidar equation are necessary to calculate the non-range-corrected NRB. Non-range-corrected specifically means the inverse range squared corrected is not applied. For the polarization MPL system, detector dead time, afterpulse, and background are distinct corrections for each of the two signals while the overlap correction, energy normalization, and range corrections applied are the same. These corrections are also unique for each MPL system and can change when a hardware component is changed or the optics of the lidar are adjusted.

The dead time is the first polarization-specific correction to be applied and is followed by the afterpulse correction. The background correction is calculated from the dead time and afterpulse-corrected raw counts and is thus the last of the three polarization distinct corrections to be applied.

The detector dead-time correction is provided by the vendor and is a scale factor directly related to the raw counts. The afterpulse is measured experimentally for each signal and is a range-dependent value subtracted from the detector-corrected measurements. The background correction is a non-range-dependent constant that is determined by averaging the 7 km of the last 10-km region (the highest altitude range) of each dead-time-corrected and afterpulse-corrected measurement profile. The last 3 km of the highest altitude is ignored because in some MPL data sets, such as those from ship deployments, lidar signal is represented as missing (-999) when the data is corrected for ship tilt. The background constant value calculated is subtracting from each range of both polarization measurements.

The polarized measurement specific corrections can be outlined as follows

$$n(r)'_{linear} = n(r)_{linear} * DTC(n(r)_{linear}) - n(r)_{linear\ ap}$$

$$n(r)'_{circular} = n(r)_{circular} * DTC(n(r)_{circular}) - n(r)_{circular\ ap}$$

$$n(r)''_{linear} = n(r)'_{linear} - n(r)_{linear\ b}$$

$$n(r)''_{circular} = n(r)'_{circular} - n(r)_{circular\ b}$$

Where

$n(r)'_{linear}$ and $n(r)'_{circular}$ are the detector dead-time- and afterpulse-corrected raw counts.

$n(r)''_{linear}$ and $n(r)''_{circular}$ are the detector dead-time-, afterpulse-, and background-corrected raw counts.

At this point the LDR can be calculated (Flynn 2007) as:

$$\frac{n(r)''_{linear}}{n(r)''_{linear} + n(r)''_{circular}}$$

A description of the SNR calculation for LDR as well as the SNR for each of the two linear and circular corrected lidar measurements components is provided in the next section of this report.

To continue the calculation of NRB, the two polarization-specific corrected signals need to be combined to give the total signal.

$$n(r)''_{combined} = 2 \times n(r)''_{linear} + n(r)''_{circular}$$

Lastly, the overlap correction and energy normalization are applied to the combined signal to give non-range-corrected NRB as

$$\frac{n(r)''_{combined}}{O(r)E}$$

The overlap is a field of view issue where the lidar signal overfills the area of the detector because of the divergence of the lidar laser. This correction is a geometric effect, is most significant in the nearest range, and converges to a correction factor of one within the first several kilometers of data for the MPL. This correction is determined experimentally and the accuracy is important because errors can introduce false features or mask cloud layers. If the overlap correction is carefully assessed, cloud detection in the first several 100 meters can be pursued more confidently; however, for this version of MPLCMASK the cloud detection is limited to above 500 meters. The energy normalization reflects the output energy of the laser pulse and is a nominal correction. It can vary from one pulse to the next but typically is within a few percent of a daily mean.

4.3 SNR

The signal-to-noise ratio (SNR) is by definition the ratio of a particular quantity identified as the “signal” over the “noise” in that same quantity computed over some averaging interval. Calculating the SNR is helpful for identifying when the measured signals are meaningful and significant. In MPLCMASK the SNR is calculated for both NRB and for the LDR. To achieve this, the (1) noise and (2) SNR for each polarization channel are calculated separately followed by the (3) effective linear copol noise and effective linear copol SNR and lastly (4) the LDR SNR. The effective linear copol noise and SNR are the appropriate form of these statistics to characterize NRB derived from polarized MPL data. Calculations for the four fields, as functions of time (t) and range (r), are outlined below.

