NIES-GOSAT2-ALG-20191008-009-00

GOSAT-2 TANSO-CAI-2 L2 Cloud Discrimination Processing Algorithm Theoretical Basis Document

May 2020

National Institute for Environmental Studies
GOSAT-2 Project

OISHI Yu[1)*]
ISHIDA Haruma[2)]
NAKAJIMA Takashi Y.[1)]

1) Tokai University Research & Information Center
2) Climate Research Department, Meteorological Research Institute
* Currently at: Research Center for Agricultural Information Technology, National Agriculture and Food Research Organization

Table of Contents

1 Introduction

Greenhouse gases Observing SATellite (GOSAT) “IBUKI” is a Japanese satellite launched on January 23, 2009 with the aim of observing the global distribution of atmospheric concentrations of CO_2 and CH_4. Greenhouse gases Observing SATellite-2 (GOSAT-2) “IBUKI-2”, which was launched on October 29, 2018 as the successor to GOSAT, carries two sensors: Thermal And Near-infrared Sensor for carbon Observation-Fourier Transform Spectrometer 2 (TANSO-FTS-2) and TANSO-Cloud and Aerosol Imager 2 (TANSO-CAI-2). GOSAT-2 estimates the concentration of greenhouse gases using FTS-2; however, one of the factors that reduces the accuracy is the existence of clouds. As clouds in the field of view of FTS-2 could cause the incorrect estimation of the concentration, it is necessary to discriminate the presence of clouds. Therefore, one of the most important roles of CAI-2 is cloud discrimination. An algorithm based on radiance values acquired in more than a dozen bands from visible (VIS) to thermal infrared (TIR) regions is generally used for cloud discrimination by multi-spectral image sensors. CAI-2, however, has only five bands from near ultraviolet (NUV) to short wavelength infrared (SWIR) regions in each of forward and backward viewing, which limits the accuracy of cloud discrimination. This document describes the CAI-2 Level 2 (L2) cloud discrimination processing algorithm [Ishida et al. 2009, 2018], which maximally utilizes the small number of bands to discriminate clouds.

1.1 Scope

Figure 1.1–1 shows the processing flow of the whole system. The part shown in yellow in the figure is the scope of this Algorithm Theoretical Basis Document.

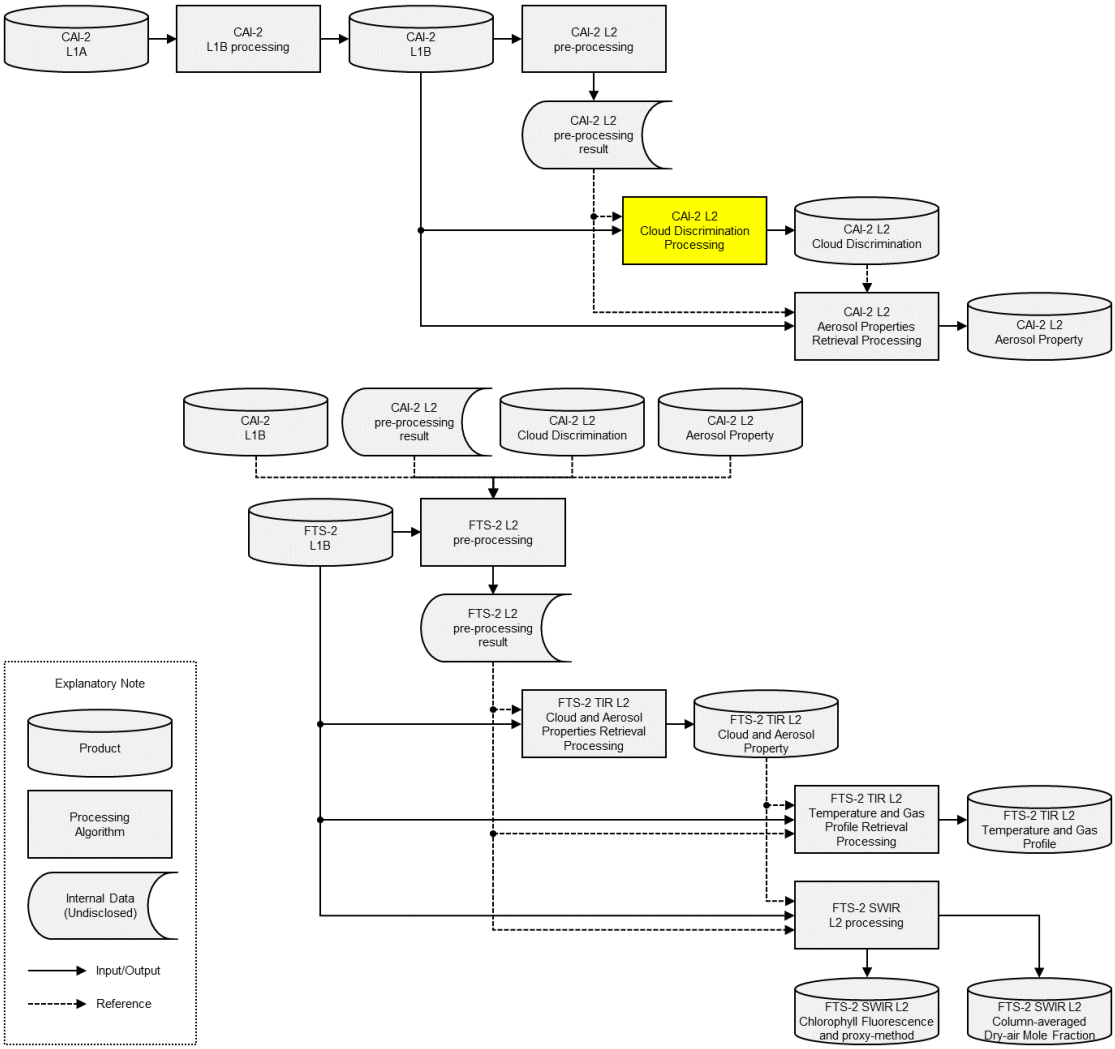


Figure 1.1–1. Scope of this Algorithm Theoretical Basis Document in the processing flow of the whole system.

1.2 Related documents

CAI-2 L2 cloud discrimination processing uses CAI-2 L2 pre-processing results and reflectance before atmospheric correction*, which are pre-processed from CAI-2 Level 1B (L1B) products. Therefore, the following documents should be referred:

1) GOSAT-2 TANSO-CAI-2 L1B Processing Algorithm Theoretical Basis Document
2) GOSAT-2 TANSO-CAI-2 L2 Pre-processing Algorithm Theoretical Basis Document.

CAI-2 L2 cloud discrimination product, which is the result of CAI-2 L2 cloud discrimination processing, is used for CAI-2 L2 aerosol properties retrieval processing and FTS-2 L2 pre-processing. To find out how CAI-2 L2 cloud discrimination product is used in GOSAT-2 operational processing, the following documents should be also referred:

3) GOSAT-2 TANSO-CAI-2 L2 Aerosol Properties Retrieval Processing Algorithm Theoretical Basis Document
4) GOSAT-2 TANSO-FTS-2 L2 Pre-processing Algorithm Theoretical Basis Document.

* Reflectance before atmospheric correction: Data created by extracting the pixels of the minimum apparent reflectance from 11 recurrences of CAI-2 L1B data (spectral radiance) to obtain the minimum reflectance and correcting the cloud shadow.

