CLOUD COVER/LAYERS

VISIBLE/INFRARED IMAGER/RADIOMETER SUITE ALGORITHM THEORETICAL BASIS DOCUMENT

Version 5: March 2002

H.-L. Huang
H. M. Woolf
R. L. Solomon
R. L. Slonaker

RAYTHEON COMPANY
Information Technology and Scientific Services
4400 Forbes Boulevard
Lanham, MD 20706

SBRS Document #: Y2392

EDR: Cloud Cover/Layers

Doc No: Y2392

Version: 5

Revision: 0

	Function	Name	Signature	Date
Prepared by	EDR Developer	H.-L. HUANG		1/19/02
Approved by	Relevant Lead	R. SLONAKER		1/31/02
Approved by	Chief Scientist	S. MILLER		2/1/02
Released by	Algorithm Lead	P. KEALY		2/4/02

TABLE OF CONTENTS

Page

LIST OF FIGURES iii

LIST OF TABLES v

GLOSSARY OF ACRONYMS vii

ABSTRACT ix

1.0 INTRODUCTION 1
- 1.1 PURPOSE 1
- 1.2 SCOPE 1
- 1.3 VIIRS DOCUMENTS 1
- 1.4 REVISIONS 2

2.0 EXPERIMENT OVERVIEW 3
- 2.1 OBJECTIVES OF CLOUD LAYER RETRIEVALS 3
- 2.2 INSTRUMENT CHARACTERISTICS 4
- 2.3 RETRIEVAL STRATEGY 10
  - 2.3.1 Iterative clustering - a K-Mean algorithm 10
  - 2.3.2 Cloud type determination 11
  - 2.3.3 Cloud Fraction Determination 11

3.0 ALGORITHM DESCRIPTION 13
- 3.1 PROCESSING OUTLINE 13
- 3.2 ALGORITHM INPUT 15
  - 3.2.1 VIIRS Data 15
    - 3.2.1.1 Cloud Mask and Phase Diagnostics 15
    - 3.2.1.2 Cloud Top Height 15
    - 3.2.1.3 Cloud Optical Thickness 15
    - 3.2.1.4 Cloud Effective Particle Size 15
  - 3.2.2 Non-VIIRS Data 16
    - 3.2.2.1 Cluster Merging Criterion 16
    - 3.2.2.2 Outlier Detection Criterion 16
- 3.3 THEORETICAL DESCRIPTION OF ALGORITHMS 16
  - 3.3.1 Physics of the Problem 16
  - 3.3.2 Mathematical Description of CC/L Algorithm 16
    - 3.3.2.1 Oblique View Correction 16
    - 3.3.2.2 CTH/Phase Initial Clustering 18

3.3.2.3 PDC/K-Means Clustering 18
3.3.2.4 Cloud Type Determination 20
3.3.2.5 Cloud Fraction Determination 21
3.3.3 Archived Algorithm Output 21
3.3.4 Variance and Uncertainty Estimates 21
3.3.4.1 Error Budget 22
3.4 ALGORITHM SENSITIVITY STUDIES 22
3.4.1 Description of Data Set and Simulation 22
3.4.2 Case Results 24
3.5 PRACTICAL CONSIDERATIONS 29
3.5.1 Numerical Computation Considerations 29
3.5.2 Programming and Procedural Considerations 30
3.6 ALGORITHM VALIDATION 30
4.0 ASSUMPTIONS AND LIMITATIONS 31
4.1 ASSUMPTIONS 31
4.1.1 Spatial Domain of Processing 31
4.1.2 Cloud Type Definition 31
4.1.3 Processing Constraints 31
4.2 LIMITATIONS 31
4.2.1 Inherent Cluster Ambiguity 31
4.2.2 Definition of Cloud Layers/Types 31
5.0 REFERENCES 33

LIST OF FIGURES

Page

Figure 1. Summary of VIIRS design concepts and heritage. 5

Figure 2. VIIRS detector footprint aggregation scheme for building "pixels." 6

Figure 3. Benefits of VIIRS aggregation scheme in reducing pixel growth at edge of scan. 6

Figure 4. VIIRS spectral bands, visible and near infrared. 8

Figure 5. VIIRS spectral bands, short wave infrared. 8

Figure 6. VIIRS spectral bands, mid-wave infrared. 9

Figure 7. VIIRS spectral bands, long wave infrared. 9

Figure 8. CC/L EDR high-level flow diagram. 13

Figure 9. The cloud location correction due to the oblique satellite view when satellite is located at equator 17

Figure 10. The cloud location correction due to the oblique satellite view when the satellite is located at 30 degrees north. 18

Figure 11. Cloud mask and Cloud Top Pressure images of the MODIS data set used for testing the CC/L processing algorithm. 23

Figure 12. Cloud Effective Particle Size and Cloud Optical Thickness images of the MODIS data set used for testing the CC/L processing algorithm 23

Figure 13. Cloud Top Height image of the MODIS data set used for testing the CC/L processing algorithm 24

Figure 14. Cloud Effective Particle Size (µm) vs. Cloud Optical Thickness scatter plots of 6 PDC/K Means clusters. 26

Figure 15. Scatter plot (COT vs. CTH) of 4 final clusters: Ice_Upper, Ice_Lower, Water_Upper, and Water_Lower. 27

Figure 16. Scatter plot (EPS vs. CTH) of 4 final clusters: Ice_Upper, Ice_Lower, Water_Upper, and Water_Lower. 27

Figure 17. Six PDC classes/layers are shown in different colors. 28

Figure 18. Cloud fraction histogram. 29

LIST OF TABLES

Page

Table 1. System Specification requirements for Cloud Cover/Layers. 3

Table 2. VIIRS VNIR bands. 7

Table 3. VIIRS SWIR, MWIR, and LWIR bands. 7

Table 4. Predefined cloud types characterized in terms of their macro (height and phase) and micro (size and optical thickness) properties. 14

Table 5. Algorithms and error sources. 21

Table 6. Pixel level cloud property character of the algorithm test data set 22

Table 7. "Initial" CTH/Phase and "final" PDC/K-means cluster/layer characteristics. 25

GLOSSARY OF ACRONYMS

ATBD	Algorithm Theoretical Basis Document
AVHRR	Advanced Very High Resolution Radiometer
CC/L	Cloud Cover/Layers
CDR	Critical Design Review
COT	Cloud Optical Thickness
CTH	Cloud Top Height
DoD	Department of Defense
EDR	Environmental Data Record
EOS	Earth Observing System
EPS	Effective Particle Size
ETM	Enhanced Thematic Mapper (Landsat)
GIFOV	Ground Instantaneous Field of View
GSD	Ground Sample Distance
HC	Horizontal Cell
HCS	Horizontal Cell Size
HSI	Horizontal Sample Interval
HSR	Horizontal Spatial Resolution
IP	Intermediate Product
LLLS	Low Level Light Sensor
LWIR	Longwave Infrared
METEOSAT	Geosynchronous METEOrological SATellite
MODIS	Moderate Resolution Imaging Spectroradiometer
MTC	Merging Test Criteria
MWIR	Middlewave Infrared
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NPOESS	National Polar-orbiting Operational Environmental Satellite System
OLS	Operational Linescan System
PD	Property Distance
PDC	Property Distance Clustering
RDR	Raw Data Record
SBRS	Santa Barbara Remote Sensing
SDSM	Solar Diffuser Stability Monitor
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
SNR	Signal-to-Noise Ratio
SRD	Sensor Requirements Document
SWIR	Shortwave Infrared
TBD	To Be Determined
TBR	To Be Reviewed

THEMIS	Thermal Emission Imaging System
TIROS	Television Infrared Observation Satellite
TMI	Type Matching Index
VIIRS	Visible/Infrared Imager/Radiometer Suite
VNIR	Visible/Near Infrared

ABSTRACT

Total cloud cover and layered cloud structure are basic components of an image analysis procedure called “nephanalysis”. Many operational users of cloud data require analysis of the extent, type, and physical characteristics of vertically distributed cloud layers. For example, in-flight aircraft refueling has stringent requirements for cloud-free visibility between aircraft at flight altitude. Icing specification and forecasts depend on accurate initial depiction of the constituent particle sizes and state (liquid or frozen) of clouds at specific altitudes. Many other uses, such as accurate prediction of lines-of-sight for aerial reconnaissance, depend on accurate classification of clouds into the classic cloud families. These diverse needs are best met by a novel algorithm called Property Distance Clustering (PDC) that adapts the K-means cluster algorithm by using the local similarity relationships among pixels to combine them in clusters.

This document includes a thorough description of the established behavior of the Cloud Cover/Layers (CC/L) Environmental Data Record (EDR) processing approach. This algorithm with initial heritage from K-means clustering requires no VIIRS SDR input; instead, input includes pixel-level information such as the Cloud Top Height, the Effective Particle Size, and the Cloud Optical Thickness. The pixel-level cloud mask and cloud phase intermediate products (IP) are used as well. Since both cloud macro (height and phase) and micro (effective particle size and optical thickness) properties are used in characterizing distinct cloud layers, the algorithm can meet the objective of the CC/L EDR to identify the vertical structure of clouds consistent with pixel-level observations within each horizontal (aggregation) cell over a VIIRS image.

1.0 INTRODUCTION

Techniques for retrieving cloud cover parameters from multispectral satellite imagery have been developed and tested with NOAA/AVHRR and METEOSAT (Arking and Childs, 1985, Desbois et al., 1982). A clustering approach, combined with the threshold method to identify cloud cover, cloud type, cloud top temperature and cloud optical thickness, is shown to be successful. The VIIRS CC/L algorithm is based on the clustering heritage using the derived cloud top height and cloud optical properties to identify cloud layer and cover.

1.1 PURPOSE

This CC/L Algorithm Theoretical Basis Document (ATBD) describes the algorithms used to determine the layered structure of cloud cover using VIIRS EDRs and IPs. These EDRs include Cloud Top Height (CTH), Cloud Optical Thickness (COT) and cloud Effective Particle Size (EPS). The algorithm also requires the IPs of the VIIRS cloud mask and phase. The primary purpose of this ATBD is to establish guidelines for the production of the CC/L EDR. This document will describe the required inputs, the theoretical foundation of the algorithms, the sources and magnitudes of the errors involved, practical considerations for post-launch implementation, and the assumptions and limitations associated with the products.

1.2 SCOPE

This Algorithm Theoretical Basis Document (ATBD) details a novel algorithm to reconstruct, on a pixel basis, the vertical distribution of cloud cover within a Horizontal Cell (HC) from recovered physical parameters. The selected algorithm fits three dimensional cloud covers to a horizontal, vertical and physically consistent cloud structure within each aggregated HC to meet and/or exceed the threshold requirements for this EDR. Section 1 describes the purpose and scope of this document; it also includes a listing of VIIRS documents that will be cited in the following sections. Section 2 provides a brief overview of the motivation for the CC/L algorithm, including the objective of the EDR, the currently designed VIIRS instrument characteristics, and the strategy for obtaining the CC/L EDR. Section 3 contains the essence of this document, a complete description of that retrieval process. Consideration is given to the overall structure, the required inputs, a theoretical description of the products, assessment of the error budget, case results of ongoing sensitivity studies, practical implementation issues, validation, and the algorithm development schedule. Section 4 provides an overview of the constraints, assumptions and limitations associated with the CC/L EDR, and Section 5 contains a listing of references cited throughout this document.

1.3 VIIRS DOCUMENTS

Reference to VIIRS documents within this ATBD will be indicated by an italicized number in brackets, e.g., *[V-1]*.

[V-1] VIIRS System Specification.

[V-2] VIIRS Geolocation ATBD.

[V-3] VIIRS Calibration/Validation Plan.

[V-4] VIIRS Cloud Mask/Phase ATBD.

[V-5] VIIRS Cloud Top Parameters ATBD.

[V-6] VIIRS Cloud Optical Thickness & Cloud Effective Particle Size ATBD.

1.4 REVISIONS

PR-08923-04-02, Version 1 Revision 0, Annotated Abstract, June 10, 1998.

PR-08923-04-02, Version 1, Revision 0.1, Annotated Outline, August 15, 1998.

Y2392, Version 1, Revision 2, Cloud Cover/Layers ATBD, October 1998.

Y2392, Version 1, Revision 3, Cloud Cover/Layers ATBD, March 1999.

Y2392, Version 2, Revision 0, Cloud Cover/Layers ATBD, June 1999.