1. Copol (circular) and Crosspol (linear) Noise.

The statistical noise of counts from a photon counter are known to follow Poisson statistics where the noise scales as the square root of the total counts. The MPL signals in each bin are reported as average counts per microsecond per laser pulse so the expression below incorporates a conversion of these averaged values to total counts, takes the square root, and then converts back to average counts per microsecond per laser pulse.

$$\begin{aligned} \text{polavg_co_nse [t, r]} &= \sqrt{\text{mpl_co[t, r]} * \text{cnv}/\text{cnv}} = \sqrt{\text{mpl_co[t, r]} / \text{cnv}} \\ \text{polavg_cross_nse [t, r]} &= \sqrt{\text{mpl_cross[t, r]} * \text{cnv}/\text{cnv}} = \sqrt{\text{mpl_cross[t, r]} / \text{cnv}} \end{aligned}$$

polavg_co_nse [t, r]: polarized average copol (circular) noise term

polavg_cross_nse [t, r]: polarized average crosspol (linear) noise term

mpl_co [t, r]: mpl_co[t,r]: corrected copol (circular) MPL signal
 mpl_cross [t, r]: mpl_cross[t,r]: corrected cross (linear) MPL signal
 cnv: cnv: conversion term

The conversion term is calculated as

$$\text{cnv} = \text{range_bin_time} * \text{shots_summed}.$$

range_bin_time: the time in seconds for each range bin
 shots_summed: the number of lidar pulses summed

The relationship between range_bin_width (width of a range bin in kilometers) is

$$\text{range_bin_time} = 2 \times (\text{range_bin_width} / \text{speed of light}).$$

Where the speed of light is ~299792 kilometers/second and multiplying by 2 to account for the round-trip path so, for example, a range_bin_time of 200 ns is equivalent to a 30-meter range bin width.

The shots_summed can be calculated as

$$\text{shots_summed} = (\text{pulse_rep} \times \text{average_interval}) / 2$$

where

pulse_rep: the repetition rate, or trigger frequency of the laser in Hz
 average_interval: in seconds

The shots summed is split between the two polarization modes and is therefore divided by two.

2. Copol (Circular) and Crosspol (Linear) SNR.

$$\text{polavg_co_snr} [t, r] = \text{mpl_co} [t, r] / \text{polavg_co_nse} [t, r];$$

$$\text{polavg_cross_snr} [t, r] = \text{mpl_cross} [t, r] / \text{polavg_cross_nse} [t, r]$$

where

polavg_co_snr [t, r]: polarized average linear (copol) snr term
 polavg_cross_snr [t, r]: polarized average circular (crosspol) snr term

3. Effective Linear Copol Noise and SNR.

$$\text{co_lin_nse} [t, r] = \sqrt{\text{polavg_co_nse}[t, r]^2 + \text{polavg_cross_nse}[t, r]^2}$$

$$\text{co_lin_snr} [t, r] = (\text{mpl_co} [t, r] + \text{mpl_cr} [t, r]) / \text{co_lin_nse}[t, r]$$

where

co_lin_nse [t, r]: effective linear copol noise.

co_lin_snr [t, r]: effective linear copol snr

4. LDR SNR.

$$\text{polavg_ldr_snr}[t][ht] = 1/\sqrt{1/\text{polavg_cross_nse}[t,r]^2 + \sqrt{1/\text{co_lin_snr}[t,r]^2}}$$

