JPL D-13397, Rev. B

Earth Observing System

Multi-angle
Imaging
Spectro-
Radiometer

MISR

Level 1 Cloud Detection Algorithm Theoretical Basis

David J. Diner[1]
Larry Di Girolamo[2]
Eugene E. Clothiaux[3]

[1]Jet Propulsion Laboratory, California Institute of Technology
[2]University of Illinois
[3]Pennsylvania State University

JPL

Jet Propulsion Laboratory
California Institute of Technology

December 7, 1999

JPL D-13397, Rev. B

Multi-angle Imaging SpectroRadiometer (MISR)

Level 1 Cloud Detection Algorithm Theoretical Basis

Approval:

David J. Diner
MISR Principal Investigator

The MISR web site should be consulted to determine the latest released version
of this document (http://www-misr.jpl.nasa.gov).
Approval signatures are on file with the MISR Project.

JPL

Jet Propulsion Laboratory
California Institute of Technology

TABLE OF CONTENTS

GLOSSARY OF ACRONYMS

A

AGP (Ancillary Geographic Product)
ASCM (Angular Signature Cloud Mask)
ATB (Algorithm Theoretical Basis)
AU (Astronomical Unit)
AZM (Azimuthal Model)

B

BRF (Bidirectional Reflectance Factor)

C

CCD (Charge-Coupled Device)
ClearHC (Clear with High Confidence)
ClearLC (Clear with Low Confidence)
CloudHC (Cloud with High Confidence)
CloudLC (Cloud with Low Confidence)
CSSC (Cloud Screening Surface Classification)

D

DSVI (D-parameter Spatial Variability Index)

E

ECS (EOSDIS Core System)
EOSDIS (Earth Observing System Data and Information System)
EPA (Environmental Protection Agency)

I

IFOV (Instantaneous Field Of View)
IR (Infrared)
ISCCP (International Satellite Cloud Climatology Project)

M

MISR (Multi-angle Imaging SpectroRadiometer)

N

NDVI (Normalized Difference Vegetation Index)
NOAA (National Oceanic and Atmospheric Administration)

R

RC (Radiometric Camera-by-camera)
RCCM (Radiometric Camera-by-camera Cloud Mask)
RDQI (Radiometric Data Quality Indicator)
RLRA (Reflecting Level Reference Altitude)

S

SCF (Science Computing Facility)
SDCM (Stereoscopically Derived Cloud Mask)
SDP (Science Data Production)
SERCAA (Support of Environmental Requirements for Cloud Analysis and Archive)
SVI (Spatial Variability Index)

T

TOA (Top-of-Atmosphere)

W

WGS (World Geodetic System)

1. INTRODUCTION

1.1 PURPOSE

This Algorithm Theoretical Basis (ATB) document describes the algorithms used to retrieve the Radiometric Camera-by-camera Cloud Mask (RCCM) within the MISR Level 1B2 Geo-rectified Radiance Product. These parameters are summarized in Table 1. In particular, this document identifies sources of input data, both MISR and non-MISR, required for parameter retrievals; includes implementation details; and provides the physical theory and mathematical background underlying derivation of the RCCM. This ATB is used by the MISR Science Data System Team to establish requirements and functionality of the data processing software.

Table 1: RCCM parameters within the Level 1B2 Geo-rectified Radiance Product

Parameter name	Units	Horizontal Sampling and Coverage	Comments
Radiometric Camera-by-camera Cloud Mask (RCCM)	none	1.1 km (Global)	• Calculated on a camera-by-camera basis • Indicates one of five designations: 　--Cloud with high confidence (CloudHC) 　--Cloud with low confidence (CloudLC) 　--Clear with low confidence (ClearLC) 　--Clear with high confidence (ClearHC) 　--No Retrieval (NR) • Also contains glitter flag: 　--Not glitter contaminated 　--Glitter contaminated
RCCM quality flag	none	1.1 km (Global)	• Associated with the RCCM • Indicates one of four designations: 　--No Retrieval 　--Secondary test used only 　--Primary test used only 　--Both primary and secondary tests used

In addition, this document describes three ancillary datasets which are used as part of RCCM generation, the Radiometric Camera-by-camera (RC) Histogram Dataset, the RC Threshold Dataset, and the Cloud Screening Surface Classification (CSSC) Dataset. The first contains histograms of observables, derived from MISR data, which are used to generate updated values of the thresholds used in deriving the RCCM. The second contains values of the static, seasonal, and monthly thresholds used to determine whether a particular 1.1-km area is to be classified as cloudy or clear. The third provides a surface classification required for allocating the appropriate thresholds for clear/cloud discrimination.

1

1.2 SCOPE

This document covers the algorithm theoretical basis for the parameters of the RCCM, the RC Histogram Dataset, and the RC Threshold Dataset, as well as a description of the CSSC Dataset. Current development and prototyping efforts may result in modifications to parts of certain algorithms. Only the algorithms which will be implemented at the DAAC for routine processing will be preserved in the final release of this document.

Chapter 1 describes the purpose and scope of the document. Chapter 2 provides a scientific and historical background. The processing concept and algorithm description is presented in Chapter 3. Chapter 4 summarizes assumptions and limitations. Literature references used throughout the document are provided in Chapter 5. Within the text, these references are indicated by a number in italicized square brackets, e.g., *[1]*.

1.3 MISR DOCUMENTS

A listing of MISR Project documents referenced within the text is provided below. References are indicated by a number in italicized square brackets as follows, e.g., *[M-1]*. The MISR web site (http://www-misr.jpl.nasa.gov) should be consulted to determine the latest released version of each of these documents.

[M-1] Experiment Overview, JPL D-13407.

[M-2] Data Product Description, JPL D-11103.

[M-3] Level 1 Radiance Scaling and Conditioning Algorithm Theoretical Basis, JPL D-11507.

[M-4] Level 1 Georectification and Registration Algorithm Theoretical Basis, JPL D-11532.

[M-5] Level 1 In-flight Radiometric Calibration and Characterization Algorithm Theoretical Basis, JPL D-13398.

[M-6] Level 1 Ancillary Geographic Product Algorithm Theoretical Basis, JPL D-13400.

[M-7] Level 1 In-flight Geometric Calibration Algorithm Theoretical Basis, JPL D-13399.

[M-8] Level 2 Cloud Detection and Classification Algorithm Theoretical Basis, JPL D-11399.

[M-9] Level 2 Top-of-Atmosphere Albedo Algorithm Theoretical Basis, JPL D-13401.

[M-10] Level 2 Aerosol Retrieval Algorithm Theoretical Basis, JPL D-11400.

[M-11] Level 2 Surface Retrieval Algorithm Theoretical Basis, JPL D-11401.

[M-12] Level 2 Ancillary Products and Datasets Algorithm Theoretical Basis, JPL D-13402.

[M-13] Algorithm Development Plan, JPL D-11220.

[M-14] In-flight Radiometric Calibration and Characterization Plan, JPL D-13315.

[M-15] In-flight Geometric Calibration Plan, JPL D-13228.

[M-16] Science Data Validation Plan, JPL D-12626.

[M-17] Science Data Processing Sizing Estimates, JPL D-12569.

1.4 REVISIONS

The original version of this document was contained within the Level 2 TOA/Cloud ATB dated March 3, 1994 and revised on December 19, 1994. Subsequently, radiometric camera-by-camera cloud masking was moved from Level 2 processing to Level 1 processing. The original release of this document was dated August 15, 1996. Revision A was released November 25, 1997. This release is Revision B. Modifications from the previous version are denoted by change bars, as shown at left.

2. EXPERIMENT OVERVIEW

2.1 MISR INSTRUMENT

The MISR instrument consists of nine pushbroom cameras. From the 705-km descending polar orbit, the overlap swath width of the MISR imaging data (that is, the swath seen in common by all nine cameras) is 360 km, which provides global multi-angle coverage of the entire Earth in 9 days at the equator, and 2 days at the poles. The cameras are arranged with one camera pointing toward the nadir (designated An), one bank of four cameras pointing in the forward direction (designated Af, Bf, Cf, and Df in order of increasing off-nadir angle), and one bank of four cameras pointing in the aftward direction (using the same convention but designated Aa, Ba, Ca, and Da). Images are acquired with nominal view angles, relative to the surface reference ellipsoid, of 0°, 26.1°, 45.6°, 60.0°, and 70.5° for An, Af/Aa, Bf/Ba, Cf/Ca, and Df/Da, respectively. Each camera uses four Charge-Coupled Device (CCD) line arrays in a single focal plane. The line arrays consist of 1504 photoactive pixels plus 16 light-shielded pixels per array, each 21 μm by 18 μm. Each line array is filtered to provide one of four MISR spectral bands. The spectral band shapes are approximately gaussian, and centered at 446, 558, 672, and 866 nm. The combination of nine cameras and four spectral bands yields a total of 36 channels within the instrument.

The crosstrack IFOV and sample spacing of each pixel is 275 m for the all of the off-nadir cameras, and 250 m for the nadir camera. Downtrack IFOV's depend on view angle, ranging from 214 m in the nadir to 707 m at the most oblique angle. Sample spacing in the downtrack direction is 275 m in all cameras. An on-board capability within the instrument enables averaging 4 x 4 arrays of samples on a channel-by-channel basis.

2.2 OBJECTIVES OF RCCM GENERATION

The RCCM is used for several purposes during MISR geophysical parameter retrievals. Within the TOA/Cloud Product, the RCCM is combined with the Stereoscopically Derived Cloud Mask (SDCM) to (1) establish final, reported values of the Reflecting Level Reference Altitude (RLRA), (2) determine whether a scene is classified as clear or cloudy for the purpose of choosing the angular integration coefficients (azimuthal models, or AZM's) which are used in estimating TOA albedos, and (3) calculate regional scene classifiers *[M-8]*. Retrieval of aerosol and surface properties within the Aerosol/Surface Product requires the absence of clouds, in order that the assumptions inherent in the retrievals are not invalidated *[M-10]*, *[M-11]*.

2.3 HISTORICAL PERSPECTIVE

Cloud screening involves discriminating between clear and cloudy pixels in an image. Reviews of cloud detection methods can be found in *[8]*, *[14]*, and *[15]*. Methods for identifying clouds are generally based on radiance threshold, radiative transfer model, or statistical techniques making use of spectral and textural features in the imagery. Radiance threshold techniques work

on a pixel-by-pixel basis, and single or multiple-channel thresholds are defined which are then used to divide clear and cloudy pixels. Radiative transfer model techniques use one or more spectral radiance measurements as input to an atmospheric radiative transfer model and retrieve a physical quantity such as cloud optical thickness or altitude. The pixels are then determined to be clear or cloudy based on thresholds in the retrieved quantity. Statistical techniques use groups of adjacent pixels. Among these are methods based on spatial coherency between adjacent pixels [2], neural networks (e.g., [17]), maximum likelihood decision rules (e.g., [6]), and clustering routines (e.g., [7]). Specific cloud detection algorithms applied to satellite data generally have features which are beneficial for a particular scene class.

2.4 MISR APPROACH

Traditional cloud screening methods applied to each MISR camera is a challenging problem given the small number of spectral channels available, none of which are longward of 1 μm. As a result, only a few simple cloud detection observables can be constructed from arithmetic operations on the camera radiances. This requires good estimates of the thresholds that will discriminate clear skies from cloudy skies.