2 Background

The CAI-2 L2 cloud discrimination processing algorithm has the same design as the GOSAT CAI L2 cloud flag processing algorithm and has been developed based on existing research on global cloud masks from satellite observation data. This chapter explains existing research on cloud masks. The cloud detection algorithm of the International Satellite Cloud Climatology Project (ISCCP) creates clear-sky composite images for both visible (0.6 μm) and infrared (11 μm) and compares the radiance of the clear-sky composite image with the radiance of visible and infrared of each scene to detect clouds [Rossow and Garder, 1993]. The AVHRR Processing scheme Over cLouds, Land and Ocean (APOLLO) creates cloud masks by performing threshold tests for reflectance and observed radiance temperature, tests for the difference in observed radiance temperature and reflectance ratio of two channels, and tests for spatial coherence, by using channels of 0.58–0.68 μm, 0.725–1.00 μm, 3.55–3.93 μm, 10.30–11.30 μm, and 11.50–12.50 μm of the Advanced Very High Resolution Radiometer (AVHRR) [Kriebel et al., 2003]. The CLouds from AVHRR (CLAVR) is an algorithm to classify clear/cloudy by using the AVHRR onboard the National Oceanic and Atmospheric Administration (NOAA) polar orbiting satellite, with added tests for nighttime cloud detection to the APOLLO [Stowe et al., 1999]. The MODerate resolution Imaging Spectroradiometer (MODIS) cloud mask algorithm performs several cloud discrimination tests using 22 of 36 bands and creates a cloud mask by combining the confidence level obtained from each test [Ackerman et al., 1998, 2010]. The Atmospheric CO_2 Observations from Space (ACOS) team of the National Aeronautics and Space Administration (NASA) team has created the own product of GOSAT. This ACOS product processing uses the A-Band Oxygen cloud screening algorithm (ABO2) based on FTS (O_2) A-band without using CAI. Parameters such as surface pressure and albedo can be estimated from (O_2) A-band. If the values that these parameters are retrieved from satellite data assuming clear-sky are not realistic, it is determined that clouds or aerosols are present [Taylor et al., 2012]. This algorithm is also used for the Orbiting Carbon Observatory-2 (OCO-2) [O'Dell and Taylor, 2014].

2.1 Overview of CAI-2

The CAI-2 is an imager which measures clouds and aerosols in the atmosphere in each five bands of forward and backward viewing (±20 degrees). The major specifications of CAI-2 are shown in Table 2.1–1. The center wavelengths of each band are 343 nm, 443 nm, 674 nm, 869 nm, and 1630 nm (forward viewing), and 380 nm, 550 nm, 674 nm, 869 nm, and 1630 nm (backward viewing). Except for Bands 5 and 10, the spatial resolution is 460 m and the swath is 920 km. For Bands 5 and 10, the spatial resolution is 920 m and the swath is 920 km. The forward viewing (Band 3–5) and the backward viewing (Band 8–10) are used for the cloud discrimination processing algorithm. The cloud discrimination is performed in both forward and backward viewing (The band used may be added in the future.).

Table 2.1–1. Major required specifications of CAI-2.

	Band	Wavelength [nm]	Spatial resolution [m]	Swath [km]
Forward viewing	1	343	460	920
	2	443		
	3	674		
	4	869		
	5	1630	920	
Backward viewing	6	380	460	920
	7	550		
	8	674		
	9	869		
	10	1630	920	

3 Input and output data

This chapter explains the input and output data for the CAI-2 L2 cloud discrimination processing. Reference data are not necessary for the cloud discrimination processing.

3.1 Input data

The input data are categorized into (Ia) CAI-2 L2 pre-processing results and (Ib) Observation-related information stored in the CAI-2 L1B product (Table 3.1–1).

Table 3.1–1. List of input data to the CAI-2 L2 cloud discrimination processing algorithm.

Input data	Description
(Ia) CAI-2 L2 pre-processing results	Apparent reflectance, Reflectance before atmospheric correction
(Ib) Observation-related information	Viewing direction (forward and backward viewing), Saturation flag, Missing flag, Land/water mask, Solar zenith angle [deg.], Satellite zenith angle [deg.], Solar azimuth angle [deg.], Satellite azimuth angle [deg.], Geodetic latitude and longitude on WGS84 [deg.], Observation time (YYYY/MM/DD hh:mm:ss at UTC)

3.1.1 CAI-2 L2 pre-processing results

The cloud discrimination processing algorithm uses the difference between the apparent reflectance and the reflectance before atmospheric correction created by the CAI-2 L2 pre-processing. It is assumed that the reflectance before atmospheric correction is performed the following processing:

1) Obtaining minimum reflectance at the top of atmosphere to remove clouds and aerosols by using the CAI-2 L1B products for 11 recurrences (one month before and after the observation date of the data for processing).
2) Cloud shadow correction.

3.1.2 CAI-2 L1B product

The following information stored in the CAI-2 L1B product are also used: viewing direction (forward and backward viewing), saturation flag, missing flag, land/water mask, solar zenith angle, satellite zenith angle, relative azimuth angle (calculated from solar azimuth angle and satellite azimuth angle), geodetic latitude and longitude on the World Geodetic System 1984 (WGS84), and observation time (YYYY/MM/DD hh:mm:ss at UTC). The land/water mask is assumed to be a 15-second mesh. At latitudes higher than ±60 degrees, the 1 km land/sea mask provided by the United States Geological Survey (USGS) is processed into a 15-second mesh and applied. At latitudes lower than ±60 degrees, the use of the land/sea mask of the Shuttle Radar Topography Mission (SRTM) can be considered. Figure 3.1.2–1 shows the definition of zenith angle and azimuth angle.

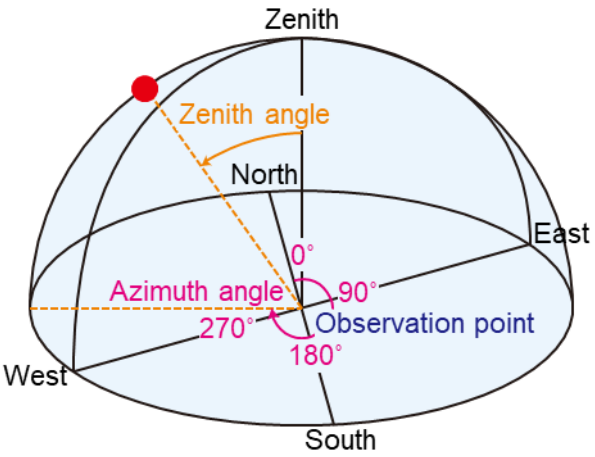


Figure 3.1.2–1. Zenith angle and azimuth angle.
The zenith angle is the angle between the zenith of the observation point and the object.
The azimuth angle is the horizontal component of the direction of the object from the observation point (clockwise with north set to 0 degrees).

3.2 Output data

The output data are cloud discrimination bit fields that store flags for cases easily misidentified in addition to the integrated clear-sky confidence level (integrated-CCL) for each pixel.

(a) Determination of integrated clear-sky confidence level
For CLAUDIA1, the integrated-CCL is determined based on the clear-sky confidence level (CCL) of each discrimination test. The integrated-CCL is stored as a 4-byte real number in the output file.
For CLAUDIA3, the cloud discrimination result is output in the form of integrated-CCL and also stored as a 4-byte real number in the output file.

(b) Output of cloud discrimination processing result
As output information from the cloud discrimination processing algorithm, cloud discrimination bit fields are created, which express the integrated-CCL and other output in bits. 32 bits are allocated to a pixel. Table 3.2–1 presents the cloud discrimination bit fields. As shown in Table 3.2–2, the integrated-CCL is expressed in 4 bits, while the cone angle is expressed in 3 bits.
The cloud discrimination bit fields are output as auxiliary information from the cloud discrimination processing. Therefore, when cloud discrimination processing is not performed, the flag results other than bit 0 (cloud discrimination executed/not executed flag), bit 5 (day/night flag), bit 6–8 (cone angle), and bit 14–18 (saturation flag) are not guaranteed.
Major flags other than the integrated-CCL are introduced as follows:

(i) Cloud discrimination Executed/Not executed flag
When cloud discrimination is not executed for some reason, this flag becomes 1. In this case, the flag results other than the day/night flag, cone angle, water/land flag, saturation flag, and band abnormality flag are not guaranteed.

(ii) Day/Night flag

When the solar zenith angle is 85 degrees or more, it is judged to be nighttime and this flag becomes 1. In this case, cloud discrimination is not executed. The calculation cannot be performed from an invalid solar zenith angle so that the flag result is not guaranteed.

(iii) Cone angle

The definition of cone angle can be referred to Eq. (4.2–1). If any one of the four angles, solar zenith angle, solar azimuth angle, satellite zenith angle, or satellite azimuth angle is an invalid value, the result is not guaranteed.