where

polavg_ldr_snr[t][ht]: polarized average linear depolarization ratio

4.4 MPL Instrument Correction Options Default Setting

The instruments corrections necessary to calculate NRB from the raw MPL data are generally provided in the mplpolfs datastream. The current version of MPLCMASK has the option to use these corrections or alternatively to receive some of the corrections as external files at the time the VAP is run. For the afterpulse correction, there is also the option to exclude it from the MPL processing. Historically, afterpulse corrections have not always been measured or been available in the mplpolfs datastreams and most recent versions of MPLCMASK have not applied it. Beginning in 2016, regular collection of this correction was instituted at all of the ARM sites on a quarterly basis. An afterpulse correction can also be determined from MPL instrument data when conditions exist where the lidar signal is fully attenuated by low optically thick clouds or precipitation, or when the view of the MPL is blocked during cleaning. It is important that the afterpulse correction only be included if it can be demonstrated to accurately characterizes the afterpulse of the system. An inaccurate correction can result in a nonphysical negative lidar signal or can cause an error in the slope of the lidar signal that interferes with distinguishing clear-sky, cloud, and attenuated signal. A preliminary analysis comparing the impact of applying or not applying the afterpulse correction for cloud detection is included in Appendix C.

In MPLCMASK there are three options for applying the afterpulse:

1. Afterpulse correction applied using external file.
2. Afterpulse correction applied from mplpolfs datastream.
3. Default option: No afterpulse correction applied.

For the dead-time correction there are two options:

1. Dead-time correction applied using external file.
2. Default option: Dead-time correction applied from mplpolfs datastream.

The dead-time correction used in MPL processing is a table of counts and corresponding correction factors. These tables are provided by the detector manufacturer and are non-trivial to measure. The maximum observed counts varies from one system and to the next and depend particularly on the laser energy output. Prior to applying the dead-time correction, MPLCMASK calculates the maximum counts to determine if the manufacturer's table is sufficient. The first several range bins include laser flash and are ignored when determining the maximum count. If the MPL raw counts on a given day exceed the

maximum table value, the correction is extended by fitting the last three points of the table to a second-order polynomial and extrapolated the fit to the daily maximum observed raw signal.

5.0 The Input Data

The primary measurement required for MPLCMASK is NRB or attenuated backscatter measured by the MPL. Historically, the ARM user facility operated single-wavelength, non-polarized MPLs at all fixed sites and mobile facilities. In 2004, polarized MPL systems were installed at all sites and mobile facilities. In addition, polarized MPL systems were upgraded starting in June 2010 to fast-switching systems to allow improved switching capability between the linear and circular polarization channels. The current version of MPLCMASK will only process fast-switching polarized MPL data. Previous versions of MPLCMASK are maintained to manage all other historical datastreams.

MPLCMASK uses balloon-borne sounding system (SONDE) data when it is available to create a Rayleigh profile to estimate molecular scattering and extinction from temperature and pressure measurements. If the SONDE datastream is not available, a static molecular backscatter profile is used based on the atmospheric state variables for a standard atmosphere.

Micropulse Lidar Instruments

sssmplXX.b1 or sssmplpolavgXX.b1, sssmplpolfsC1.b1

Where:

- sss is the site and XX is the facility
- sgpmpolfsC1.b1 is the ingested data produced after the upgrade to fast-switching MPL system.

SONDE instruments:

ssssondewnpnXX.b1, sgpsondewnpnXX.a1

Where:

- sss is the site and XX is the facility

6.0 Flow Chart

This VAP has several source files and a single header file. The purpose of the header file is to define all the input fields and the output fields in a central location. This header file is included in the source file – `mplcmask1zvang.c` – which makes it easier to access all the fields defined in the header file. Similarly, all the external and internal functions are centrally defined in the header file.

The following steps are a brief description of the logic used in the VAP:

1. Read the input data based on the retriever file.
2. If the MPL data are not available, the VAP will abort and no output will be generated.
3. A 30-second interval sample set is established based on the command line input.