The few observables that are used by the RCCM to determine clear vs. cloudy depend on whether the observations are made over water or land. In case of cloud detection over water, the observables are r_4, the bidirectional reflectance factor (BRF) in band 4 (near-IR) at 1.1 km resolution, and σ_3, the standard deviation of the 4 x 4 array of 275 m band 3 (red) BRF's within a 1.1 km area. Over land, the observables are the D parameter, derived from the band 4 and band 3 BRF's, and the D spatial variability index (DSVI), calculated from a 3 x 3 array of 1.1 km samples and assigned to the middle one.

Each observable will be tested by comparing to three thresholds in order to classify the pixel as cloud with high confidence (CloudHC), cloud with low confidence (CloudLC), clear with low confidence (ClearLC), or clear with high confidence (ClearHC). The Support of Environmental Requirements for Cloud Analysis and Archive (SERCAA) algorithm is a recent example that applies confidence levels to cloud masking [9]. The thresholds are derived from time cumulated statistics, for a particular observable, over a particular geographical region. For example, Figure 1 shows a hypothetical example of a one-dimensional frequency histogram for an observable, Q. The observables for water tend to have a peak at low values representing clear conditions, whereas the land observables tend to have a peak at low values representing cloud. In either case, the histogram may often be unimodal. The thresholds are derived from the histogram using an automated procedure. However, the thresholds are not backward-propagated through a routine reprocessing step (e.g., as done by ISCCP) due to the MISR processing strategy; instead, they are applied to the next set of observations that fall within the threshold bin.
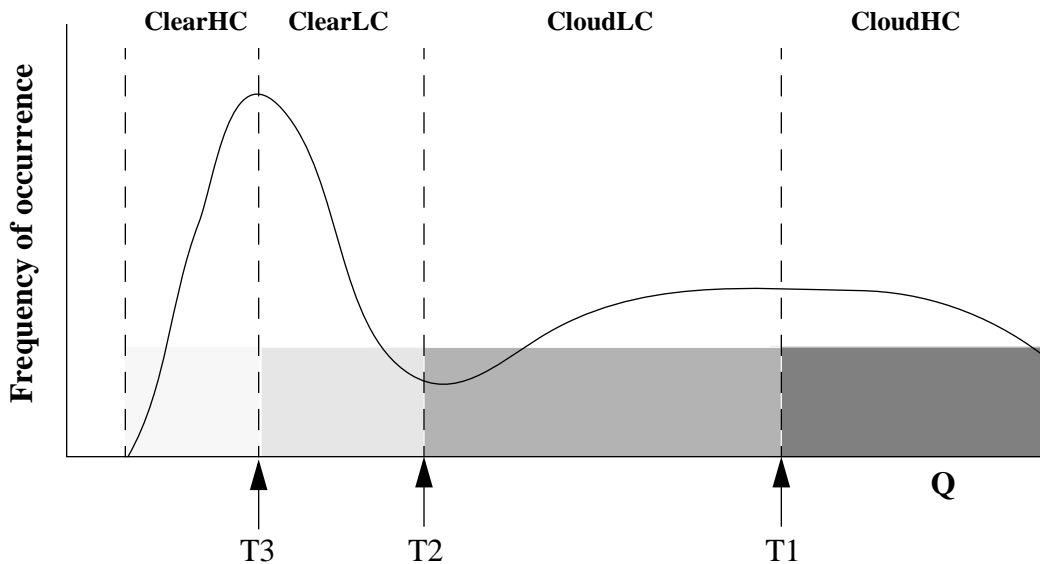
Figure 1. Schematic view of cloud thresholding

The thresholds are a function of view angle, sun angle, relative view-sun azimuth angle, time, and place on globe (or surface type). The time dependence can be of 3 kinds: (1) time independent (static), (2) dynamic (seasonal), and (3) dynamic on a less than seasonal time scale. The latter is approximately monthly, on a sliding window. The values are stored on a global grid system in which the tessellation varies with geographic position. The grid size will be based on the uniformity (in time and space) of surface properties. The global grid system is defined in an ancillary dataset known as the Cloud Screening Surface Classification Dataset.

Further background on the MISR investigation is provided in *[M-1]*.

3. ALGORITHM DESCRIPTION

3.1 PROCESSING OUTLINE

Generation of the MISR Radiometric Camera-by-camera Cloud Mask (RCCM) occurs routinely as part of Level 1B2 processing, and involves several activities. First, prior to flight, we create at the MISR Science Computing Facility (SCF) a version of the RC Threshold Dataset containing nominal values for the static thresholds which will be used early in the mission. This dataset is delivered to the DAAC. During the first several months of the mission, MISR data will be used to revise the static values of the thresholds, and a new version of the RC Threshold Dataset will be delivered from the SCF to the DAAC. In addition, dynamically updated histograms of the observables used in MISR radiance-based cloud detection will be generated at the DAAC, and stored in the RC Histogram Dataset. Once the RC Histogram Dataset has been updated, automated histogram analysis is performed in order to generate dynamic values of the cloud masking thresholds, on seasonal and monthly timescales. These updated values are stored in the RC Threshold Dataset. In all phases, the latest version of the RC Threshold Dataset is used to generate the RCCM.

All RCCM processes, occur by definition on a camera-by-camera basis. Thus, Level 1B2 processing does not include calculation of those observables used in the generation of the Stereoscopically Derived Cloud Mask (SDCM) or the Angular Signature Cloud Mask (ASCM), which make use of multiple cameras. The SDCM and ASCM are generated as part of Level 2 TOA/Cloud Product generation *[M-8]*.

Processing flow concepts are shown diagrammatically throughout the document. The convention for the various elements displayed in these diagrams is shown in Figure 2.
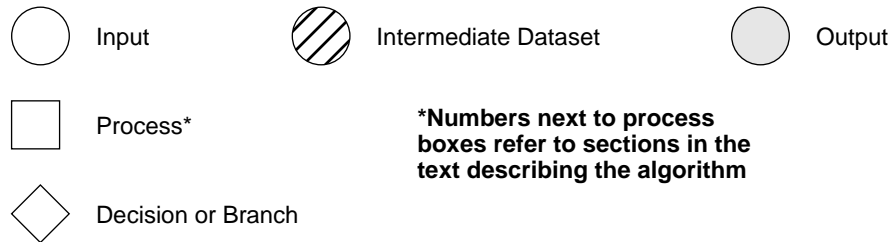
Figure 2. Conventions used in processing flow diagrams

An overview of the RCCM generation process is shown in Figure 3. Calculation of the observables for generation of the RCCM and the RC Histogram Dataset occur in parallel. The flow between update of the RC Histogram Dataset, updating of the thresholds, and generation of an updated RC Threshold Dataset is shown as a dashed line to indicate that requires accumulation of data of a period of time, and thus occurs on a different time schedule than the processing which generates the RCCM.

7

Figure 3. RCCM processing overview

3.2 ALGORITHM INPUT

The required inputs for RCCM generation come from MISR and non-MISR sources and are summarized individually in the following paragraphs. The MISR data either come directly from the MISR processing stream, or consist of relatively static inputs, generated pre-flight by the Science Team.

3.2.1 MISR data

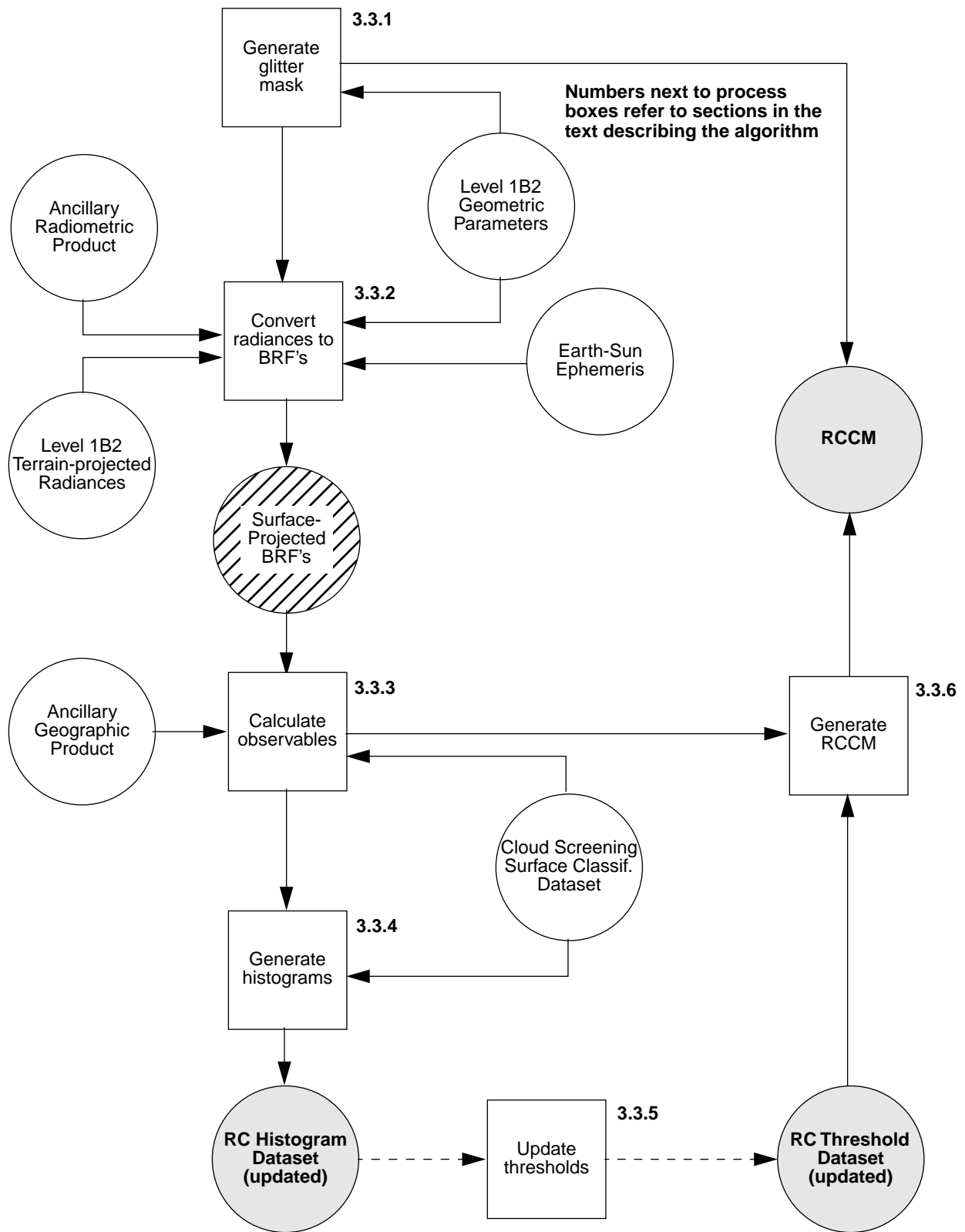Required inputs for the RCCM generation to be obtained from the MISR instrument or from the MISR team are summarized in Table 2. Further information on each of the inputs is provided below.

Table 2: RCCM Generation Inputs (MISR Data)

Input data	Source of data
Ellipsoid-referenced geometric parameters	Level 1B2 Geo-rectified Radiance Product
Terrain-projected TOA radiances	Level 1B2 Geo-rectified Radiance Product
Radiometric Data Quality Indicators (RDQI's)	Level 1B2 Geo-rectified Radiance Product
Exo-atmospheric solar spectral irradiances	Ancillary Radiometric Product
Land/water mask identifier	Ancillary Geographic Product

3.2.1.1 Ellipsoid-referenced geometric parameters

These include illumination and view zenith and azimuth angles relative to the surface normal of the World Geodetic System 1984 (WGS84) reference ellipsoid. Azimuth angles are referenced to local North. These inputs are obtained from earlier stages of the Level 1B2 processing *[M-4]*.