(iv) Snow possibility flag

The factor behind a high reflectance in the VIS region may sometimes be snow-covered surface, instead of clouds. Since snow cover increases and diminishes quickly, it is difficult to judge based solely on the snow cover database or the like. As an indicator to discriminate snow-covered surface, the Normalized Difference Soil Index (NDSI) is proposed, which is derived from reflectances in 0.674 μm and 1.630 μm (expressed r(0.674 μm) and r(1.630 μm) respectively), as shown below [Hall et al., 1995]:

$$\mathrm{NDSI} = \frac{r(0.674\ \mu\mathrm{m}) - r(1.630\ \mu\mathrm{m})}{r(0.674\ \mu\mathrm{m}) + r(1.630\ \mu\mathrm{m})}\ . \qquad (3.2-1)$$

This approach stems from the fact that the reflectance of snow-covered surface is lower in the SWIR region than in the VIS region (Figure 4.2–3); the larger the NDSI value is, the more likely it is snow-covered surface. However, snow cover in a dense forest produces a small NDSI value, and hence may not be recognized as snow. In the meantime, the discrimination based solely on the NDSI may result in mistaking cirrus clouds for snow-covered surface. For accurate snow-covered surface discrimination, it is essential to combine multiple wavelengths, including a TIR channel which provides an estimate on the surface temperature so that areas with high temperatures can be excluded. In this manner, the discrimination using NDSI does not guarantee a definitive snow-covered surface discrimination, and hence it requires certain caution to apply it. For the purpose of the CAI-2's snow possibility flag, an additional condition of the reflectance in 0.869 μm being 0.11 or higher is set in the NDSI being 0.4 or larger. If any of 0.674 μm, 0.869 μm, or 1.630 μm cannot be used, the flag result is not guaranteed.

(v) Water/Land flag

Based on the USGS land/sea mask, this flag informs whether cloud discrimination processing was performed on water or land. Basically, it matches the USGS land/sea mask, but even if the USGS land/sea mask is an invalid value, if there is no error in other apparent reflectance, it is processed as a water area. In this case, even if landWaterMask under ImageGeometry in the CAI-2 L2 cloud discrimination product is an invalid value, this flag becomes water (00).

(vi) Heavy aerosol possibility flag

When the integrated-CCL is 0.99 or more and the following conditions are satisfied, a probability of heavy aerosols is recognized (r(wavelength): apparent reflectance, R_{min}(wavelength): reflectance before atmospheric correction).

$$\text{Forward viewing: } Dif_1 = r(0.343\ \mu\text{m}) - R_{min}(0.343\ \mu\text{m})\ ,$$
$$\text{Backward viewing: } Dif_1 = r(0.380\ \mu\text{m}) - R_{min}(0.380\ \mu\text{m})\ ,$$
$$Dif_2 = r(0.674\ \mu\text{m}) - R_{min}(0.674\ \mu\text{m})\ ,$$
$$Rat = \frac{(Dif_1 - Dif_2)}{(Dif_1 + Dif_2)}\ ,$$
$$Rat < 0.1 \text{ or } 0.3 < Rat\ .$$

If either 0.343 μm or 0.674 μm cannot be used for forward viewing and 0.380 μm or 0.674 μm for backward viewing, the flag result is not guaranteed.

(vii) Cirrus possibility flag

When the reflectance ratio between the SWIR and near infrared (NIR) regions is the following,

$$0.3 < \frac{r(1.630\ \mu\text{m})}{r(0.869\ \mu\text{m})} < 0.6\ , \qquad (3.2-2)$$

a probability of cirrus clouds is recognized. If either 0.869 μm or 1.630 μm cannot be used, the flag result is not guaranteed.

(viii) Radiance Not saturated/Saturated flag

This flag informs that the radiance in each band of the CAI-2 has reached the saturation level. When saturated, this flag changes to 1.

(ix) Band abnormality flag

Based on information such as the presence or absence of data missing and whether the apparent reflectance is an invalid value, this flag is set to 1 when there is an error. For example, over water areas (except polar regions), only the solar reflectance test can be performed even if bands other than 0.869 μm are not available [Only the solar reflectance among the features can be used.] (see Chapter 4). In this case, since the integrated-CCL using only the solar reflectance test [only the solar reflectance among the features] is output, it is necessary to check the band abnormality flag together.

(x) Flags for each discrimination test result [Only when using CLAUDIA1]

The flags inform the results of the four cloudy/clear discrimination tests individually. Average values of t_1 and t_2 indicated in Sub-section 4.2.1 (b) are used as thresholds in the discrimination tests. However, if the band required for each discrimination test is not available, the flag result of the discrimination test is not guaranteed.

Table 3.2–1. Bit fields of output files of the CAI-2 L2 cloud discrimination processing algorithm.

Bit field	Description	Bit	Wavelength used, data
0	Cloud discrimination Executed/Not executed	0: Executed 1: Not executed	
1–4	Integrated clear-sky confidence level	Table 3.2–2	
5	Day/Night	0: Day 1: Night	Solar zenith angle
6–8	Cone angle	Table 3.2–3	Solar and satellite zenith angles, Solar and satellite azimuth angles
9	Snow possibility	0: No snow 1: Probable snow	0.674 μm, 0.869 μm, 1.630 μm
10–11	Water/Land	00: Water 01: Not used 10: Not used 11: Land	USGS land/sea mask
12	Heavy aerosol possibility	0: No aerosol 1: Probable aerosol	Forward viewing: 0.343 μm, 0.674 μm, Backward viewing: 0.380 μm, 0.674 μm
13	Cirrus possibility	0: No cirrus 1: Probable cirrus	0.869 μm, 1.630 μm
14–18	Radiance (Band 1–5 [6–10]) Not saturated/Saturated	0: Not saturated 1: Saturated	Saturation flag (CAI-2 L1B product)
19–23	Abnormality of Band 1–5 [6–10]	0: Normal 1: Abnormal	Missing/No missing data, Valid/invalid apparent reflectance
24	Discrimination test result (solar reflectance) [Not used in CLAUDIA3]	0: Cloudy 1: Clear	0.674 μm (polar regions and land areas), 0.869 μm (water areas)
25	Discrimination test result (solar reflectance ratio) [Not used in CLAUDIA3]	0: Cloudy 1: Clear	0.674 μm, 0.869 μm (except polar regions)
26	Discrimination test result (NDVI) [Not used in CLAUDIA3]	0: Cloudy 1: Clear	0.674 μm, 0.869 μm
27	Discrimination test result (desert areas) [Not used in CLAUDIA3]	0: Cloudy 1: Clear	0.869 μm, 1.630 μm (only land areas)
28–31	Not used		

Table 3.2–2. Expressions of integrated clear-sky confidence levels in bits.

Integrated clear-sky confidence level	Bit
not less than 0.00 – less than 0.10	0000
0.10 – 0.16	0001
0.16 – 0.22	0010
0.22 – 0.28	0011
0.28 – 0.34	0100
0.34 – 0.40	0101
0.40 – 0.46	0110
0.46 – 0.52	0111
0.52 – 0.58	1000
0.58 – 0.64	1001
0.64 – 0.70	1010
0.70 – 0.76	1011
0.76 – 0.82	1100
0.82 – 0.88	1101
not less than 0.88 – less than 0.94	1110
not less than 0.94 – not greater than 1.00	1111

Table 3.2–3. Expressions of cone angles in bits.

Cone angle [deg.]	Bit
not less than 40 –	000
35 – 40	001
30 – 35	010
25 – 30	011
20 – 25	100
15 – 20	101
10 – 15	110
not less than 0 – less than 10	111

4 Cloud discrimination processing algorithm

Two algorithms, CLAUDIA1 and CLAUDIA3, are available for the cloud discrimination processing, and can be switched. The CAI-2 L2 cloud discrimination processing basically uses CLAUDIA3. This chapter explains the both algorithms.

4.1 Overview of cloud discrimination processing algorithm

Figure 4.1–1 is a flowchart showing CLAUDIA1.