4. Establish the maximum height for the output to 20 km.
5. Apply dead time to copol and crosspol signals separately.
6. Subtract the afterpulse correction for each of the signals separately.
7. The corrections are either read from the input file or from a config file.
8. Calculate background using the corrected signal separately.
9. Combine the two signals (copol and crosspol).
10. Apply overlap correction.
11. Calculate “beam-blocked” conditions, which occur at sites where the solar zenith angle is small enough near solar noon that the disk of the sun viewed by the detecting telescope is covered. The telescope is covered during these times to avoid damaging the instrument. This procedure is performed primarily for MPLs located in the tropics and is turned off in mid- and high-latitude regions. If the beam is blocked, the corrected MPL signals are set to missing value.
12. To compute when the MPLs are covered (beam-blocked conditions), the solar zenith angle, the background signal at noontime, and the average background signal at nighttime are calculated. If the background signal during a 4-hour window around noon is less than the average background signal at nighttime, the corrected MPL signals are set to missing. The quality check (qc) flag for linear depolarization ratio, backscatter, and cloud mask were set to the appropriate bits.
13. Calculate backscatter profiles for the output resolution.
14. Calculate the signal-to-noise ratio of backscatter using a rolling window of 3 minutes without averaging the vertical. The average divided by the standard deviation of this window is the resulting signal-to-noise ratio.
15. Smooth the backscatter profiles.
16. If the sonde data are available, the Rayleigh backscatter is computed using the sonde temperature and pressure. If not, static molecular backscatter is used from standard atmosphere profiles.
17. The cloud top, cloud base, and cloud layer are calculated using the cloud boundary algorithm (Wang and Sassen 2001).
18. A cluster test is applied to remove false identification of cloud above 10 km that occurs when the backscatter is noisy.
19. Bad lidar times are flagged as missing based on input from the mentor or translator that identifies bad lidar profiles in the raw datastream if reprocessing is necessary.
20. The cloud boundaries are then recalculated after the above tests are performed.
21. If at any time the backscatter signal is not valid, it is replaced with a missing value (-9999.0), and the QC bit is set.
22. If the backscatter signal-to-noise ratio is below 1.0 between cloud top and 30 meters below a cloud top, then the attenuation flag is set to true.
23. As a final step, a single file for each day is written as output.
24. A quicklook plot is generated for backscatter profile, linear depolarization ratio, and cloud mask.

For more details on the C and Fortran code, please see the webpage for documentation:

<https://engineering.arm.gov/~sivaraman/docs/vap/mplcmask1zwang/html/>

Quality checks are applied for five variables. Since raw MPL signals can have a valid value of zero, the calculated backscatter or the linear depolarization ratio could result in NaNs due to division by zero. The NaNs are set to missing, and the appropriate QC bits are set.

Primary Output Variables:

1. linear depolarization ratio
2. cloud_mask
3. linear depolarization signal-to-noise ratio
4. backscatter
5. backscatter signal-to-noise ratio.