3.2.1.2 Terrain-projected TOA radiance

Terrain-projected TOA radiances are described in *[M-4]* and are calculated during earlier stages of the Level 1B2 processing. Over ocean, terrain-projected radiances are actually stored in the Level 1B2 data as ellipsoid-projected radiances, but they are identical in this case.

3.2.1.3 Radiometric Data Quality Indicators

Radiometric Data Quality Indicators (RDQI's) are described in *[M-4]* and are calculated during earlier stages of the Level 1B2 processing. The RDQI's are associated with each value of radiance reported at Level 1B2, and take on values of 0 - 3, as follows:

RDQI = 0: Radiometric accuracy meets all specifications

RDQI = 1: Radiometric accuracy is sufficient for certain applications but some specifications are violated (see *[M-3]* and *[M-4]*)

RDQI = 2: Radiance value is available but of insufficient accuracy to be used in Level 2 retrievals

RDQI = 3: Radiance value is unavailable.

Thus, higher quality data are associated with smaller values of RDQI.

3.2.1.4   Exo-atmospheric solar spectral irradiances

These are required to convert radiances to bidirectional reflectance factors, and are obtained from the MISR Ancillary Radiometric Product *[M-5]*. There are four values, one for each of the MISR spectral bands, and weighted by the instrument spectral response.

3.2.1.5   Land/water mask

This is a land/ocean/inland water/ephemeral water/coastline mask obtained from the MISR Ancillary Geographic Product (AGP). Additionally, the AGP identifies which ocean and inland water is classified as deep water through the "dark water algorithm suitability" designation. This is required because the thresholds used for deep water and shallow water are derived on different time scales. The AGP data are provided on 1.1-km centers. The AGP is generated at the MISR SCF and stored at the DAAC. Further details of the AGP are provided in *[M-6]*.

3.2.2   Non-MISR data

Inputs for RCCM generation obtained from non-MISR sources are summarized in Table 3.

Table 3: RCCM Generation Inputs (Non-MISR Data)

Input data	Source of data
Surface classification	Cloud Screening Surface Classification Dataset
Earth-Sun ephemeris	SDP Toolkit

3.2.2.1   Surface classification

The Cloud Screening Surface Classification (CSSC) Dataset defines unique surface types in order to allocate the appropriate thresholds for clear/cloud discrimination. Additional information on the CSSC Dataset is provided in §3.5.3.

10

3.2.2.2 Earth-Sun ephemeris

This is used to obtain the Earth-Sun distance, such that observed radiances can be normalized to the standard distance of 1 AU. The source is the Science Data Production (SDP) Toolkit, which is generated by the EOSDIS Core System (ECS) contractor.

3.3 THEORETICAL DESCRIPTION
3.3.1 Generate glitter mask

3.3.1.1 Physics of the problem

Regions determined to be possibly contaminated by sun glitter at particular view angles will be included in the generation of histograms and cloud masks; however, during RCCM generation they will also carry a glitter contamination flag. This designation will be made use of in higher-level processing. In regions flagged as glitter-contaminated, the classification assigned by the RCCM may be suspect; a classification as clear is probably valid, whereas a classification as cloudy may actually be due to the presence of glitter, rather than cloud.

For the purpose of generating a glitter mask, we define a cone of half-angle ξ_c, centered on the specular reflection direction. Any view vector which falls within this cone is flagged as glitter contaminated. The flag is included in the RCCM. The algorithm for determining whether to flag a view as glitter-contaminated is as follows.

3.3.1.2 Mathematical description of the algorithm

Define a local right-handed Cartesian coordinate system in which the +z-axis is co-aligned with the normal to the Earth's ellipsoid and points toward the center of the Earth, the x-axis is aligned with a great circle and points toward the geographic north pole, and the y-axis is orthogonal to both of these (i.e., points East). Let $\hat{s}$ be the unit vector pointing in the direction of the Sun's rays, in which θ_0 is the solar zenith angle (relative to the z-axis) and ϕ_0 is the solar azimuth angle (relative to the x-axis). Then, the vector which describes the specular reflection direction, $\hat{r}$, has polar angles 180° - θ_0 and ϕ_0. Let the view vector $\hat{v}$ point in the direction of photon travel, with an elevation angle θ relative to the z-axis and azimuth angle ϕ. Let $\mu_0 = \cos\theta_0$ and $\mu = \cos(180° - \theta) = -\cos\theta$. The cone angle between the specular direction and the view direction, ξ, is given by:

$$\cos\xi = \hat{r} \cdot \hat{v} = \mu\mu_0 + (1-\mu^2)^{\frac{1}{2}}(1-\mu_0^2)^{\frac{1}{2}}\cos(\phi-\phi_0) \qquad (1)$$

Then, if $\xi \leq \xi_c$, the view is considered to be potentially glitter-contaminated. To be conservative, we choose $\xi_c = 30°$.

11

3.3.2 Convert radiances to BRF's

3.3.2.1 Physics of the problem

The purpose of this step is to convert TOA radiances to BRF's. The conversion is the same for all averaging modes of the Level 1B2 data, and independent of whether the input radiances are projected to 30-km altitude or to the surface. Thus, these BRF's are used as intermediate datasets prior to generation of output parameters, (e.g. BRF referenced to the RLRA). A pre-calculated solar irradiance, tailored to the spectral response of the MISR instrument, is used for each of the four spectral bands. The BRF is defined as the observed radiance at a particular illumination and view geometry divided by the radiance that would be obtained under the same conditions with a target consisting of a perfectly reflecting lambertian surface. The primary process involved is normalization to the solar input.

3.3.2.2 Mathematical description of the algorithm

The standardized band-weighted solar irradiance for each of the MISR bands for an Earth-Sun distance of 1 AU is obtained from the MISR Ancillary Radiometric Product. Determination of BRF r in band b is then given by:

$$r(-\mu, \mu_0, \phi - \phi_0) = \frac{\pi L(-\mu, \mu_0, \phi - \phi_0) \cdot d^2}{\mu_0 E_{0, b}^{std}} \qquad (2)$$

where L is the spectral radiance recorded in band b, and d is the distance to the Sun, in AU, at the time of observation. Note that according to the geometric definitions provided in §3.3.1, the convention $-\mu$ indicates that the radiation is traveling in the upward direction.

3.3.3 Calculate observables

3.3.3.1 Physics of the problem

3.3.3.1.1 Establish surface type

Different observables are used depending on whether the surface is water or land. The AGP and CSSC are used for this classification. If the AGP indicates water, then this designation is used. If the AGP indicates land, the corresponding land surface classification from the CSSC is used. In the event of a conflict, i.e., if the AGP indicates land while the CSSC indicates water, a nearest neighbor search through the CSSC is performed to find the closest land class within a specified search window area.

3.3.3.1.2 Water

Several steps are used to separate haze from cloud and partial cloud. The near-IR band (band 4) is chosen for the initial thresholding steps because it is less sensitive to changes in atmospheric and oceanic constituents compared to the other MISR bands. For example, it lies outside the Chappuis bands of ozone. The aerosol effect is also larger at shorter wavelengths, and water-leaving radiance is negligible in band 4. Additionally, the standard deviation of high-resolution band 3 BRF's is used to identify partially cloudy regions.

3.3.3.1.3 Land

Over land, we take advantage of the surface's spectral signature as well as the spectral signature's spatial variability. The spectral signatures are optimized to provide the maximum separation between clear and cloudy values *[3]*, *[4]*.

3.3.3.2 Mathematical description of the algorithm

3.3.3.2.1 Establish surface type

The algorithm for resolving conflicts in surface classification between the AGP and CSSC is as follows. If the AGP classifies the surface as land, and the CSSC indicates water, a square extending N_{search} CSSC samples to the north, east, south, and west from the primary CSSC sample (i.e., the one within which the AGP sample falls) is constructed. This square is thus $2N_{search} + 1$ CSSC samples on a side. Within this square the closest CSSC land sample is used to establish the land classification. The measure of proximity is minimization of

$$distance = \sqrt{(\Delta lat)^2 + (\Delta lon)^2} \tag{2a}$$

where Δlat and Δlon, respectively, are the latitude and longitude difference between the center of each CSSC sample being searched and the center of the 1.1-km sample on the SOM grid for which the surface classification is being established. N_{search} is a configurable parameter, and is currently set to 20. If this procedure still fails to identify a land class, a default land classification is used. The most common situation when the default CSSC class is called will be for very small tropical islands. The default land identification is the same value as used for the Brazilian rainforest.

3.3.3.2.2 Water

There are two cloud screening observables, Q, for water surfaces, namely r_4, the bidirectional reflectance factor (BRF) in band 4 (near-IR) at 1.1 km resolution, and σ_3, the standard deviation of the 4 x 4 array of 275-m band 3 (red) BRF's within a 1.1 km area.

Only values of r_4 which are associated with an RDQI $\leq$ RDQI$_0$, where RDQI$_0$ is a pre-specified threshold, are used. At present, we set RDQI$_0 = 0$, that is, only the best quality data are used. In calculating a value of σ_3, up to 16 values of r_3 are potentially available for calculating the standard deviation. We include in the calculation only those values of r_3 that are associated with an RDQI $\leq$ RDQI$_1$, where RDQI$_1$ is a pre-specified threshold, and require at least n_1 out of the possible 16 values to satisfy this criterion. No weighting by RDQI is used in the calculation of σ_3. At present, we set RDQI$_1 = 0$, that is, only the best quality data are included in the calculation, and we set $n_1 = 9$. If less than n_1 values of r_3 are present, a value of σ_3 is not calculated and the secondary cloud detection test is not applied. In calculating σ_3 from n values of r_3, the following formula is used:

$$\sigma_3 = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (r_3 - \bar{r}_3)^2} \qquad (2b)$$

where $\bar{r}_3$ is the average of the 275-m r_3 values within a 1.1-km subregion. We include in the calculation of $\bar{r}_3$ only those values of r_3 that are associated with an RDQI $\leq$ RDQI$_1$. No weighting by RDQI is used in the calculation of $\bar{r}_3$.

3.3.3.2.3 Land

Over land, we use a parameter, D, derived from the top-of-atmosphere Normalized Difference Vegetation Index (NDVI) and the band 3 reflectance, and the D spatial variability index (DSVI), calculated from a 3 x 3 array of 1.1 km samples and assigned to the middle one.

The D-parameter is defined by

$$D = \frac{|\text{NDVI}|^b}{(\bar{r}_3)^2} \qquad (3)$$

where NDVI is the Normalized Difference Vegetation Index equal to $(r_4 - \bar{r}_3)/(r_4 + \bar{r}_3)$, $\bar{r}_3$ is the average of the 275-m r_3 values within a 1.1-km subregion, and b is chosen so as to maximize the distinction between clear-sky and cloudy D values [4]. The choice of b also tends to maximizes the spatial variability of clear-sky D-values, thus allowing statistical cloud detection techniques to work effectively. The statistical technique used is the Spatial Variability Index, or SVI [1]. For each sample, the index is derived by calculating the absolute value of the difference between D in a single sample and the mean value of the 3 x 3 matrix centered on that sample, i.e., DSVI = $|D_m - D_c|$, where D_m is the mean D-value of the 3 x 3 matrix centered on the value D_c. The D-values used in calculating DSVI make use of the appropriate b-value associated with the CSSC classification of each sample. For non-vegetated surfaces $b = 0.4$; whereas for vegetated surfaces $b = 0.6$. These

14

values are obtained from a limited set of AVHRR and AVIRIS imagery and may be modified in the post-launch era. Samples classified by the AGP as water are not included in the calculation for DSVI.