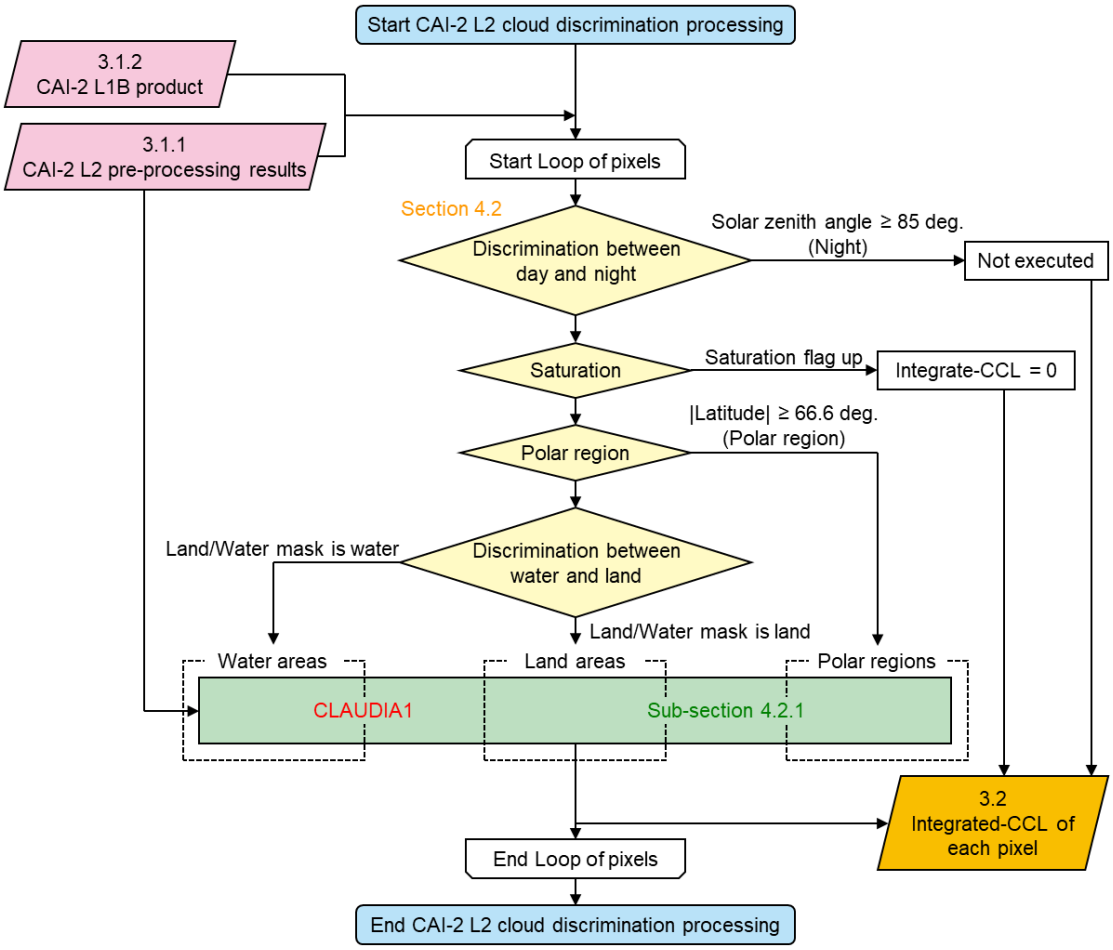


Figure 4.1–1. Processing flow of CLAUDIA1.

Next, Figure 4.1–2 is a flowchart showing CLAUDIA3.

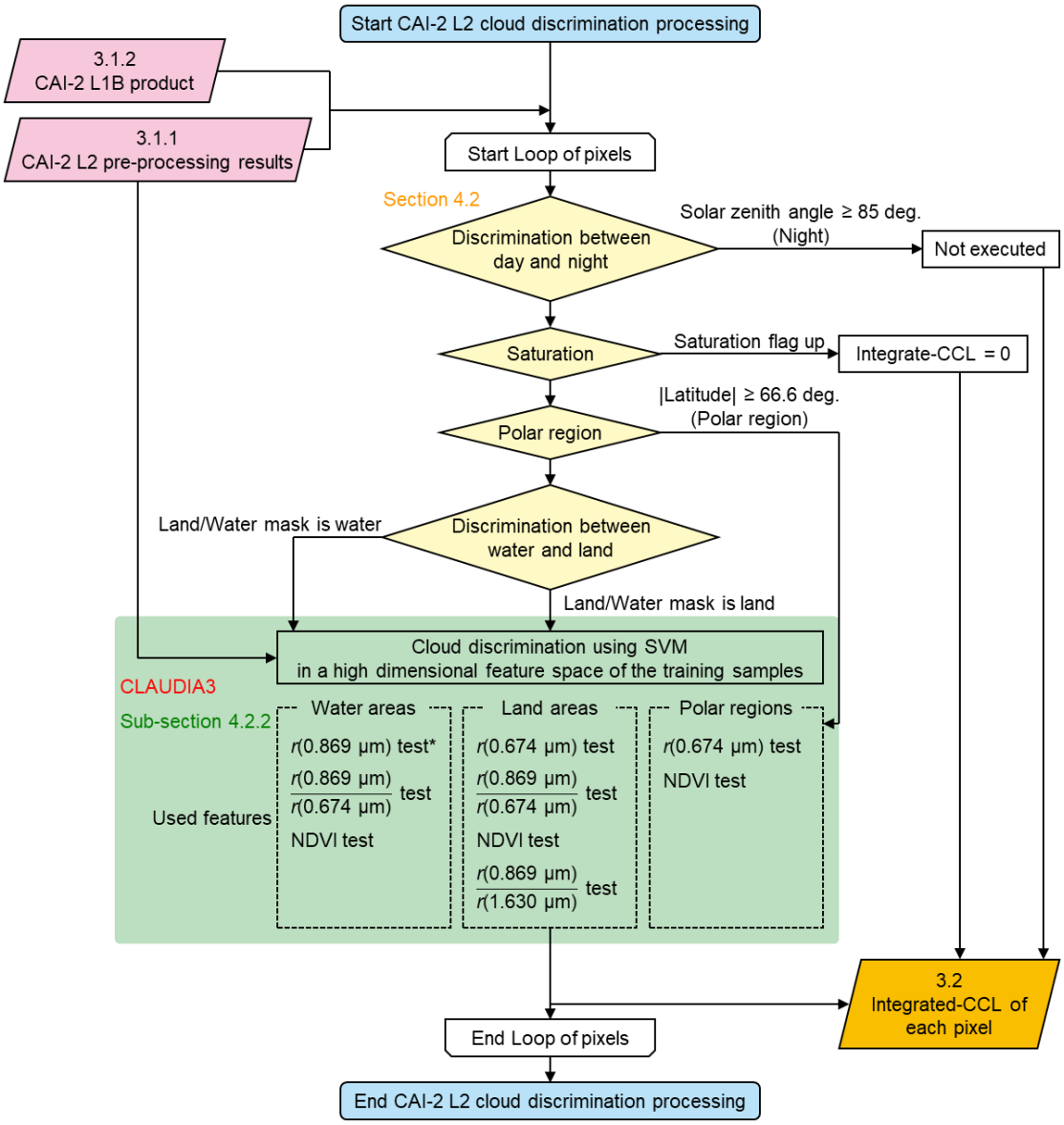


Figure 4.1–2. Processing flow of CLAUDIA3.
r(wavelength) shows the apparent reflectance of the wavelength.
* Cone angle (Eq. (4.2–1)) is considered for solar reflectance, which is one of the features used in water areas.

4.2 Cloud discrimination processing

(a) Cloud discrimination pre-processing

First, this paragraph explains the pre-processing of cloud discrimination.