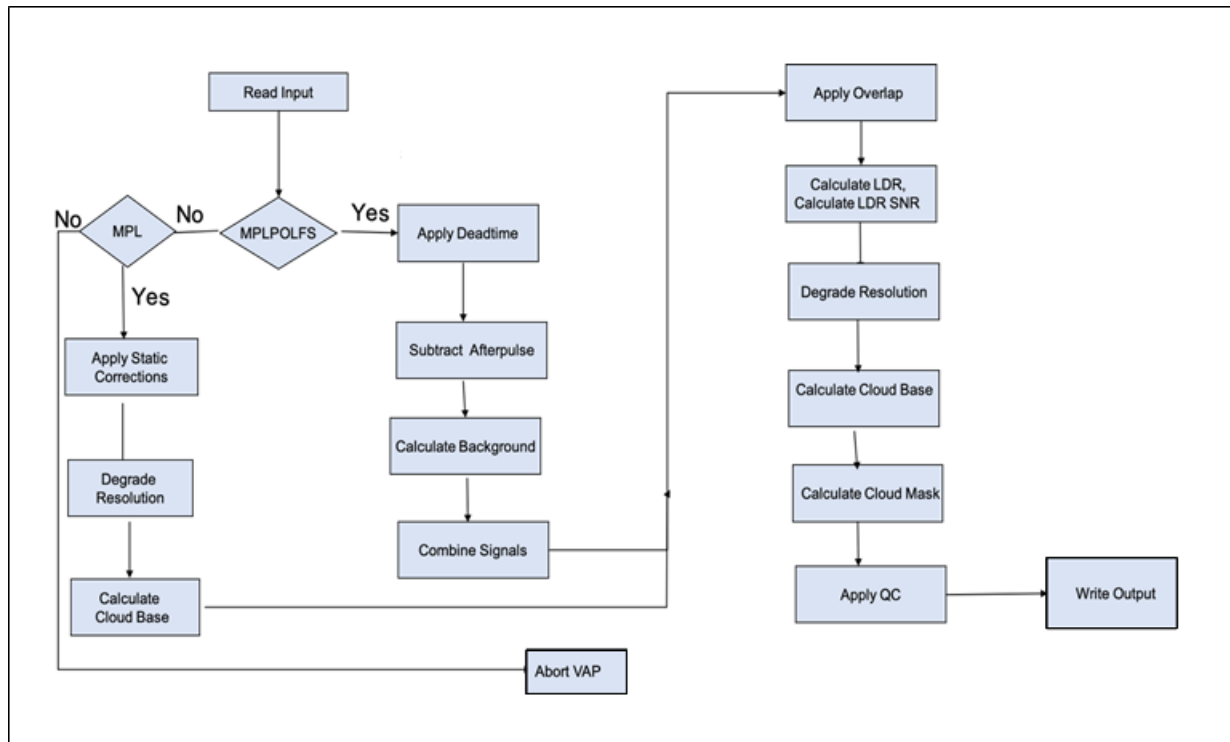


Figure 1. MPLCMASK flowchart.

7.0 Output Data

The MPLCMASK VAP produces one output file, named:

SSS30sMPLCMASK1ZWANGXX.c1.YYYYMMDD.hhmmss

Where:

- SSS is the site of the instrument (e.g., sgp)
- XX is the facility (i.e., C1, C2, etc.)
- YYYY is the year
- MM is the month of the year
- DD is the day of the month
- hh is the hour of the day
- mm is the minute of the hour
- ss is the second of the minute of data start.

base_time:

time_offset:	[1×2880	double]
time:	[1×2880	double]
time_bounds:	[2×2880	double]
height:	[667×1	single]
height_bounds:	[2×667	single]
cloud_base:	[1×2880	single]
cloud_top:	[1×2880	single]
num_cloud_layers:	[1×2880	int32]
linear_depolar_ratio:	[667×2880	single]
qc_linear_depolar_ratio:	[667×2880	int32]
linear_depolar_snr:	[667×2880	single]
qc_linear_depolar_snr:	[667×2880	int32]
cloud_mask:	[667×2880	int32]
qc_cloud_mask:	[667×2880	int32]
cloud_base_layer :	[50×2880	single]
cloud_top_layer:	[50×2880	single]
backscatter:	[667×2880	single]
qc_backscatter:	[667×2880	int32]
backscatter_snr:	[667×2880	single]
qc_backscatter_snr:	[667×2880	int32]
background_signal:	[1×2880	single]
cloud_top_attenuation_flag:	[1×2880	int32]
shots_summed:	[1×2880	single]
deadtime_correction_counts:	[16×1	single]
deadtime_correction:	[16×1	single]
overlap_correction:	[667×1	single]
afterpulse_correction_copol:	[667×1	single]
afterpulse_correction_cross:	[667×1	single]
lat:		
lon:		
alt:		