Y2392, Version 3, Revision 0, Cloud Cover/Layers ATBD, May 2000.

Y2392, Version 4, Revision 0, Cloud Cover/Layers ATBD, May 2001.

Y2392, Version 5*, Revision 0, Cloud Cover/Layers ATBD, March 2002.

* Version 5 is essentially a completely new version since the old CC/L algorithm is replaced with the new PDC/K-Means algorithm.

2.0 EXPERIMENT OVERVIEW

2.1 OBJECTIVES OF CLOUD LAYER RETRIEVALS

The CC/L Algorithm will be developed to meet System Specification requirements. This algorithm will use EDRs that have been retrieved for each image pixel rather than computed from horizontally aggregated EDRs (i.e., EDRs for horizontal cells). The objective is to identify the vertical structure of clouds consistent with pixel-level observations within each horizontal (aggregation) cell over a VIIRS image.

The VIIRS SRD provides the following definition for cloud CC/L:

> *"Cloud cover/layers consists of two data products:*
>
> *(a) fractional cloud cover, defined (TBR) as the fraction of a given area on the Earth's surface for which a locally normal line segment extending between two given altitudes intersects a cloud, and*
>
> (b) a binary (cloudy/not cloudy) map at the pixel level indicating the pixels which are deemed to contain clouds. The detection criterion for cloudiness at the pixel level is TBD.
>
> *As a threshold, fractional cloud cover is required for up to four layers of the atmosphere between the surface and an altitude of 20 km. As an objective, cloud cover is required for contiguous, 0.1 km thick layers at 0.1 km increments in altitude, from the surface of the Earth to an altitude of 30 km.."*

The VIIRS System Specification requirements for cloud cover/layers are set out in Table 1.

Table 1. System Specification requirements for cloud cover/layers.

Requirement Number	Parameter	Requirement
SSV0202	EDR CCOVR/L Fractional Cloud Cover HCS:	25 km
SSV0204	EDR CCOVR/L Fractional Cloud Cover HRI:	HCS
SSV0205	EDR CCOVR/L Fractional Cloud Cover Vertical Reporting Interval:	4 layers
SSV0206	EDR CCOVR/L Fractional Cloud Cover Horizontal Coverage:	Global
SSV0207	EDR CCOVR/L Fractional Cloud Cover Vertical Coverage:	0 to 20 km
SSV0209	EDR CCOVR/L Fractional Cloud Cover Measurement Range:	0 to 1.0 HRI
SSV0211	EDR CCOVR/L Fractional Cloud Cover Measurement Accuracy, single layers at nadir:	0.07 HRI
SSV0753	EDR CCOVR/L Fractional Cloud Cover Measurement Accuracy, single layers at EOS:	0.10 HRI
SSV0754	EDR CCOVR/L Fractional Cloud Cover Measurement Accuracy, multiple layers at nadir:	0.07 HRI
SSV0755	EDR CCOVR/L Fractional Cloud Cover Measurement Accuracy, multiple layers at EOS:	0.10 HRI
SSV0212	EDR CCOVR/L Fractional Cloud Cover Measurement Precision, single layers at nadir:	0.07 HRI
SSV0756	EDR CCOVR/L Fractional Cloud Cover Measurement Precision, single layers at EOS:	0.15 HRI

SSV0757	EDR CCOVR/L Fractional Cloud Cover Measurement Precision, multiple layers at nadir:	0.07 HRI
SSV0758	EDR CCOVR/L Fractional Cloud Cover Measurement Precision, multiple layers at EOS:	0.15 HRI
SSV0215	EDR CCOVR/L Fractional Cloud Cover Swath Width:	3000 km
SSV0203	EDR CCOVR/L Binary Map HCS:	Pixel size
SSV0896	EDR CCOVR/L Binary Map HRI:	HCS
SSV0210	EDR CCOVR/L Binary Map Measurement Range:	Cloudy /Not cloudy
SSV0759	EDR CCOVR/L Binary Map Probability of Correct Typing, day, ocean, OD ≤ 0.5:	92%
SSV0760	EDR CCOVR/L Binary Map Probability of Correct Typing, day, ocean, OD > 0.5:	99%
SSV0761	EDR CCOVR/L Binary Map Probability of Correct Typing, day, land, OD ≤ 1:	85%
SSV0762	EDR CCOVR/L Binary Map Probability of Correct Typing, day, land, OD > 1:	93%
SSV0763	EDR CCOVR/L Binary Map Probability of Correct Typing, night, ocean, OD ≤ 0.5:	90%
SSV0764	EDR CCOVR/L Binary Map Probability of Correct Typing, night, ocean, OD > 0.5:	96%
SSV0765	EDR CCOVR/L Binary Map Probability of Correct Typing, night, land, OD ≤ 1:	85%
SSV0766	EDR CCOVR/L Binary Map Probability of Correct Typing, night, land, OD > 1:	90%
SSV0767	EDR CCOVR/L Binary Map Swath Width:	3000 km

2.2 INSTRUMENT CHARACTERISTICS

The VIIRS instrument will now be briefly described to clarify the context of the descriptions of the CC/L EDR presented in this document. VIIRS can be pictured as a convergence of three existing sensors, two of which have seen extensive operational use at this writing.

The Operational Linescan System (OLS) is the operational visible/infrared scanner for the Department of Defense (DoD). Its unique strengths are controlled growth in spatial resolution through rotation of the Ground Instantaneous Field of View (GIFOV) and the existence of a Low-Level Light Sensor (LLLS) capable of detecting visible radiation at night. OLS has primarily served as a data source for manual analysis of imagery. The Advanced Very High Resolution Radiometer (AVHRR) is the operational visible/infrared sensor flown on the National Oceanic and Atmospheric Administration (NOAA) Television Infrared Observation Satellite (TIROS-N) series of satellites (Planet, 1988). Its unique strengths are low operational and production cost and the presence of five spectral channels that can be used in a wide number of combinations to produce operational and research products. In December 1999, the National Aeronautics and Space Administration (NASA) launched the Earth Observing System (EOS) morning satellite, *Terra*, which includes the Moderate Resolution Imaging Spectroradiometer (MODIS). This sensor possesses an unprecedented array of 36 spectral bands at resolutions ranging from 250 m to 1 km at nadir, allowing a wide range of satellite-based environmental measurements.

VIIRS will reside on a platform of the National Polar-orbiting Operational Environmental Satellite System (NPOESS) series of satellites. It is intended to be the product of a convergence between DoD, NOAA and NASA in the form of a single visible/infrared sensor capable of satisfying the needs of all three communities, as well as the research community beyond. As

such, VIIRS will require three key attributes: high spatial resolution with controlled growth off nadir, minimal production and operational cost, and a large number of spectral bands to satisfy the requirements for generating accurate operational and scientific products.

Figure 1 illustrates the design concept for VIIRS, designed and built by Raytheon Santa Barbara Remote Sensing (SBRS). At its heart is a rotating telescope scanning mechanism that minimizes the effects of solar impingement and scattered light. Calibration is performed onboard using a solar diffuser for short wavelengths and a V-groove blackbody source and deep space view for thermal wavelengths. A Solar Diffuser Stability Monitor (SDSM) is also included to track the performance of the solar diffuser. The nominal altitude for NPOESS will be 833 km. The VIIRS scan will extend to 56 degrees on either side of nadir.

The VIIRS SRD places explicit requirements on spatial resolution for the Imagery EDR. Specifically, the Horizontal Spatial Resolution (HSR) of bands used to meet threshold Imagery EDR requirements must be no greater than 400 m at nadir and 800 m at the edge of the scan. This led to the development of a unique scanning approach which optimizes both spatial resolution and Signal to Noise Ratio (SNR) across the scan. The concept is summarized in Figure 2 for the imagery bands; the nested lower resolution radiometric bands follow the same paradigm at exactly twice the size. The VIIRS detectors are rectangular, with the smaller dimension projecting along the scan. At nadir, three detector footprints are aggregated to form a single VIIRS "pixel." Moving along the scan away from nadir, the detector footprints become larger both along track and along scan, due to geometric effects and the curvature of the Earth. The effects are much larger along scan. At around 32 degrees in scan angle, the aggregation scheme is changed from 3x1 to 2x1. A similar switch from 2x1 to 1x1 aggregation occurs at 48 degrees. The VIIRS scan consequently exhibits a pixel growth factor of only 2 both along track and along scan, compared with a growth factor of 6 along scan which would be realized without the use of the aggregation scheme. Figure 3 illustrates the benefits of the aggregation scheme for spatial resolution.

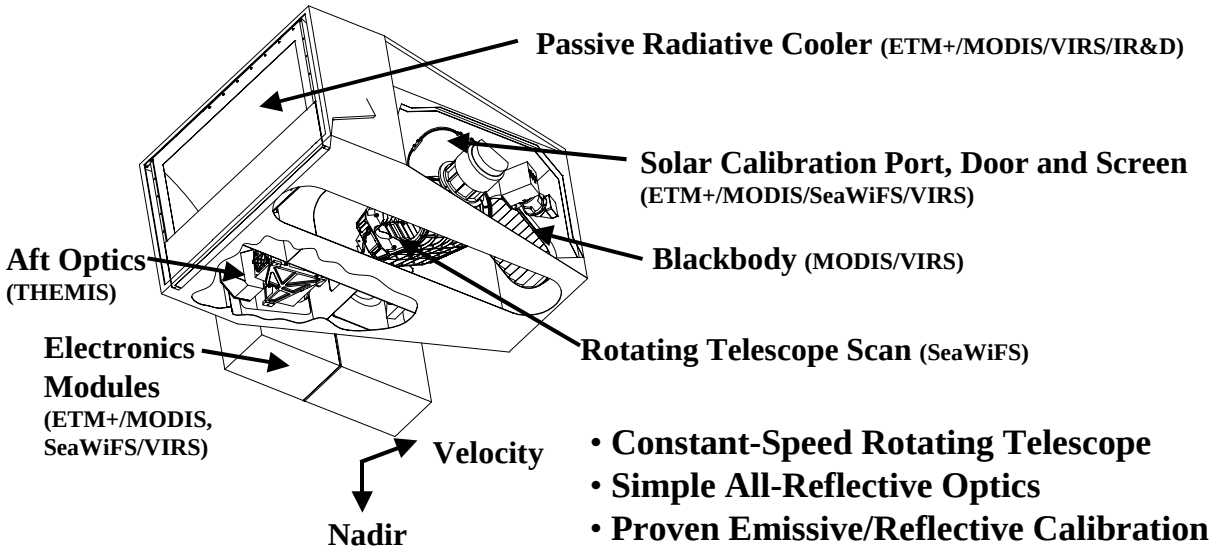

Figure 1. Summary of VIIRS design concepts and heritage.