Using the input data, the pre-processing discriminates day and night, saturation flag, missing flag, polar region, water and land areas, and sunglint area for water area except polar regions for each pixel. The discrimination between day and night is performed based on the solar zenith angle; when the angle is 85 degrees or more, cloud discrimination is not executed. If any one of the saturation flags in each band indicates saturation, the radiance is judged as saturated by clouds and the discrimination result is cloudy. When the latitude is 66.6 degrees or more north or south, the area is regarded as in the polar region. Outside the polar regions, the discrimination between water and land areas is performed by using land/water mask. The discrimination of a sunglint area is performed by calculating the cone angle (Figure 4.2–1) from the solar zenith angle, satellite zenith angle, and relative zenith angle, and then determining whether it is 35 degrees or less over a water area. Furthermore, the apparent reflectance is calculated from the observed radiance by the CAI-2 and the solar irradiation data so as to be used in the discrimination tests. The cone angle is expressed by the following equation using the solar zenith angle (za_1), the satellite zenith angle (za_2), the solar azimuth angle (aa_1), and the satellite azimuth angle (aa_2).

$$\text{Cone Angle} = \cos^{-1}\{\cos(za_1)\cdot\cos(za_2) - \sin(za_1)\cdot\sin(za_2)\cdot\cos(aa_1 - aa_2)\} \quad . \qquad (4.2-1)$$

Cone Angle is the angle between the specular reflection line of sunlight and the viewing direction line of the satellite, as shown in Figure 4.2–1.

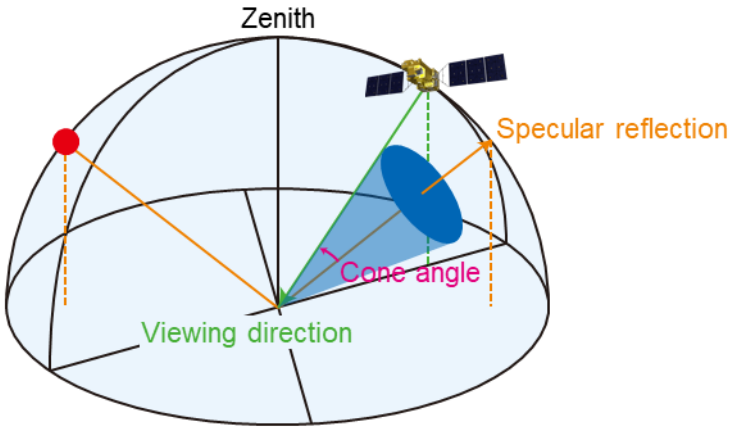


Figure 4.2–1. Cone Angle. The angle between the cone axis and the generatrix when the direction of the specular reflection of the sun is the axis of the cone and the viewing direction of the satellite contacts the cone generatrix.

The following tests are used for the cloud discrimination:

(i) Solar reflectance test
  Polar regions and land areas: *r*(0.674 μm)
  Water areas: *r*(0.869 μm), sunglint discrimination
(ii) Wavelength dependence of reflectance test
  (Except polar regions) Water areas and land areas: *r*(0.869 μm)/*r*(0.674 μm)
(iii) NDVI test
  All areas: NDVI
(iv) Cloud discrimination over desert areas test
  (Except polar regions) Land areas: *r*(0.869 μm)/*r*(1.630 μm).

(b) Solar reflection properties by clouds and ground surface and cloud discrimination
Next, this paragraph explains the solar reflection properties by clouds and the ground surface, which are used in the cloud discrimination processing algorithm.

(i) Solar reflectance
In general, clouds with certain thickness have higher reflectance than the ground surface. In the VIS and SWIR regions where the effect of absorption by atmospheric molecules is relatively small; therefore, an area showing a high reflectance is most likely cloud. When the reflectance in the region is higher than the reflectance before atmospheric correction, which can assume cloud/aerosol-free, clouds can be discriminated.
For water areas, the reflectance of Band 4 [9] (0.869 μm) in the NIR region is used because the impact of the Rayleigh scattering is minimal. On the contrary, the reflectance of Band 3 [8] (0.674 μm) in the VIS region is used for land areas, because in Band 4 [9], the reflectance of vegetation is higher and seasonal variation is greater than Band 3 [8] and thus it is not suitable for cloud discrimination over land areas. The cloud discrimination based solely on reflectance might lead to misjudging a snow and ice or desert, where the reflectance is high, as clouds. Thus, it is vital to combine this cloud discrimination approach with another one. For sunglint areas, the threshold is increased.

(ii) Wavelength dependence of reflectance
In general, the reflectance of clouds does not very likely vary with the wavelength; thus, clouds appear to be whitish. On the other hand, the reflectance of the ground surface is dependent on the wavelength, and thus land areas appear in various colors. Taking advantage of this principle, the reflectance ratio between Band 4 [9] (0.869 μm) in the NIR region and Band 3 [8] (0.674 μm) in the VIS region is taken into the cloud discrimination.

$$\frac{r(0.869\ \mu\mathrm{m})}{r(0.674\ \mu\mathrm{m})}\ , \qquad (4.2-2)$$

In this case, the reflectance ratio of clouds is usually close to 1 and that of the ground surface is often far from 1, making it possible to discriminate clouds (Figure 4.2–2).

(iii) Cloud discrimination over vegetated areas
The Normalized Difference Vegetation Index (NDVI) is often used as an indicator for the level of vegetation growth,

$$\mathrm{NDVI} = \frac{r(0.869\ \mu\mathrm{m}) - r(0.674\ \mu\mathrm{m})}{r(0.869\ \mu\mathrm{m}) + r(0.674\ \mu\mathrm{m})}\ . \qquad (4.2-3)$$

The value of NDVI is larger over vegetated areas, but with the presence of clouds, the value becomes too small to identify vegetation. Applying this characteristic inversely, the NDVI is used for cloud discrimination (Figure 4.2–2).

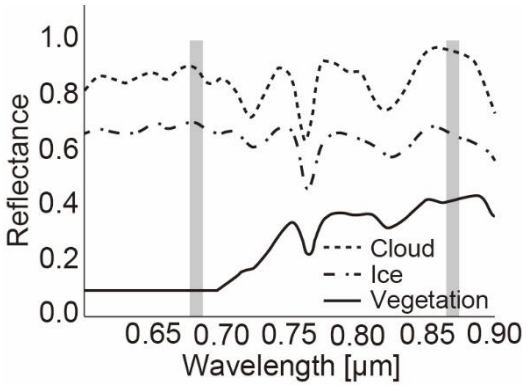


Figure 4.2–2. Spectral reflectance of vegetation and cloud [Griffin et al., 2005]. Cloud has similar reflectances at 0.674 μm and 0.869 μm, whereas vegetation has a higher reflectance at 0.869 μm than 0.674 μm.

(iv) Cloud discrimination over desert areas

The reflectance ratio between Band 4 [9] (0.869 μm) in the NIR region and Band 3 [8] (0.674 μm) in the VIS region usually becomes close to 1 in bright deserts, similarly to clouds; therefore, it is easy to misjudge a clear area as cloudy. To solve this problem, the reflectance ratio between Band 4 [9] (0.869 μm) in the NIR region and Band 5 [10] (1.630 μm) in the SWIR region, is taken into consideration.

$$\frac{r(0.869\ \mu\mathrm{m})}{r(1.630\ \mu\mathrm{m})}\ , \qquad (4.2-4)$$

In this approach, cloud discrimination builds on the tendency that, over desert areas, the reflectance in the SWIR region becomes higher with a longer wavelength (Figure 4.2–3).

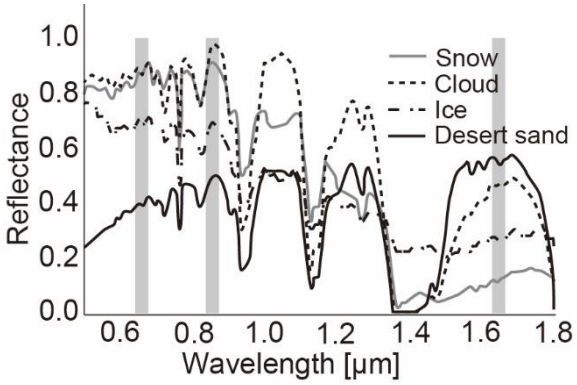


Figure 4.2–3. Spectral reflectance of desert sand and cloud [Griffin et al., 2005]. Cloud has lower reflectance at 1.630 μm than at 0.869 μm, whereas desert sand has higher reflectance at 1.630 μm than at 0.869 μm.