As noted above, the value of $\bar{r}_3$ used in Eq. (3) is required to be at 1.1-km resolution, and is calculated from a straight average of the available 275-m samples. In calculating $\bar{r}_3$, up to 16 values of high resolution data are potentially available. We include in the calculation only those values of r_3 that are associated with an RDQI $\leq$ RDQI$_2$, where RDQI$_2$ is a pre-specified threshold, and require at least n_2 out of the possible 16 values to satisfy this criterion. No weighting by RDQI is used in the calculation of $\bar{r}_3$. At present, we set RDQI$_2 = 0$, that is, only the best quality data are included in the calculation, and we set $n_2 = 9$. If less than n_2 values of r_3 are present, a value of D is not calculated. In this event, it is not possible to calculate either the primary or secondary test at this location.

If a valid value of D is calculated at a particular location, then a value of DSVI can be calculated provided there are sufficient values of D in the 3 x 3 matrix to calculate the mean, D_m. We require there to be a minimum of $n_3 = 5$ values of D out of the maximum possible of 9 in order to establish a valid DSVI. Missing surrounding values of D may result from lack of data with adequate RDQI's, or due to location at the swath edge.

3.3.4   Generate histograms

3.3.4.1   Physics of the problem

Histograms are generated for equal-length time blocks, each 16 days (or more specifically, 1 EOS cycle of 233 orbits) long. A 16-day interval is chosen to ensure enough new data to warrant the calculation of the new thresholds. An RC Histogram Dataset is generated for each 16-day block. In general, the upper and lower bounds of the histograms can depend on surface type and sun/view geometry. For each surface type, the cloud screening observables, Q, are calculated and stored in the RC Histogram Dataset in bins, partitioned as follows:

(1) There are 5 view angle bins, corresponding to the MISR camera view angles: Nadir, 26.1°, 45.6°, 60.0°, and 70.5°.

(2) There are 10 sun angle cosine (μ_0) bins: $0.0 \leq \mu_0 < 0.1$; $0.1 \leq \mu_0 < 0.2$; ...; $0.9 \leq \mu_0 \leq 1.0$.

(3) There are 12 relative azimuth ($\Delta\phi = \phi - \phi_0$, the azimuth difference between the view and illumination angles) bins: $0° \leq \Delta\phi < 15°$; $15° \leq \Delta\phi < 30°$; ...; $165° \leq \Delta\phi \leq 180°$. Symmetry about the principal plane is assumed, such a value of $\Delta\phi > 180°$ is replaced by $360° - \Delta\phi$.

(4) The histogram of each observable is broken up into 128 gray levels.

15

Surface type is established by referencing the Ancillary Geographic Product (AGP) and the Cloud Screening Surface Classification (CSSC) Dataset. If the AGP specifies that the surface is covered by ocean or inland water, the cloud screening observables for water are calculated and stored in the part of the RC Histogram Dataset corresponding to water surface. Otherwise, the cloud screening observables for land are calculated, the CSSC Dataset is used to identify the surface type, and the appropriate part of the RC Histogram Dataset is accessed.

Further information on the contents of the RC Histogram Dataset is provided in §3.5.1.

3.3.5   Update thresholds

3.3.5.1   Physics of the problem

For any single sun-view geometry, r_4 and σ_3 will have fairly constant values over all deep water covering the globe. Thus, the ocean thresholds for deep water can be established early on in the mission without a continuing update procedure. Also, nominal pre-launch thresholds can be established with a high degree of confidence using existing data sets. For shallow waters, where bottom upwelling radiance becomes important, the static thresholds will be supplemented by dynamic thresholds that are continually updated using the procedure outlined in this section. The need for this arises from the seasonal variability in water depth and/or basin vegetation content.

Because of the variability of the Earth's surface, particularly over land, it is desirable to have the ability to generate thresholds dynamically for certain sections of the cloud detection algorithm. This can be done in automated fashion at the DAAC using the histogramming approach described earlier, as applied to accumulated MISR observations over a period of time. Analyses of the histograms will be performed to insure that they contain sufficient information to derive updated thresholds.

Thresholds are a function of: View angle, sun angle, relative view-sun azimuth angle, time, and place on globe (or surface type). The time dependence can be of 3 kinds: (a) time independent (static), (b) dynamic (seasonal), and (c) dynamic on a less than seasonal time scale. The latter is approximately monthly, on a sliding window (see below). Each observable is compared to three thresholds: T1 divides CloudHC from CloudLC, T2 divides CloudLC from ClearLC, and T3 divides ClearLC from ClearHC. T1 and T3 are either static or dynamic on seasonal time scales. T2 can have any of the three time dependencies. The following table shows the time dependencies of the thresholds for the different observables used in generation of the RCCM. It is possible that the shallow water thresholds may become static at a later time. In this event, the seasonal entries will be replaced by temporally invariant thresholds.

Table 4: Time dependencies of thresholds in the RC Threshold Dataset

Observable	Threshold 1	Threshold 2	Threshold 3
r_4	Deep water: Static Shallow water: Seasonal	Deep water: Static Shallow water: Seasonal	Deep water: Static Shallow water: Seasonal
σ_3	Deep water: Static Shallow water: Seasonal	Deep water: Static Shallow water: Seasonal	Deep water: Static Shallow water: Seasonal
D	Seasonal	Pre-launch: T2 = (T1 + T3)/2 Post-launch: Monthly, sliding	Seasonal
DSVI	Seasonal	Pre-launch: T2 = (T1 + T3)/2 Post-launch: Monthly, sliding	Seasonal

The time dependencies are defined as follows:

(1) Static means that the same threshold is used throughout the mission.

(2) Seasonal means that a separate threshold is defined for Winter, Spring, Summer, and Fall (i.e., Jan-Feb-Mar, Apr-May-Jun, Jul-Aug-Sep, Oct-Nov-Dec). As described in the previous section, histograms are generated for each 16-day block. Each season is considered to contain all 16-day blocks for which at least 7 days of the block fall within the season. (It is therefore possible for a 16-day block to be contained within two seasons.) Histogram analysis is performed at the end of each season on the composite histogram generated from the blocks that are contained in the season to derive new values for the thresholds identified as seasonal in Table 4. Current data in a particular season uses the threshold derived from the previous year's data in the same season.

(3) Monthly, sliding works in this manner: For processing on a given date, the histograms generated for the two previous and completed 16-day blocks are composited together and analysis is performed to generate a new threshold T2 for the observables D and DSVI (see Table 4). This value is used in processing current data. The use of a 1 month window, coupled with the areal extent of surface classes within the CSSC data set (§3.5.3), provides sufficient sampling of data to ensure a proper statistical summary for threshold selection. The 1 month window slides in order to capture the seasonal changes in surface and atmospheric properties.

Further information on the contents of the RC Threshold Dataset is provided in §3.5.2.

3.3.5.2 Mathematical description of the algorithm

There are numerous automated threshold selection procedures that are based only on the frequency distribution of an observable Q. Several comparative studies have been performed in order to decide which is the best procedure (e.g., *[10]*, *[16]*). These studies have shown the results to be

data-dependent. For the purpose of cloud detection from satellite data, preliminary results *[5]* show that the method proposed by Otsu *[13]* (described in Rev. A of this document) is very conservative, producing thresholds that overestimate cloud amount by about 10% compared to the "truth". Consequently, it is replaced by the current method, that of Li and Lee *[11]*, which performs consistently well for a wide range of AVHRR derived histograms in both the space and time domain and is nearly unbiased. A brief description of the method is given below.

Let a histogram of Q consist of B bins (nominally, B = 128). Let the frequency of observation within a bin be denoted by n_i, $i \in \{1,..., B\}$. The histogram is normalized and can by viewed as a probability mass function, that is

$$f_i = \frac{n_i}{\sum\limits_{i=1}^{B} n_i} \tag{4}$$

Let the threshold that divides observations into clear and cloudy categories be set at bin T2, with T2 $\in \{1, ..., B\}$, such that all values of Q $\leq$ T2 defines one category, C_1, all values of Q $>$ T2 defines the other category, C_2. The probabilities of class occurrence are given by

$$p_1 = Pr(C_1) = \sum\limits_{i=1}^{T2} f_i = p(T2) \tag{5}$$

$$p_2 = Pr(C_2) = \sum\limits_{i=T2+1}^{B} f_i = 1-p(T2) \tag{6}$$

Also, we define the following quantities:

$$m_1(T2) = \sum\limits_{i=1}^{T2} i \cdot f_i \tag{7}$$

$$m_2(T2) = \sum\limits_{i=T2+1}^{B} i \cdot f_i \tag{8}$$

$$J(T2) = \sum\limits_{i=1}^{T2} i \cdot f_i \cdot \ln\left\{\frac{i \cdot p(T2)}{m_1(T2)}\right\} + \sum\limits_{i=T2+1}^{B} i \cdot f_i \cdot \ln\left\{\frac{i \cdot [1-p(T2)]}{m_2(T2)}\right\} \tag{9}$$

Then, the threshold that minimizes *J* is selected as the optimum threshold for T2.

$$\text{Equation deleted} \tag{10}$$

$$\text{Equation deleted} \tag{11}$$

The approach is founded on the assumption that well-thresholded categories are separated in Q-values and that a threshold that best separates the categories in Q-values would be the best threshold. This algorithm is a robust method of selecting the threshold T2, and does not require the histogram to be bimodal.

T1 and T3 are chosen based on seasonal or longer term statistics. Over such time scales, the histogram is expected to be bimodal. The two modal peaks define T1(peak) and T3(peak), i.e., by selecting the histogram bins with the maximum frequencies of occurrence on either side of T2. T1 and T3 are calculated by first finding T1(peak) and T3(peak), the histogram maxima on the respective sides of T2. If multiple bins on either side of T2 have the same frequency, the values of T1(peak) and T3(peak) closest to T2 are selected. We then set

$$T1 \;=\; T1(\text{peak}) + a\sigma_1 \tag{11a}$$

$$T3 \;=\; T3(\text{peak}) + b\sigma_3 \tag{11b}$$

where σ_1 is the standard deviation of the data on the cloudy side of T2 and σ_3 is the standard deviation of the data on the clear side of T2, and a and b are scene-dependent parameters with $a \geq 0$ and $b \leq 0$.

3.3.6   Generate RCCM

3.3.6.1   Physics of the problem

The parameters r_4 and σ_3 are the primary and secondary observables for water, and D and DSVI are the primary and secondary observables for land. Each primary and secondary observable, in conjunction with the thresholds obtained from the latest version of the RC Threshold Dataset, classifies a 1.1-km area into the CloudHC, CloudLC, ClearLC, and ClearHC designators. These are then combined together to form the final cloud mask for each camera.

3.3.6.2   Mathematical description of the algorithm

The primary and secondary masks are combined together to form the final cloud mask according to the following table. The designation No Retrieval is included to cover situations where a channel is inoperable, data are missing, or the Radiometric Data Quality Indicator has a value > $RDQI_0$, $RDQI_1$, or $RDQI_2$ (as appropriate) for some other reason.

19

Table 5: Logic for combining primary and secondary cloud masks

		Primary Cloud Mask (r_4 or D)				
		No Retrieval	CloudHC	CloudLC	ClearLC	ClearHC
Secondary Cloud Mask (σ_3 or DSVI)	No Retrieval	No Retrieval	CloudHC	CloudLC	ClearLC	ClearHC
	CloudHC	CloudHC	CloudHC	CloudHC	CloudHC	ClearHC
	CloudLC	CloudLC	CloudHC	CloudLC	CloudLC	ClearHC
	ClearLC	ClearLC	CloudHC	CloudLC	ClearLC	ClearHC
	ClearHC	ClearHC	CloudHC	ClearHC	ClearHC	ClearHC

3.4   PRACTICAL CONSIDERATIONS

3.4.1   Numerical computation considerations

Requirements on processing speed and data storage are described in *[M-17]*.