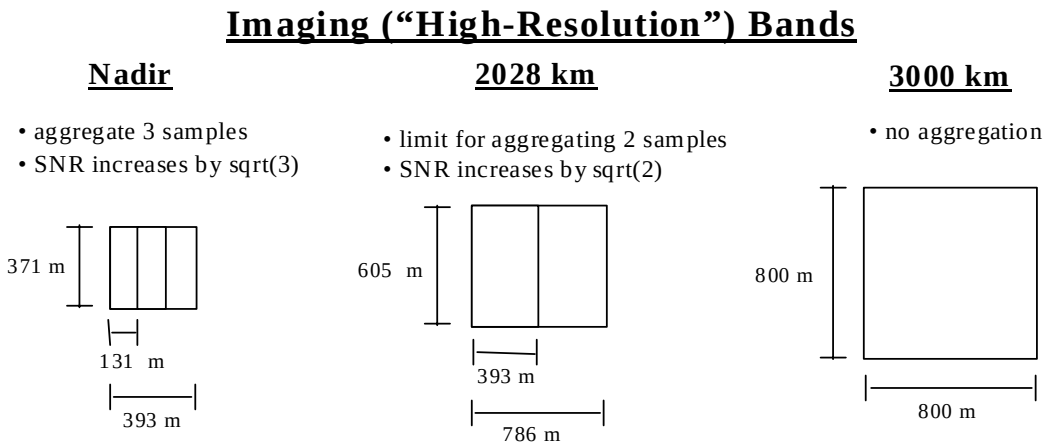

Figure 2. VIIRS detector footprint aggregation scheme for building "pixels." (Dimensions shown are approximate)

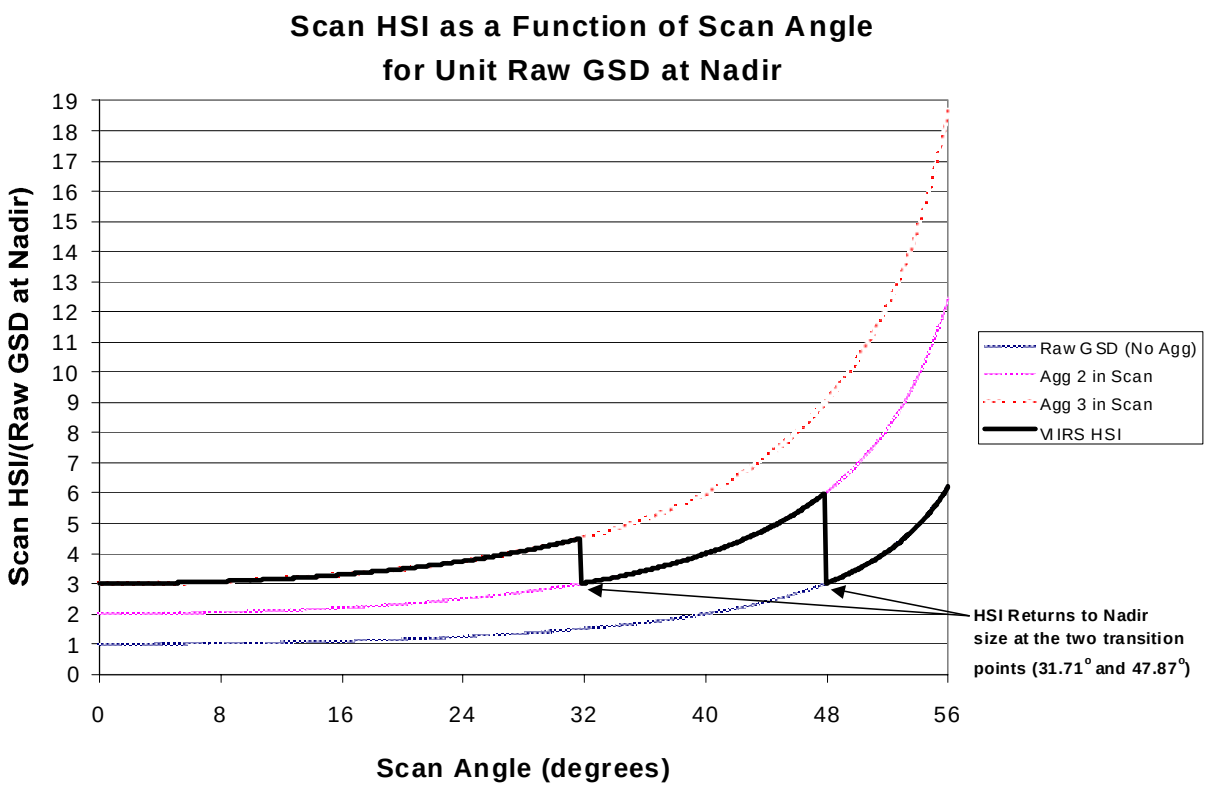

Figure 3. Benefits of VIIRS aggregation scheme in reducing pixel growth at edge of scan.

The VIIRS bands are summarized in Table 2 and Table 3. The positioning of the VIIRS spectral bands is summarized in Figures 4-7.

Table 2. VIIRS VNIR bands.

Band Name	Wavelength (μm)	Bandwidth (μm)
Day Night Band	0.700	0.400
M1	0.412	0.020
M2	0.445	0.018
M3	0.488	0.020
M4	0.555	0.020
I1	0.640	0.080
M5	0.672	0.020
M6	0.746	0.015
I2	0.865	0.039
M7	0.865	0.039

Table 3. VIIRS SWIR, MWIR, and LWIR bands.

Band Name	Wavelength (μm)	Bandwidth (μm)
M8	1.240	0.020
M9	1.378	0.015
I3	1.610	0.060
M10	1.610	0.060
M11	2.250	0.050
I4	3.740	0.380
M12	3.700	0.180
M13	4.050	0.155
M14	8.550	0.300
M15	10.7625	1.000
I5	11.450	1.900
M16	12.0125	0.950

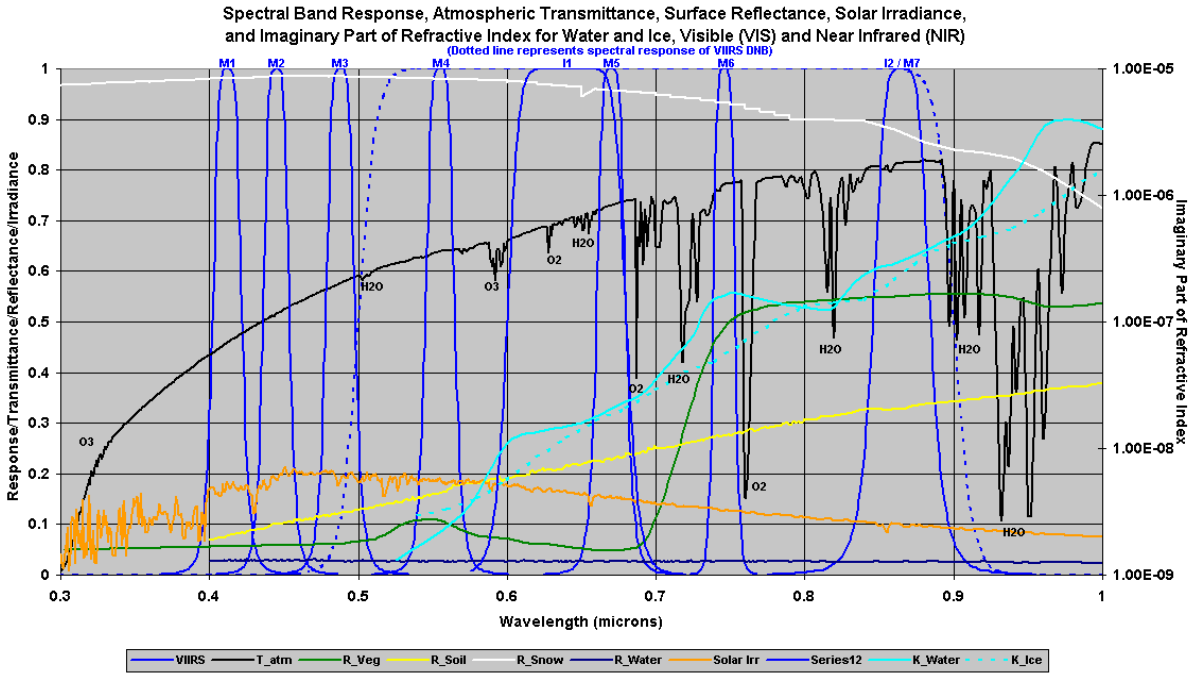

Figure 4. VIIRS spectral bands, visible and near infrared.

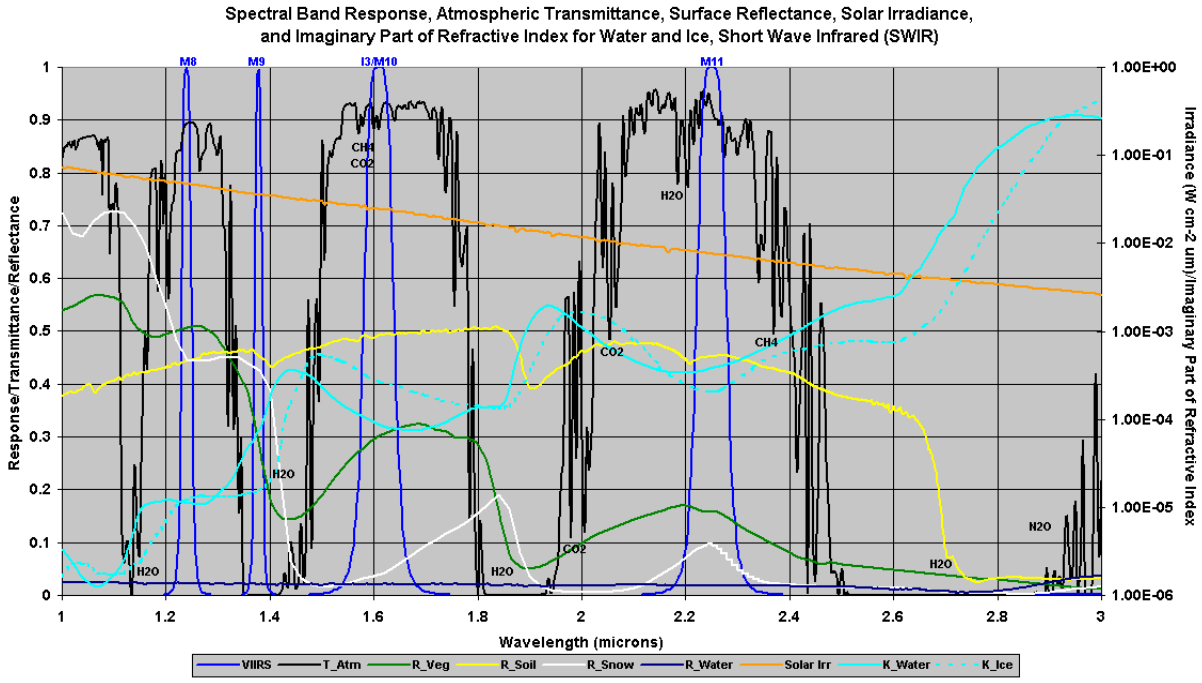

Figure 5. VIIRS spectral bands, short wave infrared.

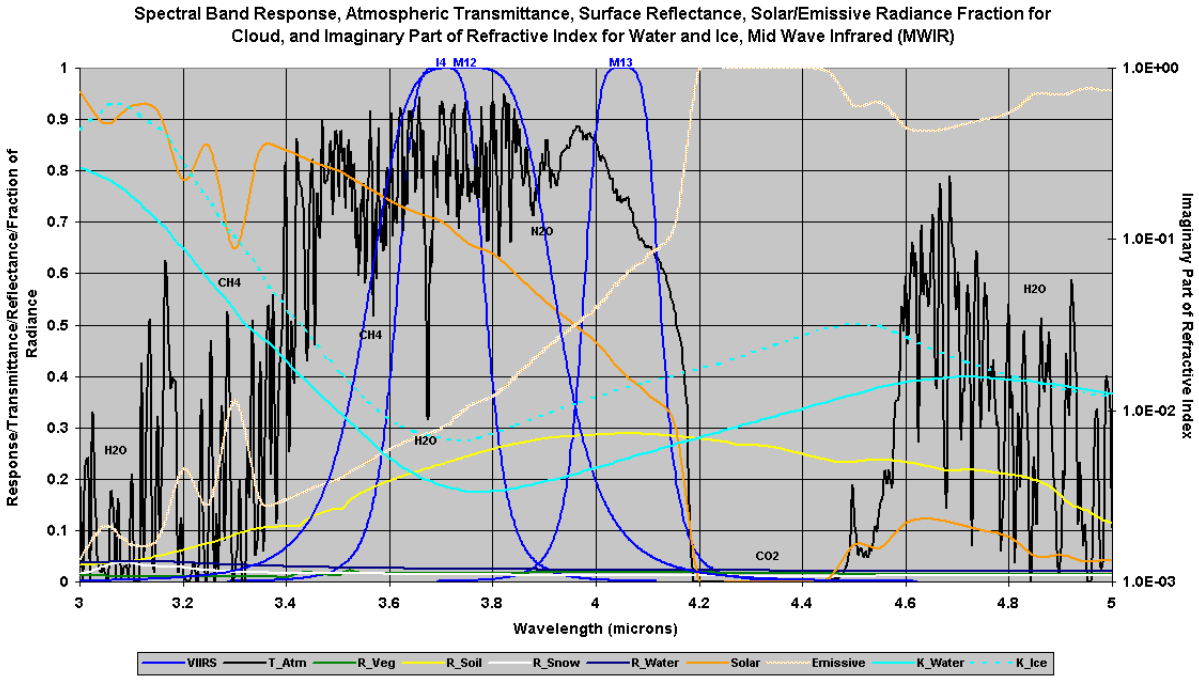

Figure 6. VIIRS spectral bands, medium wave infrared.