(c) Determination processing of integrated-CCL

Based on the solar reflection properties of clouds and the ground surface, clouds are discriminated in each pixel by comparing the apparent reflectance by the CAI-2 and the threshold. The bottleneck of this discrimination is to define a definitive threshold, because the optical thickness of clouds changes seamlessly and hence there exist gray zones where clouds are less discriminable. Existing cloud discrimination algorithms typically regard these gray zones as clear or cloudy on a case-by-case basis. However, if deemed as clear, it could increase errors in the concentration estimation by the FTS-2, while, on the other hand, if deemed as cloudy, it could reduce the number of FTS-2 data which can be used to estimate the concentration. That being the case, a parameter of CCL [Ishida and Nakajima, 2009; Ishida et al., 2018] is introduced anew in this algorithm in order to quantify the ambiguity in cloud discrimination to a certain extent, thereby making it possible to obtain unbiased and neutral discrimination results. Two algorithms to determine the integrated-CCL are described below.

4.2.1 CLAUDIA1

As illustrated in Figure 4.2.1–1, in CLAUDIA1 [Ishida and Nakajima, 2009], the CCL is determined by comparing the apparent reflectance or reflectance ratio calculated from the calibrated observed radiance against two thresholds, t_1 and t_2 (by definition that the value is greater when the confidence of clear is higher).

(a) CCL and integrated-CCL

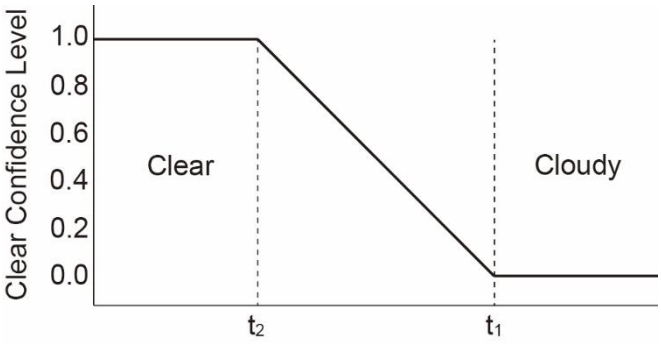


Figure 4.2.1–1. Determination of the CCL in the cloud discrimination. The horizontal axis represents the apparent reflectance or reflectance ratio that is used in the discrimination, whereas the vertical axis represents the CCL. Two thresholds, t_1 and t_2, are set forth. The ambiguous area in between clear and cloudy is expressed in numerical value from 0 to 1.

The CCL of 1 means that the area is clear and that of 0 means that the area is cloudy. Any ambiguous area in between the two levels is expressed in numerical value from 0 to 1 by linear interpolation. If it makes sense to set a range around a certain value as a criterion for determining cloud (or clear), as in the case with the reflectance ratio between the NIR and VIS regions in Eq. (4.2–2) previously mentioned, a boundary greater than the range and another boundary less than the range must be identified by setting two thresholds each, i.e., a total of four thresholds, as illustrated in Figure 4.2.1–2.

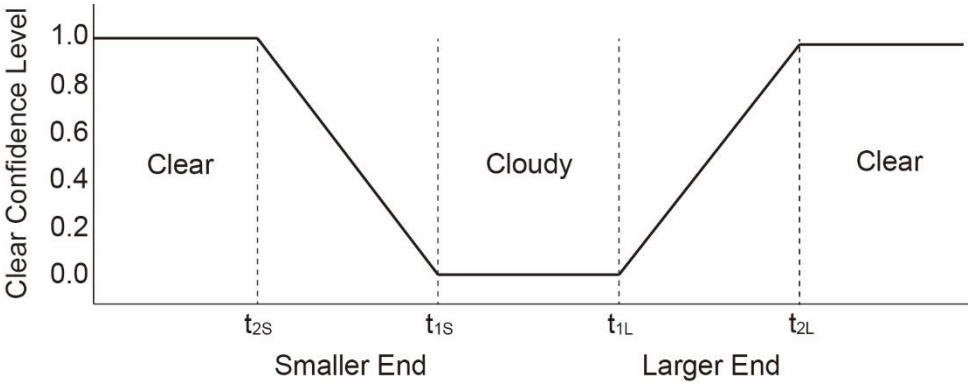


Figure 4.2.1–2. Same as Figure 4.2.1–1, except that there are two clear ranges on both sides.

The CCL determination test stated above is performed: the CCL (F_k) is determined by performing four discrimination tests for land areas, three discrimination tests for water areas, and two discrimination tests for polar regions. Based on the CCL determined by each discrimination test, $F_1 \ldots F_n$ (n: 4 for land

areas, 3 for water areas, and 2 for polar regions), the integrated-CCL (Q) is determined using the equation below. However, if there is a band that cannot be used for some reason, the integrated-CCL is determined using the CCL excluding the discrimination test using the band.

$$Q = 1 - \sqrt[n]{\prod_{i=1}^{n} 1 - F_i} \quad , \qquad (4.2.1-1)$$

The above equation reflects the fact that all discrimination tests available for CAI-2 observation have the potential of "mistaking a clear area for a cloudy area." With this equation applied, if any of the discrimination tests determines an area as clear ($F_k = 1$), $Q = 1$ is established regardless of the results of the other discrimination tests. Conversely, $Q = 0$ is established only when all discrimination tests determine an area as cloudy ($F_k = 0$). In this way, the inclination of each discrimination test to opt for cloudy and the unlikeness of Eq. (4.2.1–1) to do so are merged to maintain the neutrality of the CCL.

(b) Thresholds

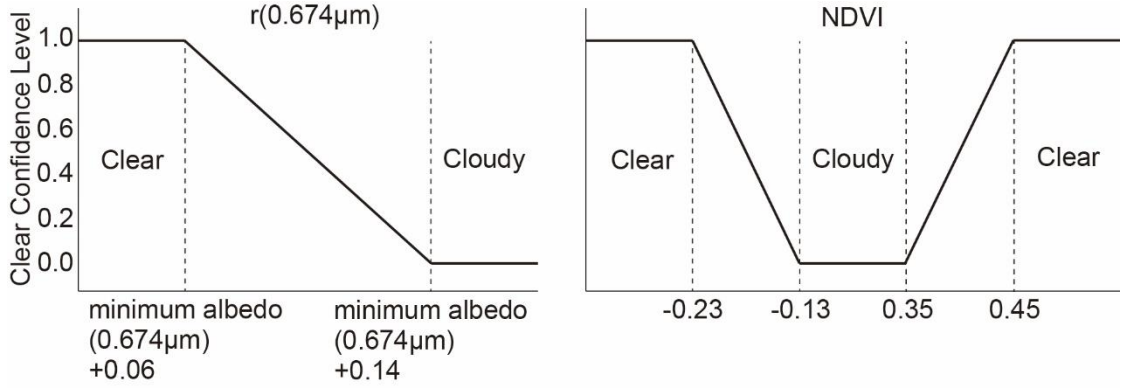
Table 4.2.1–1 summarizes the thresholds for the satellite launch in CLAUDIA1 [Ishida et al., 2011a] (It may change in the future.).

The threshold of solar reflectance test should differ depending on the R_{min} of the pixel position. If a water area is judged as a sunglint area, the threshold is increased according to the cone angle (Eq. 4.2–1). The relationship between the cone angle and the threshold increase is shown in Table 4.2.1–1 (ii).

Table 4.2.1–1. Thresholds for the CAI-2 cloud discrimination processing.

(i) Polar regions (The latitude is 66.6 degrees or more north or south.)