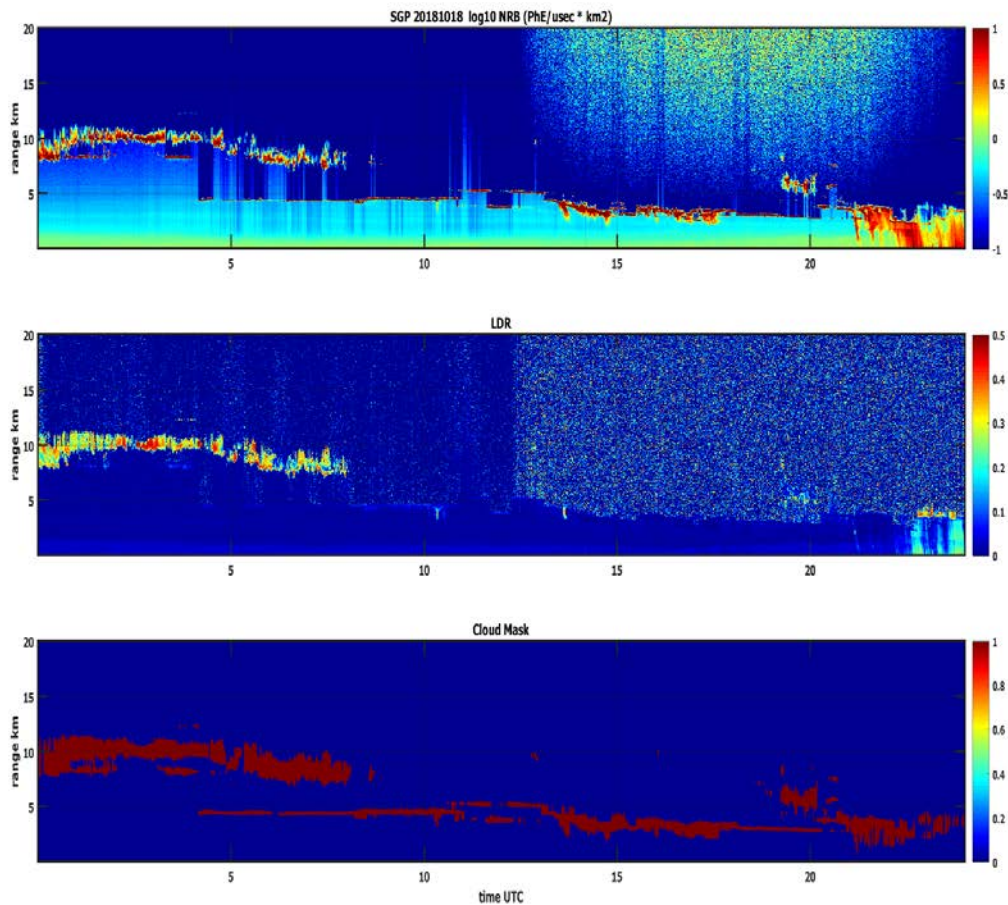


Figure 2. Example plots of NRB, LDR and the cloud mask for MPLCMASK for the ARM Southern Great Plains (SGP) observatory, October 18, 2018.

8.0 Summary and Future Work

The MPLCMASK VAP includes a cloud detection algorithm to provide cloud boundaries for cloud detected above 0.5 km and linear depolarization ratios (LDR). Currently, this VAP is used as input to the ARSCL VAP, which further combines the output of MPLCMASK with ARM vertically pointing radar reflectivity and ceilometer cloud base heights to provide a more complete cloud mask.

A strength of the cloud detection approach in MPLCMASK has been the use of a single set of thresholds within the algorithm. This feature has been important to allow the cloud output product to be provided autonomously and not require or rely on adjustments for different locations (deployment sites) or when an instrument is replaced. The cloud detection across different sites and different MPL systems is also reasonably robust even when the instrument performance varies or degrades. Full separation of the polarization channels has been an issue in some deployments of the MPL but because the cloud detection does not depend on the LDR calculations, cloud detection is still possible. While the fixed threshold is a strength and advantage as described, future work may want to test adjustment of thresholds for different

data sets to determine if improving the cloud detection accuracy outweighs this convenience. This is particularly true for boundary-layer clouds. Some discussion of this consideration is included in Appendix C.

Preliminary work has suggested that applying additional MPL instrument corrections (afterpulse) and/or improving some of the corrections (overlap and dead time) applied to the MPL signal could improve the accuracy of the cloud boundaries detected by the VAP's algorithm. A brief description of some of this work is included in Appendix B.