3.4.2   Programming and procedural considerations

Guidelines to be followed during algorithm development are described in *[M-13]*.

3.4.3   Configuration of retrievals

A Cloud Detection Configuration File is used to establish the numerical values of adjustable parameters used within the retrievals. The purpose of establishing a separate file is to avoid "hard-wiring" specific values into the software. The Georectified Radiance Product will contain information indicating what version of the configuration file was used. The contents of the Cloud Detection Configuration File are shown in Table 6. The values shown correspond to the at-launch settings. The column entitled "Section" indicates where in this ATB a description of the specific configuration parameter is found.

Table 6: Contents of the Cloud Detection Configuration File

Parameter	Value	Section
Glitter exclusion angle ξ_c	30°	3.3.1.2
Land classification search window half-width, N_{search}	20	3.3.3.2.2
Maximum value of RDQI allowable for r_4 (RDQI$_0$)	0	3.3.3.2.2
Maximum value of RDQI allowable for calculating σ_3 and $\bar{r}_3$ over ocean (RDQI$_1$)	0	3.3.3.2.2

Table 6: Contents of the Cloud Detection Configuration File (continued)

Parameter	Value	Section
Minimum number of r_3 values required for calculating σ_3 and $\bar{r}_3$ over ocean (n_1)	9	3.3.3.2.2
Maximum value of RDQI allowable for calculating $\bar{r}_3$ over land (RDQI$_2$)	0	3.3.3.2.3
Minimum number of r_3 values required for calculating $\bar{r}_3$ over land (n_2)	9	3.3.3.2.3
Minimum number of D values required for calculating DSVI (n_3)	5	3.3.3.2.3
Minimum number of days for considering a 16-day block part of a season	7	3.3.5.1
Logic for combining primary and secondary cloud masks	Table 5	3.3.6.2

3.4.4   Quality assessment and diagnostics

Several parameters will be reviewed by the MISR team for quality assessment purposes at the SCF. These include comparisons of MODIS-derived cloud masks with MISR-derived masks. In addition, a quality assessment parameter is associated with the RCCM. This parameter can take on one of four values:

(1) No retrieval. This could occur if data are missing.

(2) Secondary test used only.

(3) Primary test used only.

(4) Both primary and secondary tests used.

3.5   ANCILLARY DATASETS

3.5.1   RC Histogram Dataset

The RC Histogram is generated routinely at the DAAC during standard processing. The contents are shown in Table 1.

Table 7: Parameters within the RC Histogram Dataset

Parameter name	Units	Comments
Surface identifier	none	• Includes water and a variety of land surface types corresponding to the contents of the CSSC Dataset
Observable identifier	none	• Includes land and water primary and secondary cloud vs. clear discriminators
View angle	deg.	• For each of the nominal MISR camera view angles (5 values)
Sun angle cosine	none	• Divided into 10 bins

Table 7: Parameters within the RC Histogram Dataset (continued)

Parameter name	Units	Comments
View-illumination azimuth difference	deg.	• Divided into 12 bins
Time interval of applicability	calendar date	• Divided into 16-day duration bins
Frequency of occurrence of observable	none	• Divided into 128 gray-level bins in the histogram

3.5.2 RC Threshold Dataset

A pre-launch version of the RC Threshold Dataset will be used during initial processing. It will consist of nominal thresholds. Once automated histogram generation has been initiated at the DAAC, the RC Threshold Dataset will be updated on a regular basis. The contents are described in Table 1.

Table 8: Parameters within the RC Threshold Dataset

Parameter name	Units	Comments
Surface identifier	none	• Includes water and a variety of land surface types corresponding to the contents of the CSSC Dataset
Observable identifier	none	• Includes land and water primary and secondary cloud vs. clear discriminators
View angle	deg.	• For each of the nominal MISR camera view angles (5 values)
Sun angle cosine	none	• Divided into 10 bins
View-illumination azimuth difference	deg.	• Divided into 12 bins
Time interval of applicability	calendar date	• Divided into 16-day duration bins
Threshold T1	none	• Divides CloudHC from CloudLC
Time dependency indicator of T1	none	• Static or seasonal
Threshold T2	none	• Divides CloudLC from ClearLC
Time dependency indicator of T2	none	• Static, seasonal, or monthly
Threshold T3	none	• Divides ClearLC from ClearHC
Time dependency indicator of T3	none	• Static or seasonal

3.5.3   Cloud Screening Surface Classification Dataset

A pre-launch version of the CSSC Dataset has been derived from the WE1.4D version of Olson's global ecosystem database *[12]*. Version WE1.4D contains 59 ecosystem classes mapped onto a 10-arcmin grid over the globe. Each unique surface cover type within the CSSC Dataset is derived based on the temporal and spatial properties of the surface being relatively uniform over the grid. This allows for accurate threshold determination since the variance in the properties within a surface cover type is reduced. Thus, the 59 ecosystem classes are further subdivided into a larger number of surface cover types. The CSSC Dataset is derived based on the following requirements:

(1) Each unique surface cover type must be greater than 12100 km^2 in area (about 36 10-arcmin grid points about the equator);

(2) A surface cover type is made up of the same ecosystem class that is 8-connected (defined for a square-tessellation to be a region that is connected to at least one of the eight surrounding regions);

(3) 8-connected regions of the same ecosystem class that are less than 12100 km^2 in area are grouped into its same-class nearest-neighbor that is within a $\pm5°$ latitude band.

Note that the third requirement allows a surface cover type within the CSSC Dataset to be non-contiguous. With these requirements, 1580 surface cover types make up the pre-launch CSSC Dataset. Each of these 1580 surface types has an associated indicator specifying whether the surface is classified as vegetated or non-vegetated. Each surface type is also associated with a parameter value that is used as part of cloud screening over land. At present, there are only two values of this parameter in the CSSC, depending on whether the surface is vegetated or non-vegetated. The vegetated/non-vegetated designation is also used in Level 2 processing *[M-9]*. The CSSC Dataset will be updated post launch using data derived from MISR and MODIS.

3.6   CLOUD DETECTION PERFORMANCE

In this section, we present some examples of the RCCM algorithm as applied to AVHRR and Landsat data. The results should not be interpreted as validation of the RCCM algorithm, but rather as a guide to showing its functionality. True validation of this algorithm requires actual MISR data, especially when considering threshold development. Validation also needs to be performed globally and over a sufficiently long time scale to permit the algorithm to be tested over a rich variety of surface and atmospheric conditions.The examples presented below, along with hundreds of other scenes tested but not shown provide us with confidence in the method used to generate the RCCM.

Three scenes are considered: ocean, land covered with vegetation, and land covered with desert. (The RCCM is expected to perform poorly over snow/ice covered scenes. For this reason

the SDCM and ASCM, which are generated at Level 2, are used without the RCCM to summarize the state of cloudiness over snow/ice surfaces *[M-8]*.) The approach taken in generating the land examples does not include generating the histograms using temporal data (§3.3.4). This is presently under development using AVHRR data. Instead, the histogram is generated based only on the data within the scene. This requires the scene to be partially cloudy in order to get a statistical representation of both the clear and cloudy parts of the scene. Only T2 is generated from the histogram. T1 and T3, which are static within a season (see Table 4), cannot be determined from the histogram as long-term statistics are required. Instead, they are chosen based on our experience in analyzing many AVHHR scenes.

3.6.1   Water

Over water, the RCCM makes use of data at 1.1 km resolution for the 866-nm band and 275 m resolution for the 672-nm band. The only appropriate data available to us was Landsat data. We use only the Landsat 830-nm band for both the primary and secondary observables at a degraded resolution of 912 m and 228 m, respectively, with the assumption that the standard deviation of the 866-nm and 672-nm reflectances are similar. The Landsat scene is shown below is 58.4 km per side. The solar zenith angle is 2° and the view is nadir. Figures 4 - 7 show the reflectance at 114 m resolution, the 11 μm brightness temperature at 114 m resolution, the reflectance at 912 m resolution, and the standard deviation of the reflectance at 912 m resolution. Note that the scene is characterized by cumulus clouds of a wide variety of sizes.

The thresholds are static (see Table 4), and for this sun view geometry are equal to $T1(r_4) = 0.056$, $T2(r_4) = 0.036$, $T3(r_4) = 0.031$, $T1(\sigma_3) = 0.0040$, $T2(\sigma_3) = 0.0025$, and $T3(\sigma_3) = 0.0012$. $T1(r_4)$ is consistent with the conservative thresholds derived in *[4]*. With these thresholds the application of the RCCM algorithm produced the mask shown in Figure 8. By carefully comparing the mask with the reflectance and temperature fields, we see that the algorithm performs very well. In this particular example, all pixels have been classified with high confidence, demonstrating the usefulness of the secondary observable in detecting subpixel cloudiness.
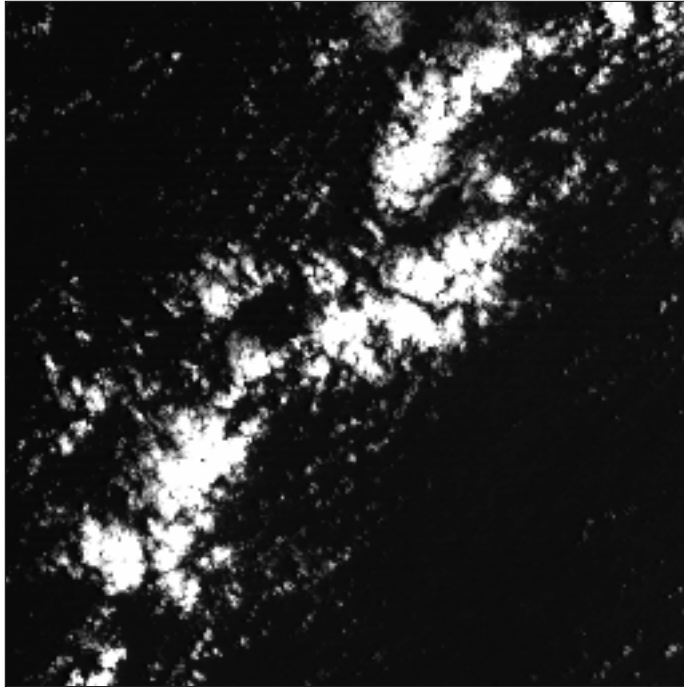
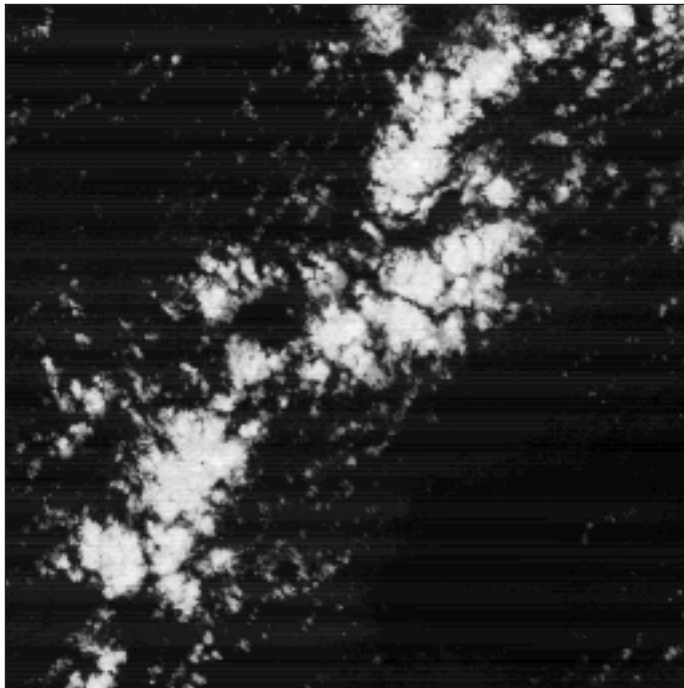**Figure 4. 860 nm reflectance of the ocean at a resolution of 114 m**