Threshold test		t_1	t_2
r(0.674 μm)		minimum albedo +0.14	minimum albedo +0.06
NDVI	Smaller End	−0.13 (t_{1S})	−0.23 (t_{2S})
	Larger End	0.35 (t_{1L})	0.45 (t_{2L})



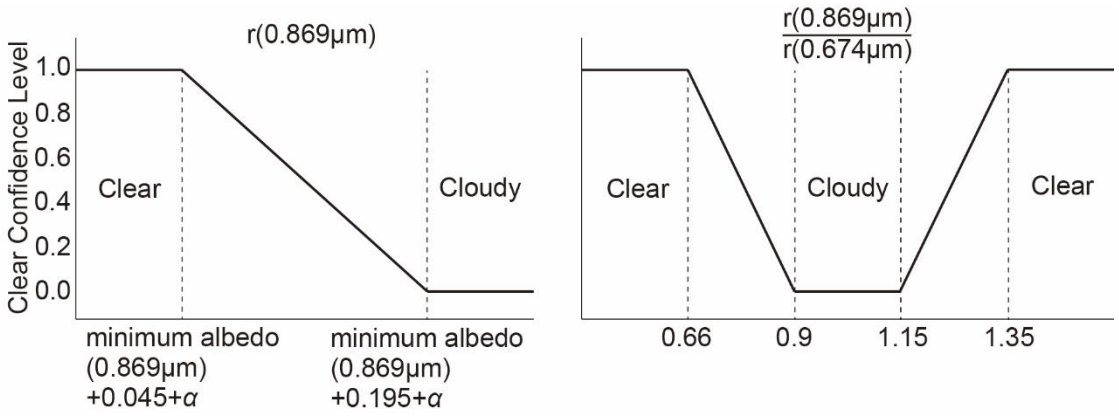
(ii) Water areas (except polar regions)

Threshold test		t_1	t_2
r(0.869 μm)*		minimum albedo +0.195	minimum albedo +0.045
$\frac{r(0.869\ \mu m)}{r(0.674\ \mu m)}$	Smaller End	0.90 (t_{1S})	0.66 (t_{2S})
	Larger End	1.15 (t_{1L})	1.35 (t_{2L})
NDVI	Smaller End	−0.10 (t_{1S})	−0.22 (t_{2S})
	Larger End	0.22 (t_{1L})	0.46 (t_{2L})

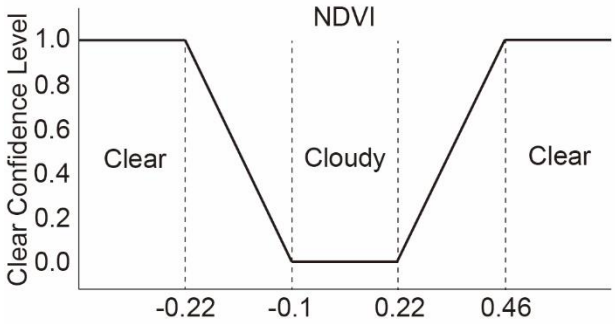
* For sunglint areas, the threshold is further increased.

Degrees of threshold increase for solar reflectance test over sunglint areas.
The cone angles in between the angles shown in the table, the degree of threshold increase is determined by linear interpolation.

Cone angle [deg.]	Degree of threshold increase
35	0.00
30	0.01
25	0.02
20	0.10
15	0.15
10 or less	0.20

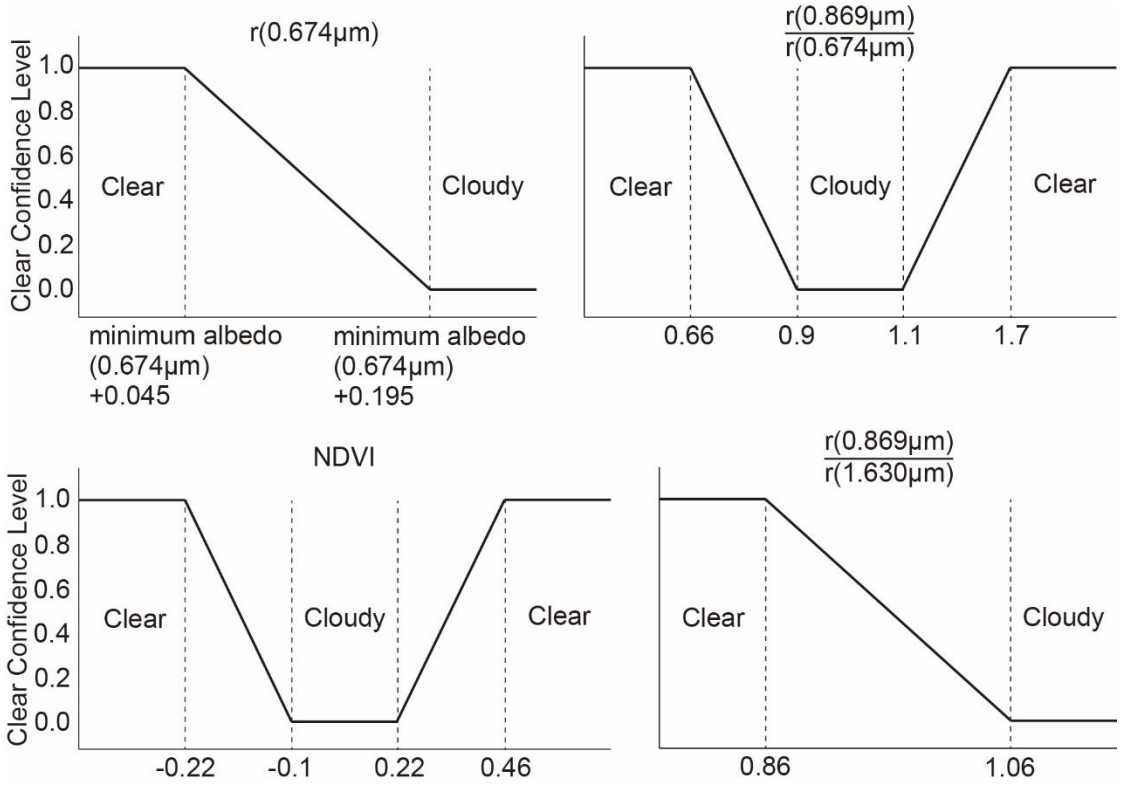


α is the degree of increase of the threshold according to the Cone Angle.



(iii) Land areas (except polar regions)

Threshold test		t_1	t_2
r(0.674 μm)		minimum albedo +0.195	minimum albedo +0.045
$\frac{r(0.869\ \mu m)}{r(0.674\ \mu m)}$	Smaller End	0.90 (t_{1S})	0.66 (t_{2S})
	Larger End	1.10 (t_{1L})	1.70 (t_{2L})
NDVI	Smaller End	−0.10 (t_{1S})	−0.22 (t_{2S})
	Larger End	0.22 (t_{1L})	0.46 (t_{2L})
$\frac{r(0.869\ \mu m)}{r(1.630\ \mu m)}$		1.06	0.86



4.2.2 CLAUDIA3

CLAUDIA1 performs the cloud discrimination by setting thresholds based on experience. CLAUDIA3 performs the cloud discrimination by Support Vector Machine (SVM) in order to objectively determine thresholds using multivariate analysis.

(a) Use of SVM for CLAUDIA3

(i) Decision function and integrated-CCL

CLAUDIA3 [Ishida et al., 2018] applies the kernel trick to soft-margin SVM. The kernel uses a second-order polynomial (The kernel may change as appropriate.).

$$K\ \mathbf{x}_i, \mathbf{x}\ = \frac{\mathbf{x}_i \cdot \mathbf{x} + 1\ ^2}{2}\ ,$$

thus, the decision function is expressed as

$$y = sign\left(\sum_{i \in S} \alpha_i^* t_i \frac{\mathbf{x}_i \cdot \mathbf{x} + 1\ ^2}{2} - h^*\right)\ , \qquad (4.2.2\text{–}1)$$

where $t_i = 1$ is clear and $t_i = -1$ is cloudy, and clear/cloudy is discriminated by the distance from the separating hyperplane (decision function) shown below:

$$D\ \mathbf{x}\ = \sum_{i \in S} \alpha_i^* t_i K\ \mathbf{x}_i, \mathbf{x}\ - h^*\ . \qquad (4.2.2\text{–}2)$$

The decision function $D(\mathbf{x})$ discriminates as follows:

$$\begin{cases} D\ \mathbf{x}\ > 0 & \text{clear} \\ D\ \mathbf{x}\ < 0 & \text{cloudy} \\ D\ \mathbf{x}\ = 0 & \text{Undeterminable} \end{cases} .$$

Since SVM determines the separating hyperplane in a multidimensional space using each discrimination test as a feature without individual discrimination tests, there is no CCL of individual discrimination tests in CLAUDIA1. However, the integrated-CCL can be obtained by taking the absolute value of the decision function shown in Eq. (4.2.2–2).