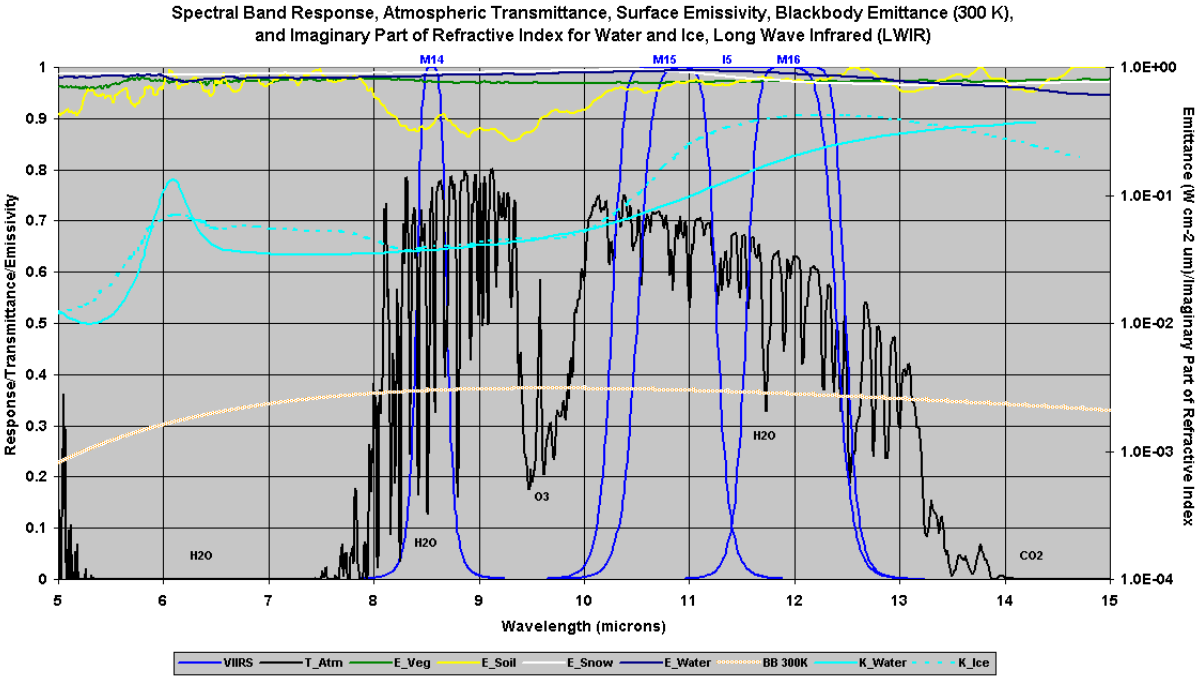

Figure 7. VIIRS spectral bands, long wave infrared.

The VIIRS instrument will exhibit approximately square pixel footprints that increase in size from roughly 750m at nadir to about 1600m at the edge of scan. The lower pixel resolution off-nadir may impair the cloud cover/layers product by offering fewer samples per horizontal cell. Therefore, the use of a variable Horizontal Cell Size (HCS) with a constant number of pixels per HC might be the best strategy to preserve consistent CC/L performance from nadir to edge of scan.

Additionally, the off-nadir pixels will have been sensed at oblique look angles. In this case, adjacent cloud elements will tend to obscure cloud-free regions between them, increasing the measured amount of cloud in each horizontal cell. Finally, the edges and sides of larger clouds will be sensed more preferentially, and these may show as spurious small cloud layers in the analysis.

2.3 RETRIEVAL STRATEGY

The strategy is to use supervised CTH and phase (ice, water, and mixed cloud) to obtain an initial cluster, followed by unsupervised EPS and COT K-means (MacQueen, 1967; Selim and Ismail, 1984; Theiler and Gisler, 1997) clustering to group pixels into distinct layers. Initially each distinct layer represents a vertical slab entity that is uniquely separated by CTH. Pixels are grouped into three vertical layers -- high, middle and low for ice, water, and mixed clouds, respectively. COT and EPS properties are used to define their cluster attributes in terms of cluster means. Individual pixels in each layer can then, upon consideration of property similarity and vertical location proximity, be linked to traditional cloud types. The resulting pixel distributions are used to infer cloud fraction for each layer, and an estimate is made of the total amount of cloud within the HC.

The CC/L strategy can be summarized as follows:

2.3.1 Iterative clustering - a K-Mean algorithm

Step 1: Start with an initial CTH/Phase partition with K clusters

An initial partition can be formed by first specifying similar vertically located pixels that are members of the same vertical slab (with similar cloud altitude range and phase). These pixels are grouped together into two pre-defined clusters for each of the ice, water, and mixed cloudy pixels (initially, up to 9 clusters can be formed).

Step 2: Generate a new partition by assigning each pixel to its closest cluster center

Step 3: Compute new cluster centers as the centroid of the clusters

A set of K patterns/clusters that are well separated from each other can be obtained by taking the centroid of the data as the first seed point and selecting successive seed points which are at least a certain distance away from those already chosen.

Step 4: Repeat steps 2 and 3 until an optimum value of the criterion function is found

Partitions are updated by reassigning pixels to clusters in an attempt to reduce the distance-error. Limited iteration may be sufficient if the change of distance-error is small.

The Euclidean metric is the most common metric for computing the distance between a pixel and cluster centers.

Distance-error is defined as the distance between a single pixel and a cluster centroid. Greater distance implies larger separation from the cluster and hence higher distance-error.

Step 5: Adjust the number of clusters by merging and splitting existing clusters or by removing small, or outlier, clusters.

2.3.2 Cloud type determination

Classic cloud family members can be linked to the cloud clusters/layers according to their cloud height, phase and other micro-properties. Each distinct cloud layer is identified as a well-defined cloud type.

The VIIRS IPs of CTH, EPS, COT, cloud mask and cloud phase, all at single pixel resolution, will be used for CC/L EDR retrieval.

2.3.3 Cloud Fraction Determination

Cloud fraction is defined as the ratio of the number of cloudy pixels for a given cluster/layer to the total number of pixels (clear and cloudy) within each HC. It is the cloud fraction of a cluster/layer valid for that particular HC.

3.0 ALGORITHM DESCRIPTION

3.1 PROCESSING OUTLINE

Figure 8 shows the high-level flow diagram for the CC/L EDR iterative processing. The CC/L algorithm operates on the VIIRS cloud mask and assigns cloudy pixels to various layers by selecting small groups of pixels that share common physical parameters. These parameters include CTH, phase, COT, and EPS. These four parameters permit discrimination via different altitudes and types of clouds. The resulting small layers are statistically merged as needed into a layered structure (initially sorted by cloud top height and phase). A pre-defined cloud type model is then used to determine the type of cloud in each layer. Its fractional coverage can be estimated from the population of cloudy (identified by cloud mask) and total pixels within each layer of the HC.

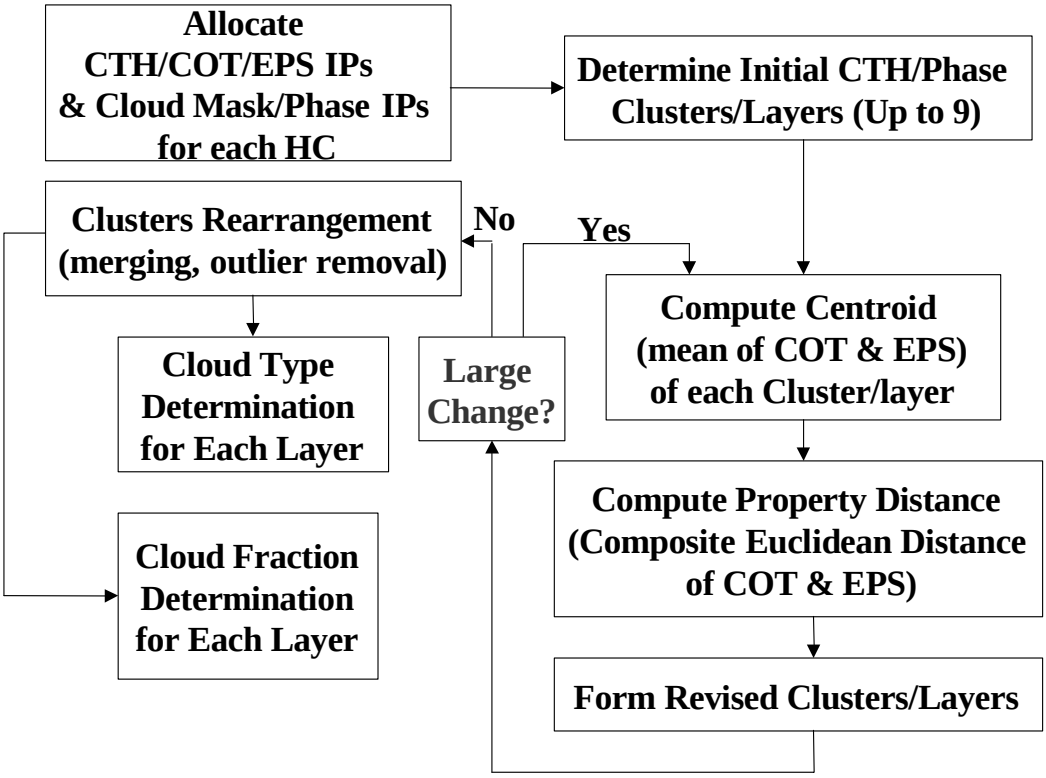

Figure 8. CC/L EDR high-level flow diagram.

Before starting the clustering algorithm, CC/L location is corrected for oblique views. The CC/L ground location is adjusted to remove the parallax positioning error that is due to the viewing geometry. For cloud-covered pixels that fall within the same HC, a K-means clustering procedure is used to determine the pixels that have similarity in COT and EPS values to the cluster/layer centroid under consideration. For each pixel, the property distance (PD) defined as the composite Euclidean distance (Burrough, 1986) of COT and EPS from the cluster centroid is calculated. The variance of COT and EPS (µm) are roughly the same, rendering these variables of approximate equal importance in the clustering algorithm. A future enhancement under consideration is to use normalized COT and EPS (using variance) when determining Euclidean

distance yielding an exactly equal weighting to both variables. For each pixel, the "correct" cluster is defined as that with minimum PD. If this cluster is different from the original based on CTH the change is logged. When all pixels have been processed, centroids of any clusters that have logged changes are re-computed. Each pixel's PDs are then computed with respect to the updated cluster centroid. As in the previous step a pixel may be re-assigned on the basis of these new PDs. This procedure is repeated until fewer than 10% of the pixels in a given cluster are re-assigned. All distinct cloud clusters/layers within an HC are tested for possible merging and outlier removal. If any one of the clusters/layers has fewer pixels than the predefined number (i.e. less than 1% of total pixel population within any single HC), this cluster is merged with another cluster/layer. The merging principle is based on which cluster/layer (other than the one itself) has the minimum PD. Outlier pixels within each cluster/layer are deleted. An outlier is a pixel whose PDs are larger than a pre-defined threshold (5 times the mean PD within each cluster/layer). The average pixel value of a parameter (e.g., cloud top height) in a specified cloud layer can be assigned as the layer value using the association between pixels and layers. Cloud type determination is performed by matching the average value of CTH, COT, and EPS of each cluster/layer with the cloud table. Hence each distinct layer within each HC will be identified as a cloud type with attributes described in Table 4 (Weickmann and Aufm Kampe, 1953; Heymsfield and Platt, 1984; Dowling and Radke, 1990; Liou, 1992).

Table 4. Predefined cloud types characterized in terms of their macro (height and phase) and micro (size and optical thickness) properties.

Cloud Type	Height (km)	Sizes (μm)	Optical Thickness	Phase
Stratus (St, Sc)	<2.5	2-25	1-10	Water
Alto Cumulus/Stratus (Ac, As)	1.5-5.5	4-30	2-32	Water/Ice
Cumulus (Cu, Cb)	0.2-6.5	5-50	3-50	Water/Ice
Cirrus (Ci)	6-12	10-100	0.01-5	Ice
Cirrocumulus (Cc)	6-15	30-120	1-8	Ice

3.2 ALGORITHM INPUT

3.2.1 VIIRS Data

The CC/L EDR uses other VIIRS cloud IPs as its primary inputs. The accuracy of the retrieval of these IPs will directly impact the performance of the retrieval. Gross errors (e.g., spatially consistent, large magnitude errors) in these other parameters will have little effect on the layered cloud retrieval algorithm due to the robust K-means PDC algorithm used to aggregate pixels into layers. Entirely random errors (e.g., instrument noise) will have a slight impact on the algorithm, but will be mitigated by the action of the clustering algorithm as a noise filter. The most significant impact on effective layering is the presence of spatially correlated errors. These errors may be interpreted by the algorithm as breaks in otherwise contiguous clusters, or even as entirely distinct, erroneous clusters consisting only of errors from the input data. Therefore, the essential requirement for all of the input IPs is spatial consistency.