Lastly, efforts have begun to develop a machine learning (ML) MPL cloud detection VAP data product that uses the data products produced by MPLCMASK as input. This new VAP, tentatively called MPLCMASKML, employs the cloud masks produced by its predecessor to pre-train a model to identify clouds from the lidar NRB and LDR images. The model is then fine-tuned with further training using hand-labeled cloud masks. A description of the approach and results are presented in Cromwell and Flynn (2019). Evaluation data for a test period is expected to become available mid-2020. As both VAPs continue to develop, future work could consider a synergy of the MPLCMASK algorithm and the MPLCMASKML model to provide an MPL best cloud mask data product.

9.0 References

- Campbell, JR, DL Hlavka, EJ Welton, CJ Flynn, DD Turner, JD Spinhirne, VS Scott, and IH Hwang. 2002. "Full-Time, Eye-Safe Cloud and Aerosol Lidar Observation at Atmospheric Radiation Measurement Program Sites: Instruments and Data Processing." *Journal of Atmospheric and Oceanic Technology* 19(4): 431–442, [https://doi.org/10.1175/1520-0426\(2002\)019<0431:FTESCA>2.0.CO;2](https://doi.org/10.1175/1520-0426(2002)019<0431:FTESCA>2.0.CO;2)
- Clothiaux, EE, GG Mace, TP Ackerman, TJ Kane, JD Spinhirne, and VS Scott. 1998. "An Automated Algorithm for Detection of Hydrometeor Returns in Micropulse Lidar Data." *Journal of Atmospheric and Oceanic Technology* 15(4): 1035–1042, [https://doi.org/10.1175/1520-0426\(1998\)015<1035:AAAFDO>2.0.CO;2](https://doi.org/10.1175/1520-0426(1998)015<1035:AAAFDO>2.0.CO;2)
- Cromwell E, and DM Flynn. 2019. "Lidar Cloud Detection with Fully Convolutional Networks." In IEEE Winter Conference on Applications of Computer Vision 2019, 619–627. PNNL-SA-138402. [doi:10.1109/WACV.2019.00071](https://doi.org/10.1109/WACV.2019.00071)
- Flynn CJ., A Mendozaa, Y Zhengb, and S Mathurb. 2007. "Novel Polarization-Sensitive Micropulse Lidar Measurement Technique." *Optics Express* 15(6): 2785–2790, <https://doi.org/10.1364/OE.15.002785>
- Wang, Z, and K Sassen. 2001. "Cloud Type and Macrophysical Property Retrieval Using Multiple Remote Sensors." *Journal of Applied Meteorology and Climatology* 40(10): 1665–1682, [https://doi.org/10.1175/1520-0450\(2001\)040<1665:CTAMPR>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1665:CTAMPR>2.0.CO;2)

Appendix A

Release History and Updates

Table 1. Release history and updates.

Date	Version	Change Log
2011	0.1-0	Initial Release
2011	0.2-0, 0.3-0, 0.4-0	Updated to run at North Slope of Alaska (NSA) and Tropical Western Pacific (TWP) and other minor changes
2014	0.5, 0.6, 0.7	Application data interface (ADI) port
2019	1.0.0	Release with updates to MPL instrument corrections and LDR calculations.

Appendix B

Version Comparison

A year of data from SGP (20180901 to 20190830) was evaluated to compare the newest VAP release version 1.0.0 (2019) to previous VAP release 0.6 (2014) to assess the impact of updating the order of applying lidar corrections in release 1.0.0 that were not applied in the correct order in release 0.6. The evaluation was based on a comparison of the reported cloud base that is dimensioned against time. The cloud base field reports the lowest detected cloud base in kilometers above ground level or -1 for clear sky. The comparison was focused primarily on two questions:

1. Do the two versions agree when conditions are clear (cloud free)?
2. Do the two versions detect a similar cloud base?

The report of clear-sky conditions was first compared for both versions and the agreement was found to be just over 93% averaged over the entire year. On any given day the average agreement was slightly lower, at 90%. Overall, the new release tended to identify more clear sky than the older version and about 5% of the time reported clear sky when the older version reported at least one cloud layer. In contrast, the older version reported clear sky just over 2% of the time when the new version reported a cloud base. In principle, such disagreement may occur when only one version erroneously reports a cloud base where no cloud is present, or, alternatively, when only one version misses a cloud when a cloud is present. Based on detailed examination of time-height backscatter plots, however, most disagreement is due to missing or under-reporting actual clouds, not false cloud reporting.