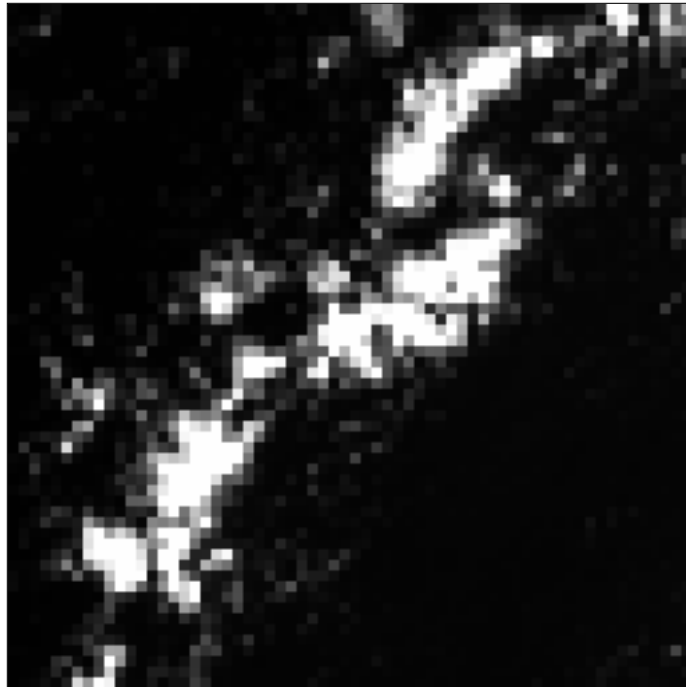**Figure 5. 11 μm brightness temperature of the ocean scene at a resolution of 114 m**

Figure 6. 830 nm reflectance for the ocean scene at a resolution of 912 m
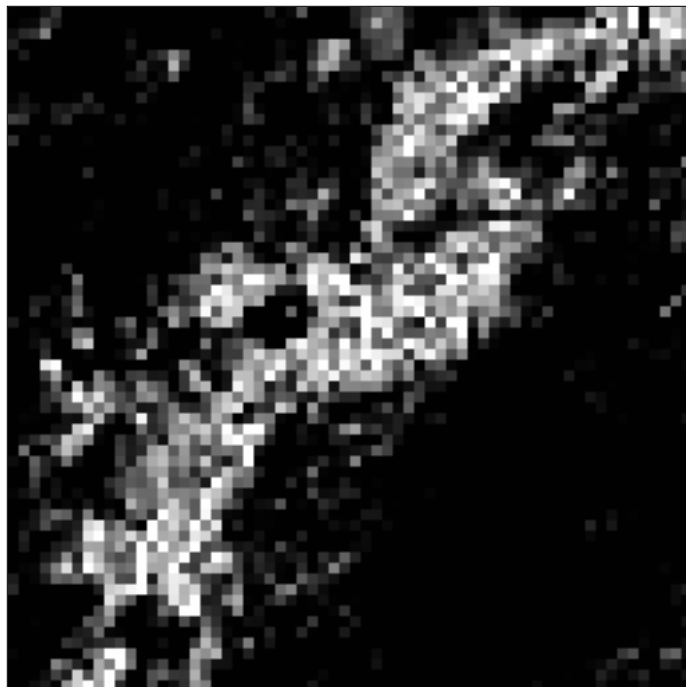

Figure 7. Standard deviation of the 830 nm reflectance for the ocean scene

26

Figure 8. Cloud mask for the ocean scene

3.6.2 Land

3.6.2.1 Vegetated scene

The AVHRR LAC (1.1 km sub-satellite ground-track resolution) scene shown below is 256 x 256 pixels, centered at (2°N, 68.5°W), which makes up part of the Brazilian rain forest. It was taken on August 26, 1993. At the center of the scene the solar zenith angle is 55°, the view angle is 39°, and the relative azimuth angle between sun and view is 21°. Figures 9 - 12 show the 670-nm reflectance, the 11 μm brightness temperature, the D field and the DSVI field. Note that the scene is characterized by low level cumulus clouds, deep convective plumes and some cirrus clouds.

The histograms for D and DSVI are shown in Figures 13 and 14, respectively. The automated threshold selection procedure (§3.3.5.2) applied to these histograms has chosen threshold values of T2(D) = 82 and T2(DSVI) = 34. We have chosen for the seasonal thresholds T1(D) = 15, T3(D) = 120, T1(DSVI) = 25, and T3(DSVI) = 65. With these thresholds the application of the RCCM algorithm produced the mask shown in Figure 15. The mask contains 4 gray levels: white = Cloud-HC, light gray = CloudLC, dark gray = ClearLC, and black = ClearHC. By carefully comparing the mask with Figures 9 and 10, we see that the algorithm performs very well. Note that the low confidence clear/cloud pixels tend to exist near the edges of the CloudHC pixels, indicating that pixels at cloud boundaries are only partially filled or that the clouds simply get thinner at the cloud "edges."
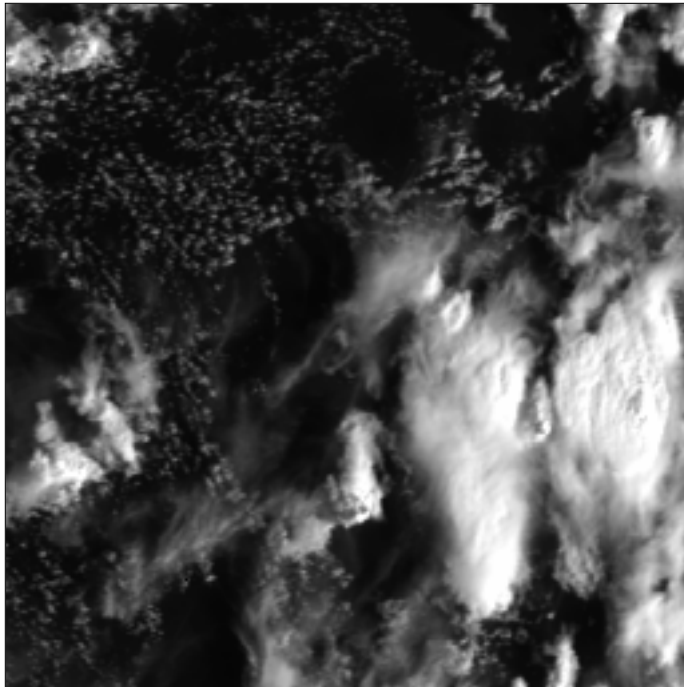
Figure 9. 670 nm reflectance field for the vegetated scene
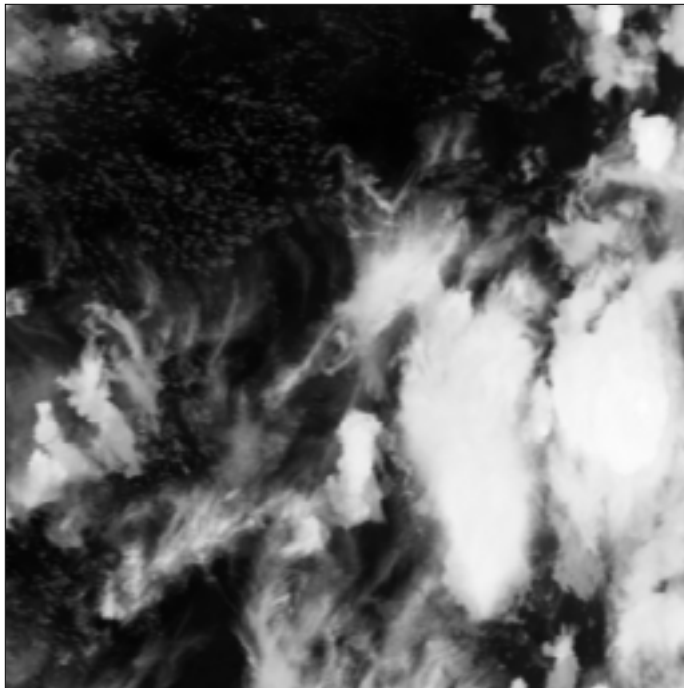
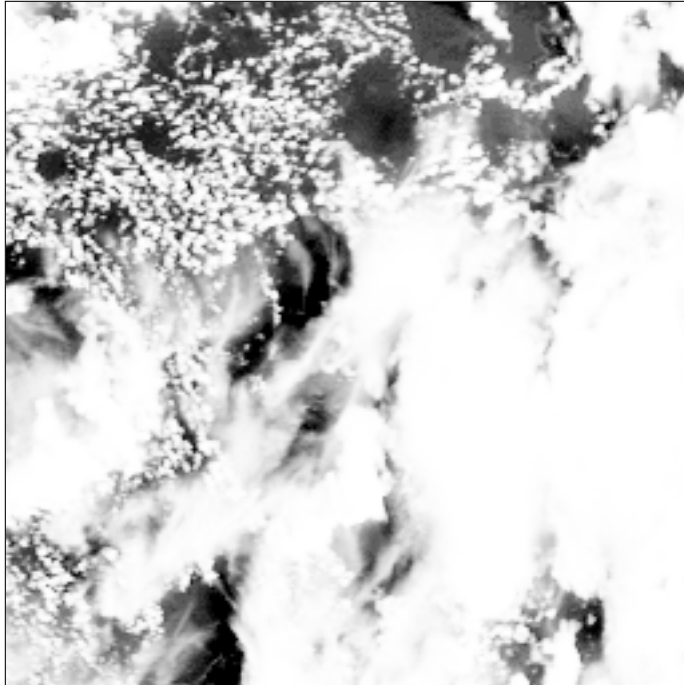**Figure 10. 11 μm brightness temperature field for the vegetated scene**
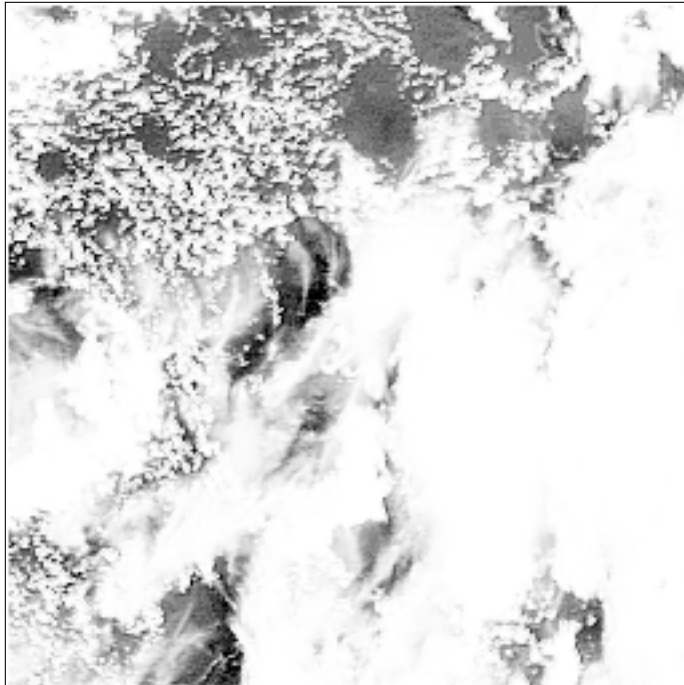
Figure 11. D field for the vegetated scene