The IPs that will be relied upon are CTH, COT and Cloud EPS. Although these EDRs will be available on the aggregation cell, the CC/L Algorithm requires them on a VIIRS pixel basis. In addition, the CC/L process requires the VIIRS Cloud Mask and phase, which are used to identify and distinguish cloudy pixels from clear pixels and water from ice phase pixels.

3.2.1.1 Cloud Mask and Phase Diagnostics

The VIIRS Cloud Mask will provide the fundamental spatial depiction of clouds. Only those pixels that are identified as being cloudy are processed by the CC/L algorithm. Cloud phase is also used to distinguish cluster layers.

3.2.1.2 Cloud Top Height

CTH serves as the primary physical characteristic used to distinguish layers. It provides the PDC algorithm with the initial clusters to start the K-means clustering process. Clouds, which stratify at different altitudes, are almost universally accepted to be in different layers, regardless of other, more superficial similarities.

3.2.1.3 Cloud Optical Thickness

In a further distinction after CTH, thin clouds may be considered distinct from opaque clouds. Thus, despite perceptual inclinations to the contrary, thin edges of cirrus layers may be classified as layers separate from the main mass of cirrus clouds.

3.2.1.4 Cloud Effective Particle Size

Finally, Cloud EPS will be used to distinguish layers that occur at the same altitudes but which consist of different cloud types. Since PD used in the K-means clustering algorithm is defined as sum of square of COT and EPS distance to the centroid, it is effectively using both COT and EPS characteristics at the same time to separate distinct clusters and layers.

3.2.2 Non-VIIRS Data

In addition to a cloud type table, some constants and data not derived from the VIIRS instrument will be required. User-specified control parameters will include the following:

3.2.2.1 Cluster Merging Criterion

A cluster or layer which has a pixel population smaller than 1% of the total pixel population of each HC, is subject to being merged with another nearby cluster.

3.2.2.2 Outlier Detection Criterion

Any pixel whose PD is 5 times larger than the mean PD for its cluster/layer is classified as an outlier and is deleted.

3.3 THEORETICAL DESCRIPTION OF ALGORITHMS

This section outlines the basic principles for obtaining CC/L.

3.3.1 Physics of the Problem

Rather than being a retrieval in the conventional sense of using a physical model to recover unknown parameters from measurements, the CC/L algorithm is statistical in nature. Multiple cloud IPs are used to guide the CC/L PDC algorithm to statistically group high similarity pixels within a single cluster. Unique physical attributes within a cluster/layer are defined since a logical link to classic cloud types is established. CC/L EDR uncertainty is heavily dependent on accurate CTH, COT, EPS, cloud mask, and phase IPs.

3.3.2 Mathematical Description of CC/L Algorithm

The CC/L algorithm consists of five main components: the Oblique View Correction, the CTH/Phase Initial Clustering, the iterative PDC/K-means clustering procedure, the cloud fraction, and the link of cluster to cloud type. The mathematical basis of each of these algorithms will be examined in turn.

3.3.2.1 Oblique View Correction

An oblique view correction is performed to correct observed pixel locations for parallax viewing errors. The pixel location correction is dependent on cloud height, view angles, and measurement location (latitude). Shown in figures 9 and 10 are corrections for clouds located at 2, 5, 10, and 20 km at the equator and 30ºN, respectively. These corrections are to be applied to the retrieval navigation location of the cloudy layer.

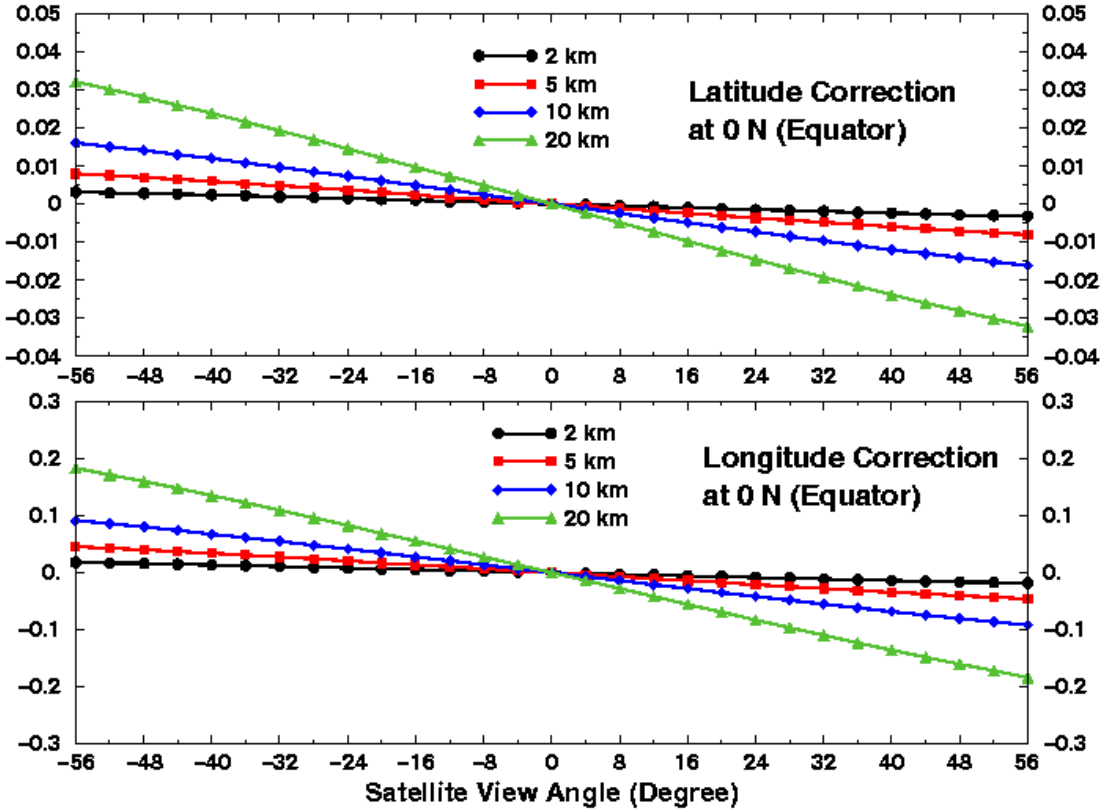

Figure 9. The cloud location correction due to the oblique satellite view when satellite is located at equator. Correction is greatly dependent on cloud altitude.

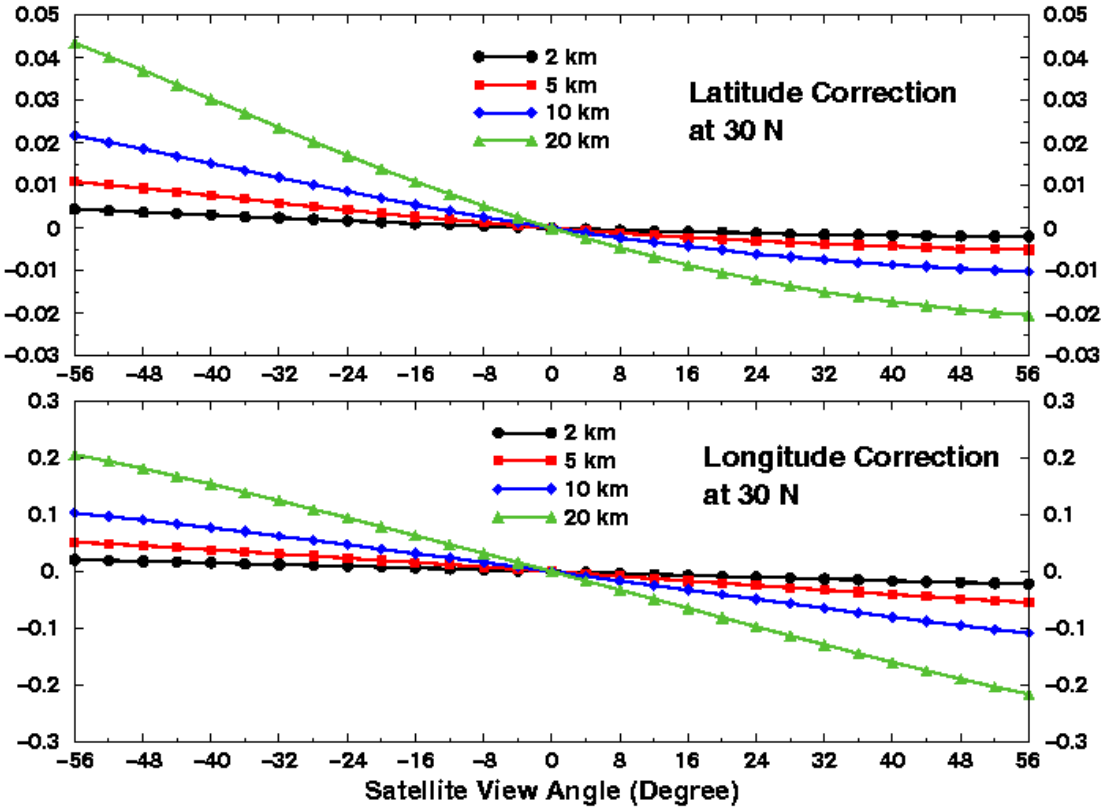

Figure 10. The cloud location correction due to the oblique satellite view when the satellite is located at 30 degrees north. Correction is greatly dependent on cloud altitude.

3.3.2.2 CTH/Phase Initial Clustering

The PDC/K-Means clustering algorithm requires initial partitions. VIIRS single pixel CTH and cloud phase EDR and IP are ingested to identify these initial clusters. Three partitions, namely high, middle, and low cloud are formed if CTH is greater than 6 km, between 2.5 and 6 km and below 2.5 km, respectively. The threshold values for cloud altitude partitioning can be geographically or seasonally dependent. However, these threshold values might not be significant, since the PDC/K-Means algorithm potentially will reassign pixels from their respective initial partitions to other existing partitions. Nevertheless, the use of dynamical CTH threshold values might speed up the clustering process since improved initial clustering would lead to the reassignment of fewer pixels. Three CTH clusters are further partitioned to form ice or water phase clusters using pixel level cloud mask IP information. In the end, CTH/Phase can produce up to 6 initial clusters such as Ice-High, Ice-Mid, Ice-Low, Water-High, Water-Mid, and Water-Low cloud layers. If mixed phase/multiple layer clouds exist within a single VIIRS field of view, then more than 6 classes are possible.

3.3.2.3 PDC/K-Means Clustering

Property Distance Clustering (PDC)/K-Means clustering is an effective iterative clustering technique designed to operate on sparse, randomly distributed and pre-partitioned data. It consists of cluster property mean/centroid estimation, pixel property distance estimation, cluster

assignment, and cluster rearrangement. If necessary the first 3 steps will be repeated until some condition is satisfied (less than a minimum number of pixels changing cluster).

- **Cluster Property Mean/Centroid Estimation**

  Cluster property mean/centroid is the center of gravity of each initial cluster defined by 3.3.2.2. Mean/Centroid of the cluster is defined as

  $(\text{COT Mean/Centroid})_j \equiv 1/N_j\ (\Sigma COT_i)$;

  $(\text{EPS Mean/Centroid})_j \equiv 1/N_j\ (\Sigma EPS_i)$

  $(\text{CTH Mean/Centroid})_j \equiv 1/N_j\ (\Sigma CTH_i)$

  Where N_j is the total number of cloudy pixels in cluster j, and i is pixel index.

- **Pixel PD Estimation**

  PD between cloudy pixel i and cluster j is defined as

  $PD_{i,j} \equiv \{[COT_i - (\text{COT Mean/Centroid})_j]^2 + [EPS_i - (\text{EPS Mean/Centroid})_j]^2\}^{1/2}$

  Mean-PD_j within each cluster j is also computed as following for cluster rearrangement

  Mean-$PD_j \equiv 1/N_j\ (\Sigma PD_{i,j})$

  This PD is defined in the Euclidean sense. That is, PD is the geometric distance in the multidimensional space defined by the cloud micro properties of effective particle size and optical thickness.