The cloud bases reported for each of the two VAP versions were compared next. Agreement of cloud bases to within 60 meters (two range bins) was found to be 81% on average daily or 91% averaged over the entire year. Agreement to better than 120 m (four range bins) but less than 60 meters was found to be just over 1% for both the yearly and daily averages. The two versions therefore both report a cloud base but disagree on the height by 120 meters or more 18% of the time on average for any given day or 8% for the entire year. Examining the data, we note that typically if the two VAP versions both report a cloud base but disagree on the height by a moderate amount, there are multiple layers in the lidar profile and one of the version failed to identify the lower layer and instead reported the base of the second layer as its cloud base.

To further understand the differences between the two VAP versions, the days that showed the most significant disagreement between the clear-sky reporting and/or had cloud base height disagreement greater than 300 meters were identified for further scrutiny. The resulting subset of data was 140 days where on each of these days there were 100 or more instances (time bins) of significant disagreement in cloud or clear-sky reporting. Note that a full day of data consists of 2880 time bins at 30-second

resolution. Between September 2 and November 19, 2018, the dominant pattern that emerged within this selected data was that the new release version was detecting low boundary-layer clouds that the older version was not identifying. Between November 25, 2018 and August 27, 2019, however, the dominant pattern reversed and instead the old release version detected all or parts of low boundary-layer clouds that the new version did not identify. Corresponding with this shift in performance was an instrument swap on November 21, 2018 at SGP when MPL serial number 107 was swapped for 4212.

The shift in the performance of the two versions of the VAP is challenging to interpret, particularly with the instrument swap. We do not know what factor or factors might account for why the algorithm performance seemingly improved for MPL 107 but not for MPL 4212 when improved lidar NRB was provided as input. As mentioned previously, it can generally be said that most of the time if one version of the VAP detects a cloud base, then it is identifying a legitimate cloud that the other VAP version is failing to detect. If the year of data is broken into two subsets, we find for MPL 107 the newer version potentially missed 2% of the clouds compared with 6% for the old version and for MPL 4212 the newer version potentially missed 6% compared with 1% missed for the older version.

The accuracy of the instrument corrections and the sensitivity of the threshold in the algorithm are potential factors that could be contributing to the differences observed in the performances of the two versions of the VAPs for two different MPL systems. The detection of boundary-layer clouds is particularly sensitive to the MPL dead-time and overlap instrument corrections because the magnitude of these corrections is highest for the lower-altitude portion of the lidar signal. Distinguishing cloud features from other atmospheric features in the boundary layer relies on a strong cloud-peak-to-cloud-base ratio, which is one of the key algorithm thresholds. Instrument corrections that are not accurate or that are applied out of order could therefore have the potential to alter this ratio and either negatively impact the performance of the VAP or even inadvertently improve it. It is also unknown how well the thresholds in the algorithm are necessarily optimized for ideal MPL lidar data. While differences were identified between the two versions of the VAP, further evaluation of the accuracy of the instrument corrections and threshold tolerances might be necessary to fully understand the comparison.

Appendix C

Applying Afterpulse Correction

The current version of MPLCMASK includes the option to apply afterpulse corrections to the MPL backscatter prior to implementing the cloud detection algorithm. Preliminary evaluation of three months of SGP data (January 1, 2018 to March 31, 2018) suggests including the correction can provide a more accurate cloud top identification for both boundary-layer and upper-level clouds in some cases. Because the full impacts of including the afterpulse correction has not been thoroughly evaluated at different sites or for different systems, the current default mode of the VAP is to run without it. Data users may request a version of MPLCMASK output with the afterpulse applied.



U.S. DEPARTMENT OF
ENERGY

Office of Science