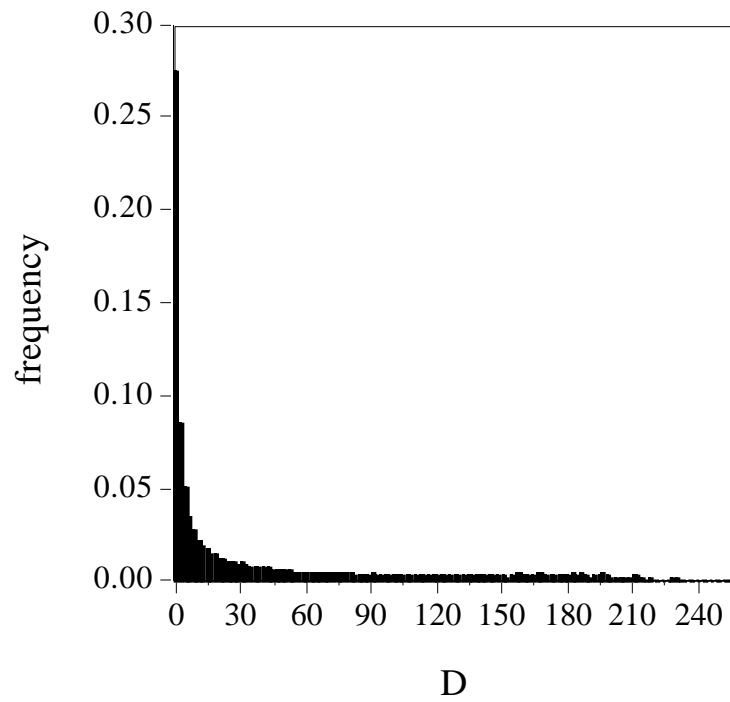**Figure 12. DSVI field for the vegetated scene**

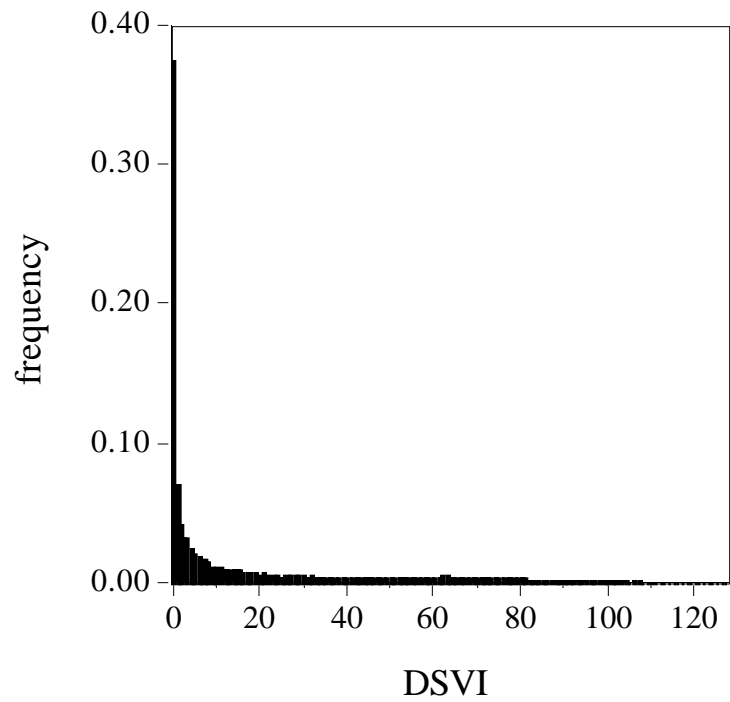**Figure 13. D histogram for the vegetated scene.**
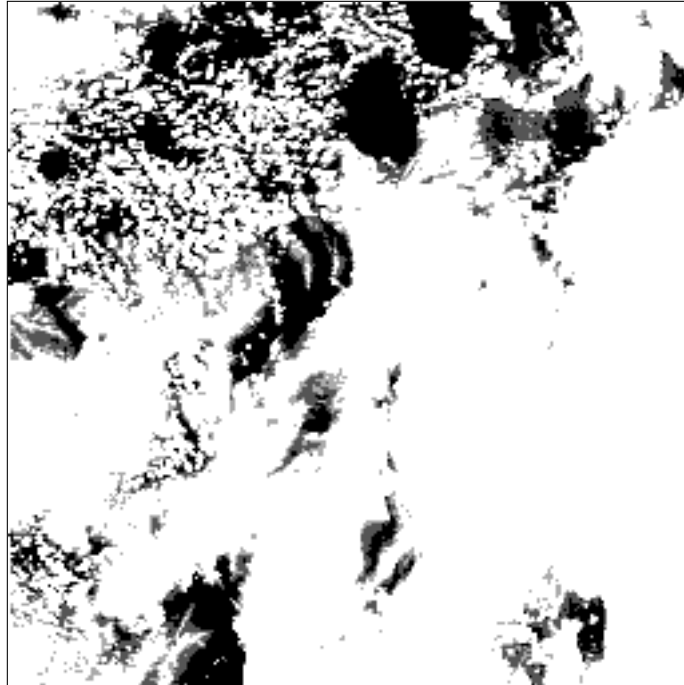

Figure 14. DSVI histogram for the vegetated scene

Figure 15. Cloud mask for the vegetated scene

3.6.2.2   Desert scene

The AVHRR LAC scene used is 256 x 256 pixels, centered at (16°N, 11°W), belonging to the desert region in south-central Mauritania. It was taken on September 26, 1993. At the center of the scene the solar zenith angle is 71.5°, the view angle is 54°, and the relative azimuth angle between sun and view is 179°. Figures 16 - 19 show the 670-nm reflectance, the 11 μm brightness temperature, the *D* and the DSVI fields. The scene is characterized by a wide range of cloud types.

The histograms for *D* and DSVI are shown in Figures 20 and 21, respectively. The automated threshold selection procedure applied to these histograms has chosen a threshold value of T2(*D*) = 2.6 and T2(DSVI) = 1.05. We have chosen for the seasonal thresholds T1(*D*) = 1.5, T3(*D*) = 4.5, T1(DSVI) = 0.5, and T3(DSVI) = 1.5. With these thresholds the application of the RCCM algorithm produced the mask shown in Figure 22. As in the vegetated scene, the low confidence clear/cloud pixels tend to exist near the edges of the CloudHC pixels. By carefully comparing the mask with Figures 16 and 17, we see that the algorithm performs very well. However, it appears the algorithm has slightly overestimated the number of ClearHC pixels in the top-right corner of the scene. Upon closer examination of the terrain within the scene, the scene is not entirely desert; it ranges from barren desert in the bottom-left corner to semi-desert/shrubs in the top-right corner. This emphasizes the need to accurately partition the histograms by land type (as is done with the CSSC Dataset). No partitioning was done in this example in order to demonstrate the errors that can be incurred.
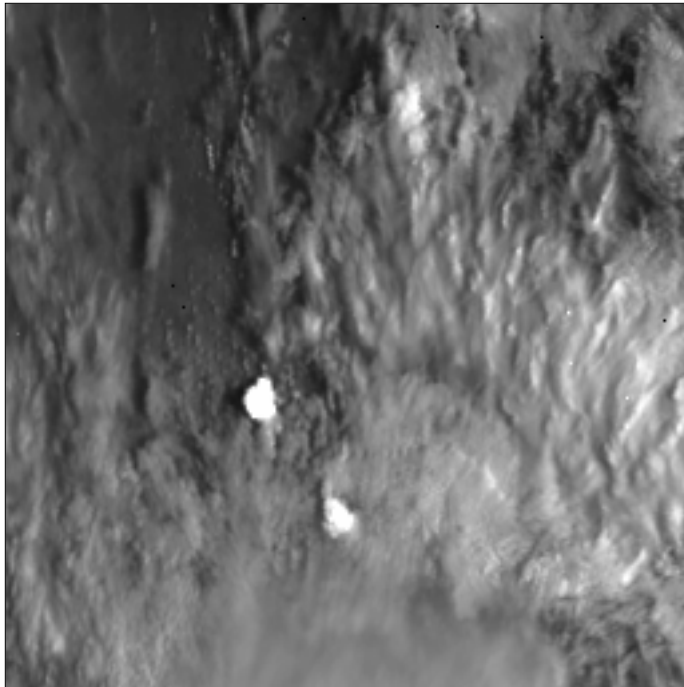
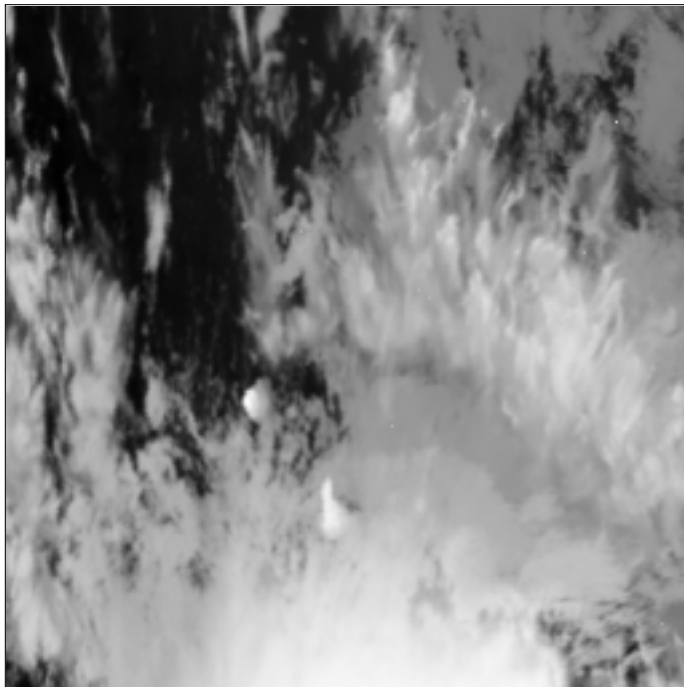**Figure 16. 670 nm reflectance field of the desert scene**