- **Cluster Assignment**

  Each cloudy pixel is assigned to cluster j if its $PD_{i,j}$ is smallest among $PD_{i,j}$ themselves where j is potentially from 1 to 6. This physically means pixel i is similar to cluster j since its pixel cloud micro-properties are close to the cluster centroid, and this pixel should be assigned to this particular cluster. The iterative decision will be made after the number of cloudy pixels which switch cluster is tallied. If the tallied switching pixel number is greater than the pre-defined threshold (10 %) then it will return to the first step of the PDC/K-Means cluster procedure and repeat the entire process.

- **Cluster Rearrangement**

  As soon as cluster assignment is done, clusters are rearranged for merging. For example, any cluster that has a total number of pixels less than 0.5 % of the total population of its HC is considered not representative, and should be merged with another nearby cluster. One more possible merging test is used to identify groups of cloud-covered pixels with the same cloud phase that are sufficiently similar to be considered as the same cloud cluster/layer. We will never consider the merging of any two different cloud phase

clusters/layers, even though their macro and micro properties might be very similar, and the Merging Test Criteria (MTC) defined below are satisfied.

The criteria for merging clusters/layers j1 and j2 can be simply defined in terms of fractional absolute distance as

$MTC_CTH \equiv [abs(Mean_CTH_{j1}-Mean_CTH_{j2})/(Mean_CTH_{j1}+Mean_CTH_{j2})]$

$MTC_COT \equiv [abs(Mean_COT_{j1}-Mean_COT_{j2})/(Mean_COT_{j1}+Mean_COT_{j2})]$

$MTC_EPS \equiv [abs(Mean_EPS_{j1}-Mean_EPS_{j2})/(Mean_EPS_{j1}+Mean_EPS_{j2})]$

If all three MTC (_CTH, _COT, and _EPS) are smaller than $0.25e^{-1}$ (=0.09197) then clusters/layers j1 and j2 should be merged.

Also, outlier cloudy pixels are identified as those whose PDs are greater than 5 times the cluster mean-PD. These outlier pixels are too distant from the cluster centroid and should not be accepted. Cluster/Layer means of CTH, COT and EPS are recalculated after the removal of outlier pixels for use in cloud type determination.

3.3.2.4 Cloud Type Determination

Up to 9 unique clusters might be available for matching with a cloud table to decide which cloud type the cluster/layer belongs to. Cluster/Layer means of CTH, COT and EPS derived from the previous step are used for the matching. The matching approach is

- Use cluster/layer phase information first
- For phase determined cluster/layer, compute Type Matching Index (TMI) as

  If cluster/layer is in ice phase

  $TMI_Ice(i) \equiv [(CTH_Mean - CTH_i) / CTH_i]^2$

  $+ [(EPS_Mean - EPS_i) / EPS_i]^2$

  $+ [(COT_Mean - COT_i) / COT_i]^2$

  for i=1 as Ci, or 2 as Cc, or 3 as Cu/Cb, or 4 as Ac/As

  If cluster/layer is water phase

  $TMI_Wat(i) \equiv [(CTH_Mean - CTH_i) / CTH_i]^2$

  $+ [(EPS_Mean - EPS_i) / EPS_i]^2$

  $+ [(COT_Mean - COT_i) / COT_i]^2$

  for i=1 as St/Sc, or 2 as Ac/As, or 3 Cu/Cb

- Select cloud type i (Ci, Cc, Cu/Cb, or Ac/As) for which TMI_Ice(i) is smallest among all TMI_Ice or if cluster/layer is pre-determined as water select cloud type i (St/Sc, Ac/As, or Cu/Cb) for which TMI_Wat(i) is smallest among all TMI_Wat.

3.3.2.5 Cloud Fraction Determination

Within each cluster/layer, cloudy pixels are re-counted and divided by the total (clear and cloudy) HC pixel population to determine cloud fraction. In other words, the cloud fraction of each unique cluster/layer is defined as the percent of its cloudy pixels in that HC.

3.3.3 Archived Algorithm Output

For each HC, Quality Control flag, CTH, COT, EPS, Cloud phase, Cloud layer/type, and cloud cover are outputs. Other ancillary data such as oblique view corrected geo-location should also be an output parameter.

3.3.4 Variance and Uncertainty Estimates

Errors in the CC/L EDR arise from several sources and in several places. Individual algorithms are sensitive to SDR measurement noise, band-to-band registration errors, input EDR and IP errors, and other effects. Table 5 summarizes possible error sources.

Table 5. Possible error sources for each algorithm.

Algorithm	Affected by SDR Noise?	Affected by Registration Errors?	Affected by IP Errors?	Other Error Sources
Initial CTH/Phase Clustering	Yes	Yes	Yes	Cloud Range Bin
PDC/K Means Clustering	Minimal	Minimal or No Effect	Yes	Centroid/Mean Definition Number of Iteration PDC Definition
Cluster Rearrangement	Minimal	Minimal or No Effect	Minimal or No Effect	Outlier Definition Merge Test Criterion
Cloud Type Determination	Minimal or No Effect	Minimal or No Effect	Minimal or No Effect	Cloud Type Definition
Cloud Fraction Determination	Yes	Yes	Yes	Cloud mask Definition

3.3.4.1 Error Budget

For a complete description of the errors that impact the CC/L Algorithm see the Raytheon VIIRS Error Budget, Version 5 (Y3249). Those error budgets are predicated on the linearity and independence of errors. In the Cloud Cover/Layers Algorithm, the contributing components are strongly coupled (i.e., non-independent) and act nonlinearly.

3.4 ALGORITHM SENSITIVITY STUDIES

The algorithm, being almost entirely statistical in nature, is less sensitive to band calibration errors. Systematic errors in IPs feeding the layered cloud algorithm resulting from mis-calibration are unlikely to result in significant layered cloud errors due to the robust statistical process (Initial CTH/Phase and PDC/K Means Clustering) being employed. Only variations in results directly attributable to errors in the input IPs should be expected. For example, initial CTH/Phase cluster is in error due to CTH IP and/or cloud phase IP in error, or the total sum of cloud coverage in all layers might be in error, but only because the cloud mask was in error.

The computation of PDC and K-Means algorithm acts like a noise suppressor, and is therefore less sensitive to SDR noise propagated through the IPs.

3.4.1 Description of Data Set and Simulation

Table 6 describes pixel level cloud property (height, phase, optical thickness and effective particle size) characteristics of the dataset derived from one granule (5 minutes) of Terra-MODIS measurements during 1600 to 1605 UTC on June 1 of 2001.

Table 6. Pixel level cloud property characteristics of the algorithm test data set

Layer		COT_Mean	EPS_Mean (μm)	Number of Samples	Note
Ice	**High**	24.0	25.4	31621	No Ice cloud is lower than 2.5 km
	Mid	1.89	20.3	36	
Water	**High-Mid**	13.5	17.5	11144	High & Mid water cloud are clustered into a single layer
	Low	4.6	17.3	7699	
Mixed	**High-Mid**	9.24	21.5	11517	High & Mid mixed phase cloud are clustered into a single layer
	Low	2.45	16.2	168	

62185 total pixels of MODIS cloud products derived from the MODIS science team algorithm form the dataset used in this ATBD. The following case results section presents the CC/L cluster analysis using this dataset. Figures 11, 12 and 13 are 2-D images of mask, cloud top pressure (in mb), EPS, COT, and CTH of this dataset, respectively.

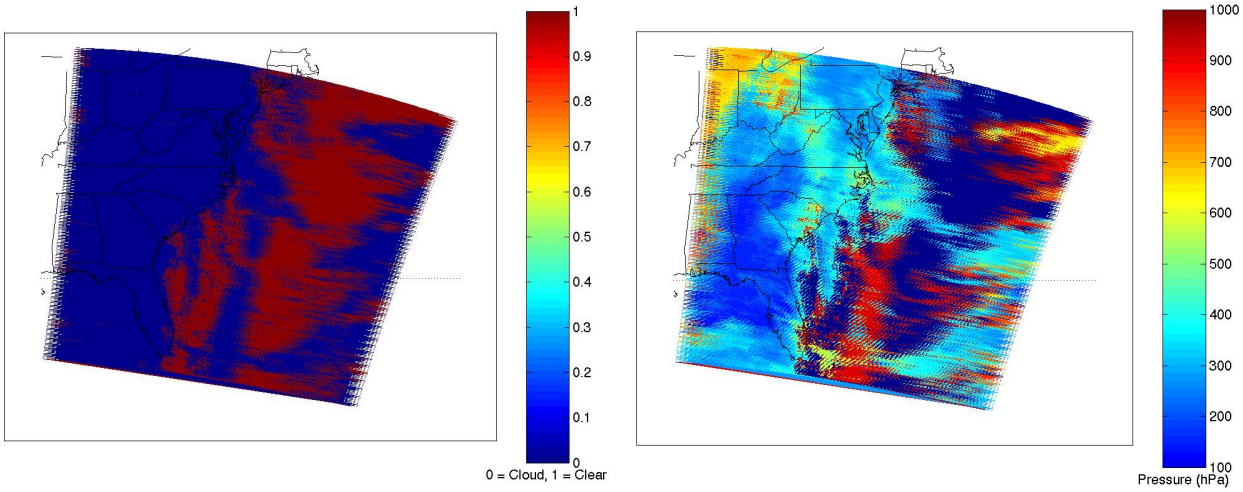

Figure 11. Cloud mask (left panel, Blue – cloudy, Dark Red – clear) and Cloud Top Pressure (right panel, mb) images of the MODIS data set used for testing CC/L processing algorithm.

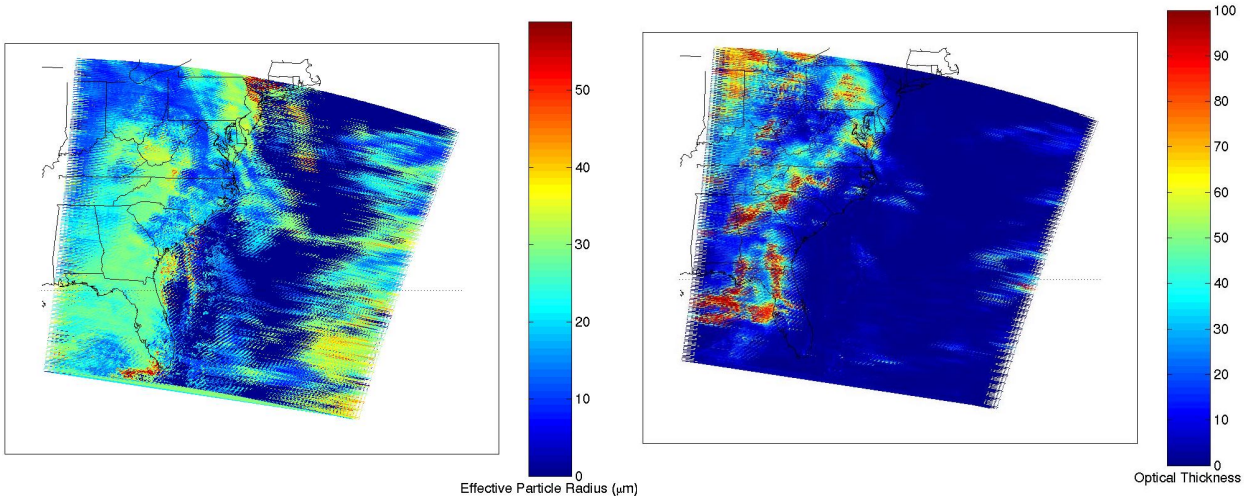

Figure 12. Cloud Effective Particle Size (left panel) and Cloud Optical Thickness (right panel) images of the MODIS data set used for testing the CC/L processing algorithm.

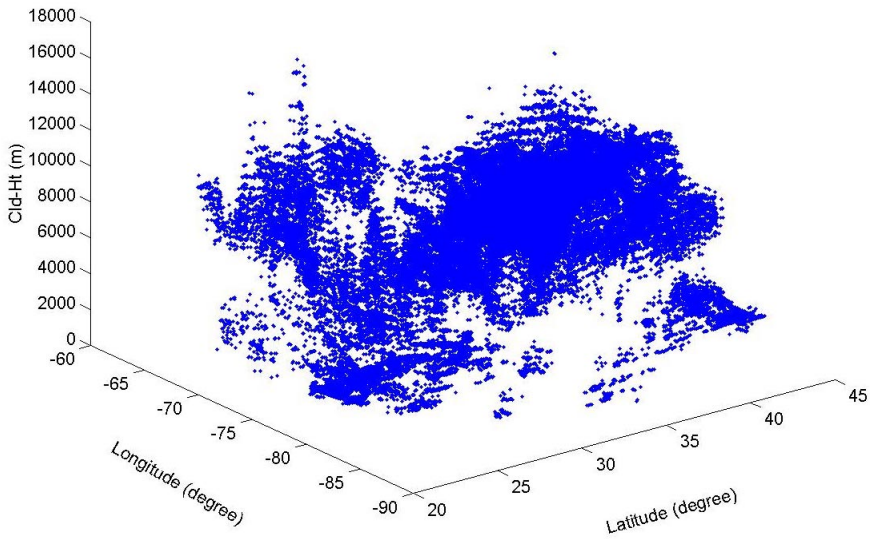

Figure 13. Cloud Top Height image of the MODIS data set used for testing the CC/L processing algorithm

3.4.2 Case Results

CC/L reads in cloudy pixel level cloud property parameters (i.e., cloud mask, cloud phase, cloud top height, cloud optical thickness, and cloud effective particle size) and aggregates them into a HC (with 121 single pixels in this analysis). Note that a final HC size is not determined so a fixed pixel number within each HC is assumed to demonstrate CC/L processing capability. For each HC, CC/L algorithm will perform clustering analysis described in algorithm section 3.3.2. The oblique view correction described in 3.3.2.1 is not performed since this simulated dataset contains no cloud location error. Table 7 displays the cluster/layer results of both initial CTH/Phase and final PDC/K-Means clustering for all pixels of the MODIS granule. As can be seen from the statistics of the initial and final cluster results, the number of pixels switching cluster/layer is limited, indicating that the initial CTH/phase cluster approach is efficient and a single iteration of PDC/K-Means re-clustering is adequate. However, much more rigorous analysis using global multiple scenes that represent diverse atmospheric conditions are necessary to permit the drawing of firm conclusions.

Table 7. "Initial" CTH/Phase and "final" PDC/K-means cluster/layer characteristics.

Cluster/Layer		COT Mean	EPS Mean (μm)	CTH Mean (meter)	Number of Samples	Percentage of cloudy pixels within clust/layer (%)
Ice Upper	**Initial Guess**	24.31	25.17	10576	30542	51.54
	Final Result	25.03	25.38	10589	29103	49.11
Ice Lower	**Initial Guess**	1.91	20.02	4001	35	0.059
	Final Result	9.40	20.09	10168	1475	2.49
Water Upper	**Initial Guess**	11.86	17.68	6257	10311	17.40
	Final Result	11.66	17.90	5732	9048	15.27
Water Lower	**Initial Guess**	4.45	17.11	1147	7140	12.05
	Final Result	5.78	16.96	2480	8403	14.18
Mixed Upper	**Initial Guess**	9.30	21.18	8035	11080	18.70
	Final Result	9.82	21.96	8006	10304	17.39
Mixed Lower	**Initial Guess**	2.57	14.95	4363	153	0.26
	Final Result	2.45	11.50	7755	929	1.57

Scatter plots of EPS and COT for the 6 final clusters are shown in figure 14. These 6 clusters are labeled as Ice_Upper, Ice_Lower, Water_Upper, Water_Lower, Mix_Upper, Mix_Lower, respectively. Although there is no particular significance to this naming convention it does imply that the basic clustering principle is based on the input cloud macro- (CTH and phase) and micro- (COT and EPS) properties. For cluster Ice_Upper it exhibits the largest variance of both COT and EPS. Ice clouds formed in the lower altitude have much smaller variation of COT. Water_Upper cluster, in addition to the altitude attribute, seems to have COT/EPS clustering characteristics similar to those of Ice_Upper. Cluster Water_Lower displays a bi-modal distribution of COT, and EPS is confined to the 10-30 micron range. Cluster Mix_Lower has small COT but has a broad extent of EPS. Although this case study analysis provides some insightful depictions of the cluster attributes, it can by no means represent the broad spectrum of complex cloud systems that exist in the real atmosphere. Continuing efforts to refine this VIIRS CC/L baseline cluster algorithm using a variety of simulated and real datasets are absolutely essential. Results presented here should be considered as algorithm concept demonstration only, and should not be generalized to draw any premature conclusions.

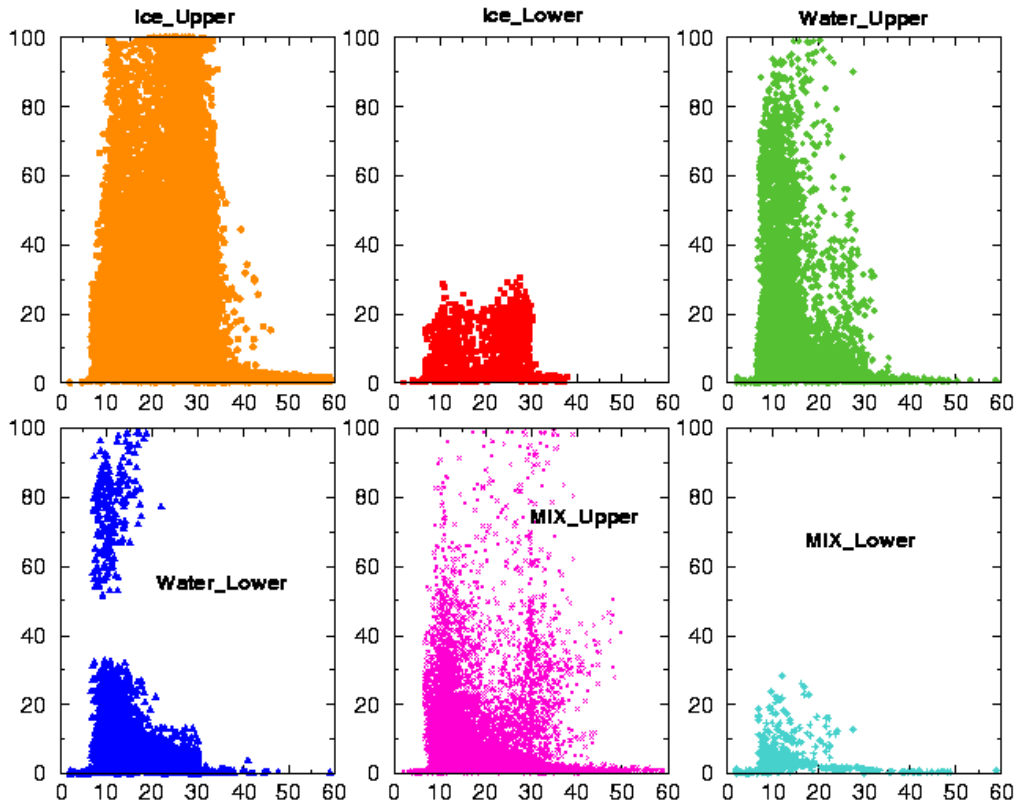

Figure 14. EPS (μm) vs. COT scatter plots of 6 PDC/K Means clusters. Horizontal and vertical axes are EPS and COT, respectively.

Figures 15 and 16 demonstrate two more scatter plots of COT vs. CTH and EPS vs. CTH. The stratification of CTH due to the finite vertical coordinate interval is obvious. In both plots CTH provides a great deal of cluster information, while COT and EPS supply some supplemental information. However, it is difficult to quantify their contribution.

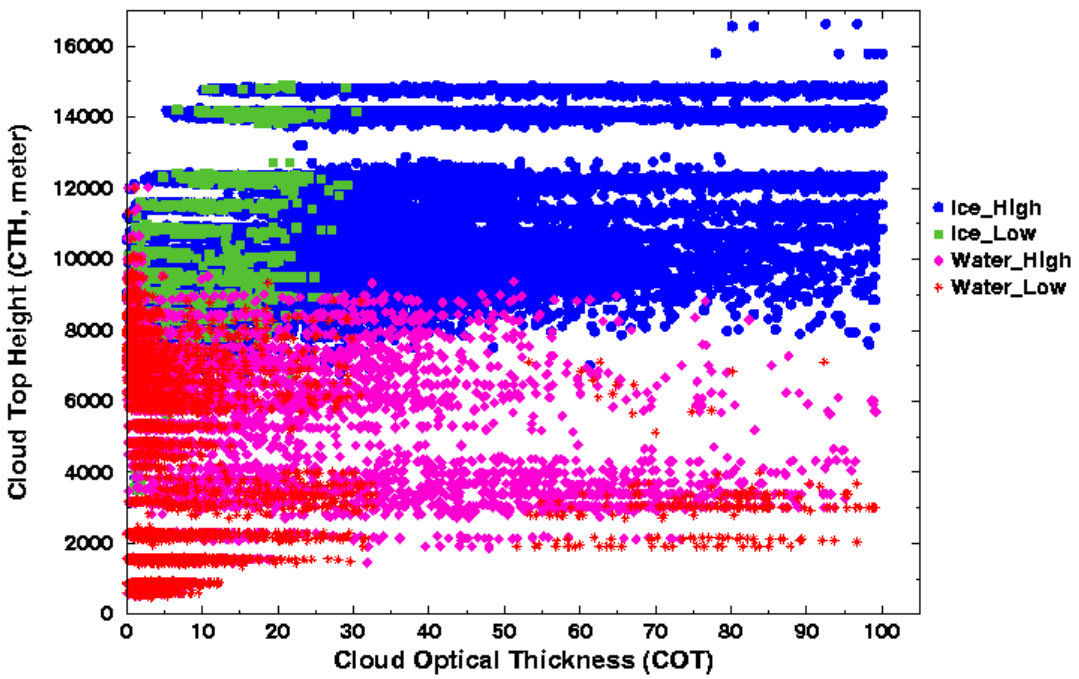

Figure 15. Scatter plot (COT vs. CTH) of 4 final clusters: Ice_Upper, Ice_Lower, Water_Upper, and Water_Lower.

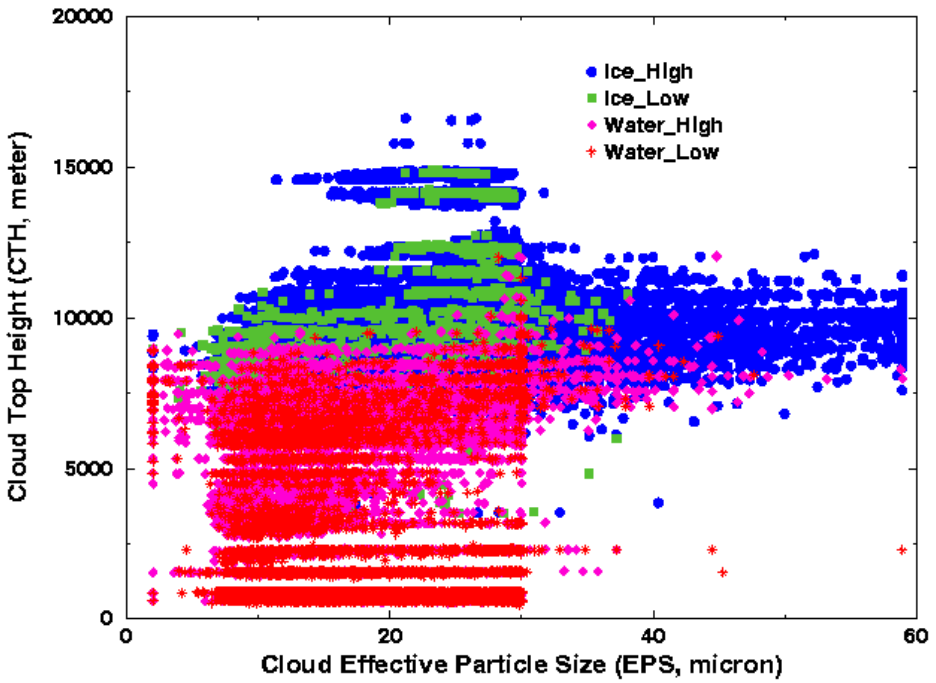

Figure 16. Scatter plot (EPS vs. CTH) of 4 final clusters: Ice_Upper, Ice_Lower, Water_Upper, and Water_Lower.

Figure 17 displays the PDC classification result. Each class is uniquely defined as a distinct layer by PDC/K-Means cluster algorithm. Six classes are represented by different colors, where highest ice clouds are in cyan and lower ice clouds are in blue. Upper and lower water clouds are coded by magenta and red, respectively. Mixed phase clouds are shaded in yellow and green. White areas correspond to clear pixels.