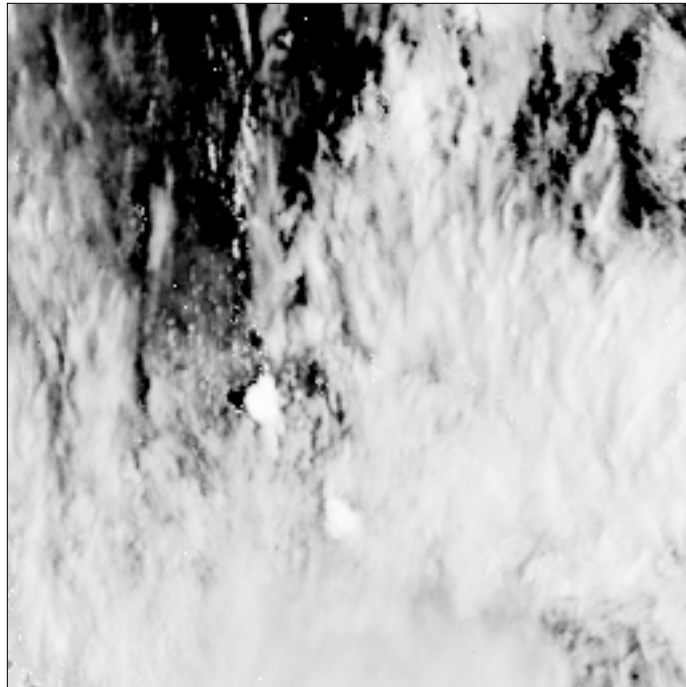**Figure 17. 11 μm brightness temperature of the desert scene**
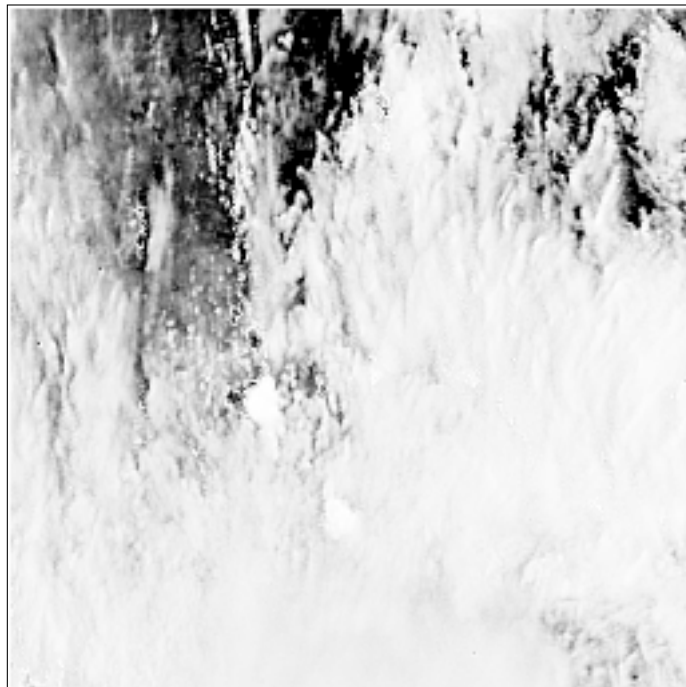
Figure 18. D field for the desert scene
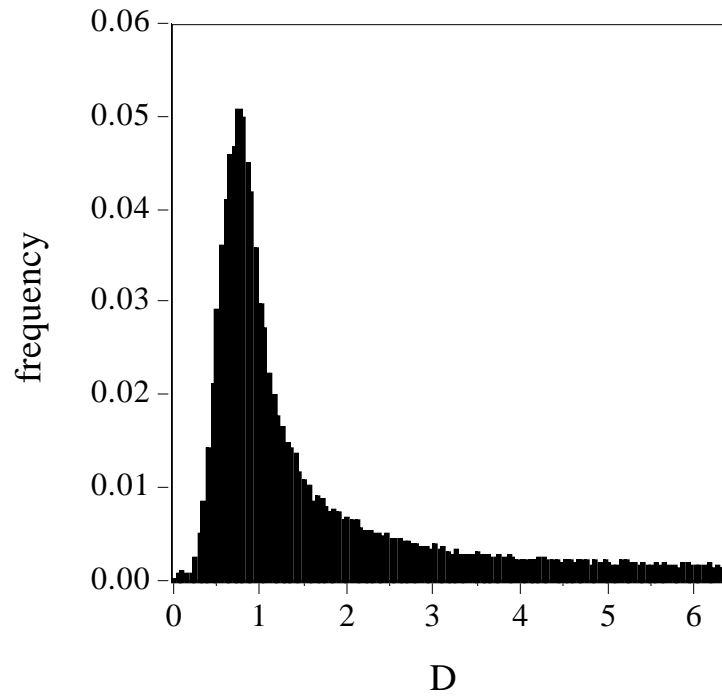

Figure 19. DSVI field for the desert scene

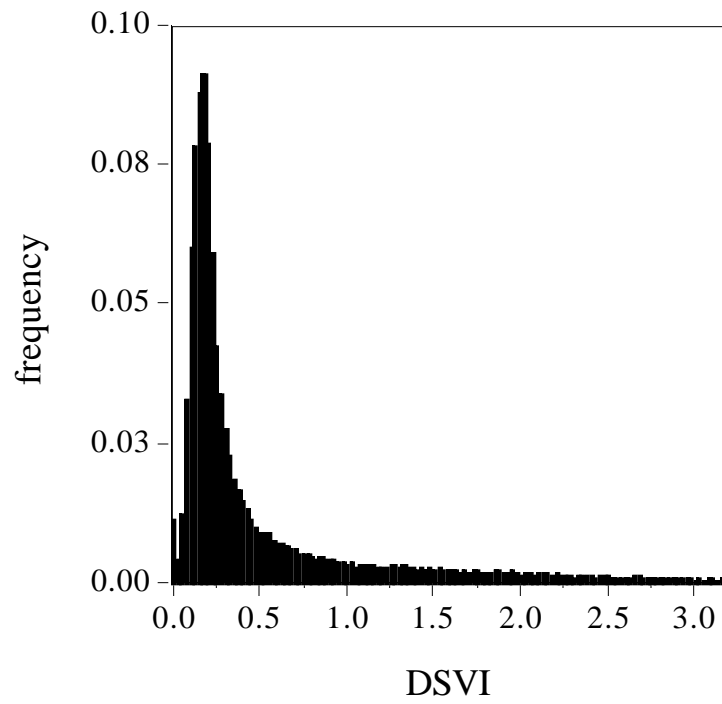**Figure 20. D histogram for the desert scene**

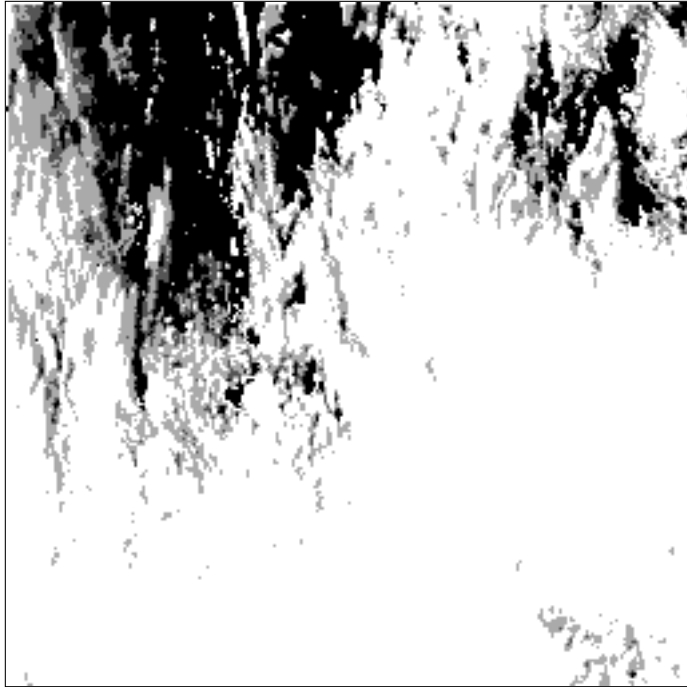**Figure 21. DSVI histogram for the desert scene**

Figure 22. Cloud mask for the desert scene

4. ASSUMPTIONS AND LIMITATIONS

4.1 ASSUMPTIONS

The following assumptions are made with respect to the algorithms described in this document:

(1) A cone half-angle of 30° is sufficient to insure that views are free of sun glitter contamination.

(2) MODIS Level 2 scene identification and classification masks will be available for intercomparisons and quality checks of MISR data.

4.2 LIMITATIONS

The following limitations apply to the algorithms described in this document:

(1) The RC Threshold Dataset will initially contain nominal values. These will be re-evaluated within the first few months of the Terra mission. After this, the RC Threshold Dataset will continually be modified based on automated update procedures.

(2) Rapid changes in surface reflectivities, e.g., due to agricultural changes, local flooding, etc. over land will temporarily affect the appropriate thresholds for cloud detection.

5. REFERENCES

[1] Brest, C.L. and W.B. Rossow (1992). Radiometric calibration and monitoring of NOAA AVHRR data for ISCCP. *Int. J. Rem. Sens.* **13**, 235-273.

[2] Coakley, J. A. and F. P. Bretherton (1982). Cloud cover from high-resolution scanner data: Detecting and allowing for partially filled fields of view. *J. Geophys. Res.* **87**, 4917.

[3] Di Girolamo, L. and R. Davies (1994a). Optimizing the use of 0.67 μm and 0.86 μm radiometric data for cloud detection. *Proc. 8th Conf. Atmos. Radiation* (AMS), Jan. 23-28, Nashville, TN, 466.

[4] Di Girolamo, L. and R. Davies (1994b). The image navigation cloud mask for the Multi-angle Imaging SpectroRadiometer (MISR). *J. Atmos. Oceanic Tech.* **12***, 1215-1228.*

[5] Di Girolamo, L. (1998). A comparison of 15 global, non-parametric, automated threshold selection procedures for cloud detection. *Proc. 9th Conf. Sat. Meteo. Ocean.,* May 24-29, Paris, France, 197-200.

[6] Ebert, E.E. (1987). A pattern recognition technique for distinguishing surface and cloud types in the polar regions. *J. Clim. Appl. Met.* **26**, 1412.

[7] Gallaudet, T.C. and J.J. Simpson (1991). Automated cloud screening of AVHRR imagery using split-and-merge clustering. *Remote Sens. Environ.* **38**, 77

[8] Goodman, A.H. and A. Henderson-Sellers (1988). Cloud detection analysis: A review of recent progress. *Atm. Res.* **21**, 203.

[9] Gustafson, G.B., R.G. Isaacs, R.P. d'Entremont, J.M. Sparrow, T.M. Hamill, C. Grassotti, D.W. Johnson, C.P. Sarkisian, D.C. Peduzzi, B.T. Pearson, V.D. Jakabhazy, J.S. Belfiore, A.S. Lisa (1994). Support of Environmental Requirements for Cloud Analysis and Archive (SERCAA): algorithm descriptions. **PL-TR-94-2114**, Phillips Laboratory, AFMC, Hanscom AFB, MA.

[10] Lee, S.U., S.Y. Chung, and R.H. Park (1990). A comparative performance study of several global thresholding techniques for segmentation. *Comput. Vision Graphics Image Process.* **52**, 171-190.

[11] Li, C.H. and C.K. Lee (1993). Minimum cross entropy thresholding. *Pattern Recogn.* **26**, 617-625.

[12] NOAA-EPA Global Ecosystems Database Project, 1992: *Global Ecosystems Database*

Version 1.0 User's Guide, Documentation, Reprints, and Digital Data on CD-ROM. U.S. DOC/ NOAA National Geophysical Data Center, Boulder, CO, 36 pp.

[13] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst., Man, Cybern.* **SMC-9**, 62.

[14] Rossow, W.B. (1989). Measuring cloud properties from space: A review. *J. Clim.* **2**, 201.

[15] Rossow, W.B., F. Mosher, E. Kinsella, A. Arking, M. Desbois, E. Harrison, P. Minnis, E. Ruprecht, G. Seze, C. Simmer, and E. Smit (1985). ISCCP cloud algorithm intercomparison. *J. Clim. Appl. Met.* **24**, 877.

[16] Sahoo, P.K., S. Soltani, A.K.C. Wong, and Y.C. Chen (1988). A survey of thresholding techniques. *Comput. Vision Graphics Image Process.* **41**, 233-260.

[17] Tovinkere, V.R., M. Penaloza, A. Logar, J. Lee, R.C. Weger, T.A. Berendes, and R.M. Welch (1993). An intercomparison of artificial intelligence approaches for polar scene identification. *J. Geophys. Res.* **98**, 5001.