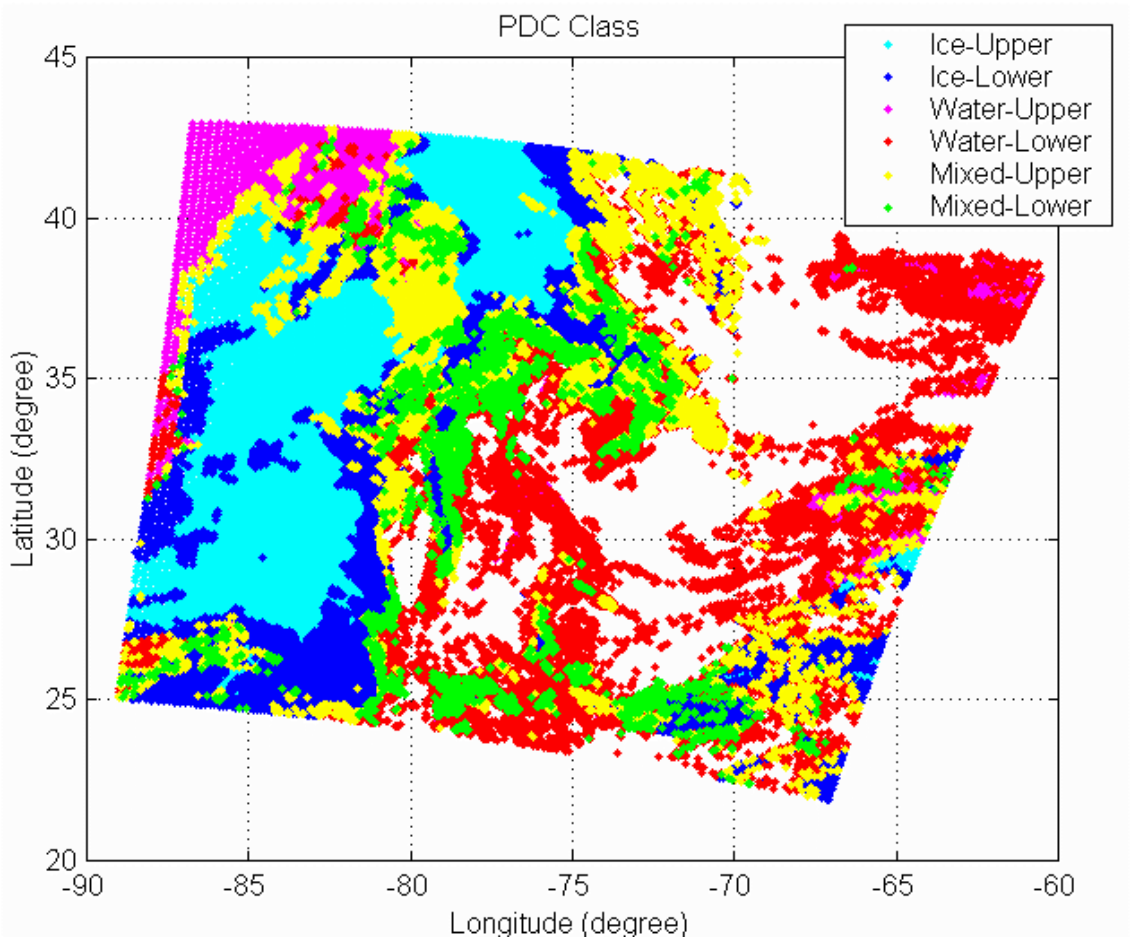

Figure 17. Six PDC classes/layers are shown in different colors.

Figure 18 shows the histogram of cloud fraction derived for HCs (121 single pixels per HC). For this case, 8.7% of all HCs are classified as clear, 62.5% of HCs are partly cloudy, and 28.8% of HCs are overcast (100% cloud fraction).

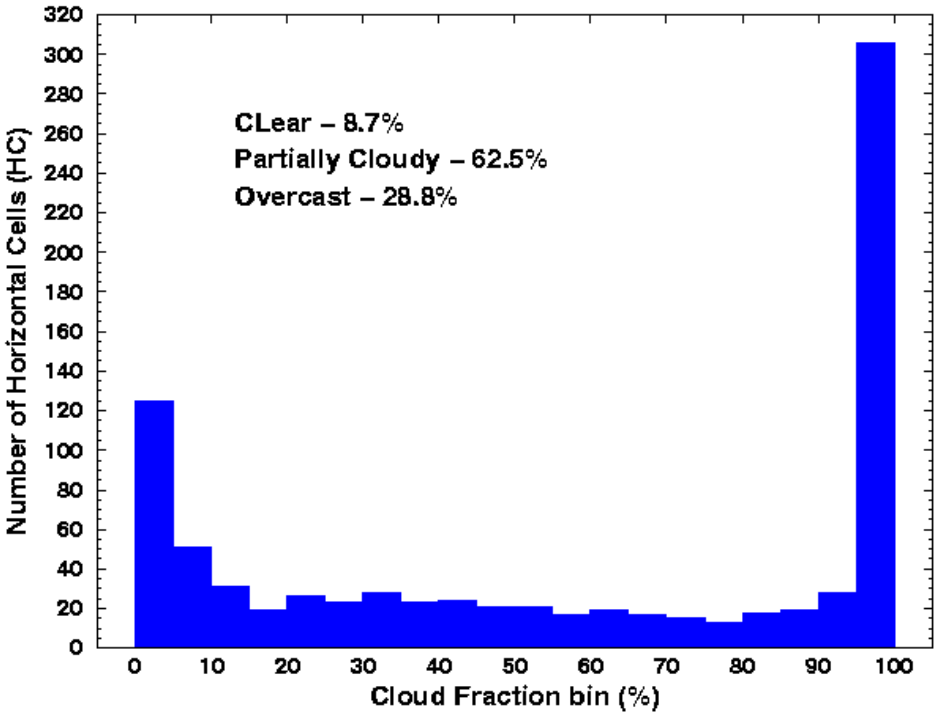

Figure 18. Cloud fraction histogram.

3.5 PRACTICAL CONSIDERATIONS

3.5.1 Numerical Computation Considerations

Paragraph SRDV3.2.1.5.4-1 of the VIIRS SRD states the following:

> "The scientific SDR and EDR algorithms delivered by the VIIRS contractor shall be convertible into operational code that is compatible with a 20 minute maximum processing time at either the DoD Centrals or DoD field terminals for the conversion of all pertinent RDRs into all required EDRs for the site or terminal, including those based wholly or in part on data from other sensor suites."

RDR here stands for Raw Data Record. This essentially means that any and all EDRs must be completely processed from VIIRS raw data, including calibration and geo-referencing within 20 minutes from the time the raw data are available. This requirement is a strong reminder that VIIRS is an operational instrument.

In the CC/L algorithm, numerical approximations to various statistical tests are performed. None of these, however, have unstable numerical properties nor are they applied in an iterative fashion, so as to amplify numerical errors. Therefore, the CC/L algorithm is resistant to numerical problems associated with finite precision arithmetic on computers. CC/L is designed to trade accuracy with computing resources. If the initial CTH/Phase does not provide accurate enough starting partitions, the PDC/K-Means cluster algorithm iterates to improve layer and cloud typing accuracy.

3.5.2 Programming and Procedural Considerations

Research grade code has been developed for the CC/L EDR processing. The algorithm itself is straightforward and easily implemented, and continued refinement will be applied in phases after Critical Design Review (CDR).

3.6 ALGORITHM VALIDATION

The CC/L algorithm will be validated against MAS data and MODIS data along the imagery track centers where lidar cloud profiling data are available. By inspection, a set of ground truth layered cloud amounts will be determined. These will then be compared to CC/L layered assessments and a qualitative indication of CC/L performance can be attained. At least one case of MODIS data will be used to validate CC/L algorithm performance. The EDR will be validated with independent cloud measurements either from space or other indirect means. The required validation data and procedure that can be used for validating algorithm performance can be briefly summarized as:

- Collect statistically significant samples of co-located in-situ cloud layer/type measurements and VIIRS-like measurements.
- Modify/create VIIRS-like measurements with VIIRS instrument specification noise.
- Perform EDR retrieval using ATBD described algorithms.
- Co-register in-situ data and EDR retrievals by taking into account spatial, temporal, and viewing discrepancy.
- Perform statistical accuracy, precision, and uncertainty estimates of EDR using retrievals and in-situ data.

4.0 ASSUMPTIONS AND LIMITATIONS

4.1 ASSUMPTIONS

4.1.1 Spatial Domain of Processing

An assumption of the appropriate domain of processing must be made. This assumption will dictate many characteristics of an operational production system and needs to be investigated further. Currently the processing domain is determined by a HC size of 11 pixels by 11 pixels. Under this processing assumption, the computationally largest blocks of data are passed to the algorithm for processing. These data blocks may correspond to fractional orbits of imagery-derived EDRs for a polar-orbiting sensor, each covering a very large geographic region. Each stage of the algorithm is applied to the entire region at once. Smaller sections of the available imagery are processed in turn. These may correspond to blocks of scan lines received in short, real-time increments. Processing is confined to these regions, and slight discontinuities between regions may result. Individual horizontal cells or small sets of horizontal cells are processed as entities. Maximal parallelization of processing may be achieved. Adjacent horizontal cells may not be entirely consistent.

4.1.2 Cloud Type Definition

Cloud typing definition is based on current general knowledge of cloud microphysical properties. Definition of cloud type is subject to being updated when newer and more sophisticated cloud property measurements can reveal any improved cloud type information.

4.1.3 Processing Constraints

The PDC/K-Means cluster approach is designed to be iterative. However, because processing resources are limited and the baseline processing CC/L EDR can meet the EDR requirements, the procedure will initially be implemented without iteration.

4.2 LIMITATIONS

4.2.1 Inherent Cluster Ambiguity

There is no best mathematical classification of data into clusters. Clustering is a fundamentally soft subject, where many different and equally justifiable interpretations of data are present. Any single partition of a dataset is inherently ambiguous. Rescaling, rotations, arbitrary coordinate transforms, choices of distance metrics, selection of the cluster integrity measures, and other design decisions determine the performance of an algorithm.

4.2.2 Definition of Cloud Layers/Types

Various definitions of cloud layers/types are defined and discussed. The choice of cloud layer definition dramatically affects the design and implementation of a cloud layer algorithm. The algorithm inputs ultimately limit the clustering capability.

5.0 REFERENCES

Arking A., and J. D. Childs, 1985: Retrieval of Cloud Cover Parameters from Multispectral Satellite Images. Journal of Climate and Applied Meteorology, 24, 322-333.

Burrough, P. A., 1986: Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press.

Desbois, M., G. Seze and Szejwach, 1982: Automatic classification of clouds on METEOSAT imagery: Application to high-level clouds. Journal of Applied Meteorology, 21, 401-412.

Dowling, D.R., and L.F. Radke, 1990: A summary of the physical properties of cirrus clouds, J. Appl. Meteor., 29, 970-978.

Heymsfield, A.J., and C.M.R. Platt, 1984: A parameterization of particle size spectrum of ice clouds in terms of the ambient temperature and ice water content. J. Atmos. Sci., 41, 846-855.

Liou, K.N., 1992: Radiation and cloud processes in the atmosphere, Oxford University Press, New York, pp.487.

MacQueen, J., 1967: Some methods for classification and analysis of multivariate observations. In Le Cam, L. M. and Neyman, J., editors, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume I: Statistics, pages 281-297. University of California Press, Berkeley and Los Angeles, CA.

Planet, W.G. (ed.), 1988: Data extraction and calibration of TIROS-N/NOAA radiometers. NOAA Technical Memorandum NESS 107 – Rev. 1, Oct. 1988. 130 pp.

Selim S.Z., and M A Ismail, 1984: "K-means-type algorithms: A generalized convergence theorem and characterisation of local optimality", IEEE Transactions on Pattern Analysis and Machine Intelligence., 6, 1, pp. 81-87.

Theiler J., and G. Gisler, 1997: A contiguity-enhanced k-means clustering algorithm for unsupervised multispectral image segmentation, Proc SPIE 3159, 108-118.

Weickmann H.K., and H.J. Aufm Kampe, 1953: Physical properties of cumulus clouds. J. Meteor., 10, 204-211.