(ii) Features to use

The features used over polar regions, water areas (except polar regions), and land areas (except polar regions) are shown in Table 4.2.2–1. This is the same as Table 4.2.1–1 for CLAUDIA1 (Some features may be added in the future.).

Table 4.2.2–1. Features to use.

	Water areas	Land areas	Polar regions
Feature	NDVI r(0.869 μm)[*1*2] $\frac{r(0.869\ \mu m)}{r(0.674\ \mu m)}$	NDVI r(0.674 μm)[*1] $\frac{r(0.869\ \mu m)}{r(0.674\ \mu m)}$ $\frac{r(0.869\ \mu m)}{r(1.630\ \mu m)}$	NDVI r(0.674 μm)[*1]

*1 The reflectance before atmospheric correction is used.

*2 The cone angle is taken into consideration.

(iii) Input information specific to CLAUDIA3

The following pre-processing is required to determine Eq. (4.2.2–2).

1) Obtaining training samples required for cloud discrimination.
2) Calculating the support vectors and parameters of Eq. (4.2.2–2) using the training samples.

It is assumed that the processing will be performed in six cases by performing the processing 1) in each of summer and winter, and the processing 2) over three areas: water areas, land areas, and polar regions (It may change in the future.).
Therefore, the support vectors and parameters obtained in the processing 2) are input information specific to CLAUDIA3.

5 Verification of algorithm

This chapter explains the verification results and verification methods of algorithms CLAUDIA1 and CLAUDIA3 used in the CAI-2 L2 cloud discrimination processing.

(a) CLAUDIA1

For CLAUDIA1, accuracy evaluation [Taylor et al., 2012; Ishida et al. 2011b] based on comparison with the MODIS cloud mask algorithm [Ackerman et al., 1998, 2010] and accuracy evaluation based on comparison with visual inspection [Oishi et al., 2014] are performed. Their results show that bare land, semi-vegetation, and the water side of the land-inland water boundary tend to be judged as cloudy even when it is clear, and that cloud edges and thin cirrus clouds are likely to be overlooked.

(b) CLAUDIA3

For CLAUDIA3, an overview of the algorithm [Ishida et al., 2018] and initial results [Oishi et al., 2018] have been reported. Verification based on comparison with visual inspection and CLAUDIA1 in the rainforest, and verification based on comparison with CLAUDIA1, the MODIS cloud mask algorithm, and the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) onboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) at the global level will be performed in the future. CLAUDIA3 is currently being improved and already shows higher accuracy of cloud discrimination than that of CLAUDIA1 at this time. On the other hand, it has been reported that CLAUDIA3 misjudges muddy rivers as cloudy even when it is clear, and it tends to judge clouds excessively [Oishi et al., 2018].

After the launch, the similar verification will be performed by using the actual data.

6 Prerequisites and constraints

(a) How to utilize integrated-CCL

The integrated-CCL for CLAUDIA1 expresses a cloudy area with 0, a clear area with 1, and an ambiguous area with numerical value between 0 and 1. Users can set an arbitrary numerical value between 0 and 1 as a boundary between cloudy and clear areas, so that the area with less than the set value is regarded as cloudy and that with greater than the set value as clear. The closer to 0 the threshold is, the lager the area to be discriminated as clear. The closer to 1 the threshold is, the smaller the area to be discriminated as clear. For reference, when using the CAI-2 L2 cloud discrimination product in the GOSAT-2 operational processing, the threshold is 0.33 (It may change in the future.).

The integrated-CCL for CLAUDIA3 is similarly expressed in numerical value, where 0 means cloudy area, 1 means clear area, and 0.5 means on the separating hyperplane. For reference, when using the CAI-2 L2 cloud discrimination product, the threshold is 0.5 (It may change in the future.).

The CAI-2 L2 cloud discrimination product stores the 4 bits integrated-CCL (16 levels in total). The cloud discrimination can be performed by setting a threshold based on this dataset. Users adjust the threshold between cloudy and clear areas with the integrated-CCL and discriminate clouds.

(b) Cloud determination using FTS-2

Since the CAI-2 does not have a TIR channel, the cloud discrimination based on the cloud top temperature is not feasible. Thus, clouds cannot be discriminated over areas on the night side. Even in areas on the day side, if the clouds are thin, it may well be mistaken for clear. Furthermore, if the solar zenith angle is large, it is difficult to accurately discriminate. In general, cirrus cloud is discriminated by values observed in multiple TIR channels which include water vapor absorption bands; thus, discriminating cirrus cloud using the CAI-2 alone is difficult.

Since the FTS-2 has a 2 μm band including water vapor absorption bands and TIR bands, cirrus clouds can be removed by performing cloud determination using these bands in addition to the CAI-2 L2 cloud discrimination product [Yoshida, 2020].

On the other hand, when analyzing using mainly the CAI-2, FTS-2 data may not be present in the target area. Thus, when it is necessary to use data that does not include cirrus clouds, the cirrus possibility flag in the bit fields of output files of the cloud discrimination processing algorithm (Table 3.2–1) should be referred although the accuracy is lower than that of cirrus cloud removal using the FTS-2.

(c) Cases where cloud discrimination is relatively difficult

(i) Snow and ice areas

As described in Section 3.2 (b) (iv), the CAI-2 does not have a TIR channel; hence the cloud discrimination is relatively difficult over snow and ice areas. Therefore, the snow possibility flag in the bit fields of output files of the cloud discrimination processing algorithm (Table 3.2–1) should be referred.

(ii) Sunglint areas

In CLAUDIA1, the threshold is changed according to the cone angle. In CLAUDIA3, the features with consideration of the cone angle is used. However, the cloud discrimination is relatively difficult over sunglint areas because of the high reflectance. Therefore, the cone angle in the bit fields of output files of the cloud discrimination processing algorithm (Table 3.2–1) should be referred.

(iii) Other high reflectance areas

Other than snow and ice areas and sunglint areas, the cloud discrimination is relatively difficult over high reflectance areas such as dry salt lakes and urban areas.

(iv) Shadows and clouds of reflectance before atmospheric correction

As described in Section 4.2 (a) (i), the solar reflectance test discriminates clouds by comparing the apparent reflectance calculated from the calibrated observed radiance by the satellite with the reflectance before atmospheric correction. In this processing of creating the reflectance before atmospheric correction, cloud shadows and clouds are removed; however, they may remain. If shadows (including cloud shadows) remain in the reflectance before atmospheric correction, it is determined to be high reflectance even if there is no cloud; thus, it is likely to discriminate clouds even if there is no cloud. On the other hand, if clouds remain in the reflectance before atmospheric correction, it is not determined to be high reflectance even if there are clouds; thus, it is likely to discriminate cloud-free even if there are clouds.

References

Ackerman, S. A., K. I. Strabala, W. P. Menzel, R. A. Frey, C. C. Moeller, and L. E. Gumley (1998), Discriminating clear sky from clouds with MODIS, *J. Geophys. Res.*, **103**, 32141–32157, https://doi.org/10.1029/1998JD200032.

Ackerman, S., R. Frey, K. Strabala, Y. Liu, L. Gumley, B. Baum, and P. Menzel (2010), Discriminating Clear-Sky from Cloud with MODIS Algorithm Theoretical Basis Document (MOD35); Version 6.1, Available at: https://modis.gsfc.nasa.gov/data/atbd/atbd_mod06.pdf [accessed on April 30, 2020].

Griffin, M. K., S. M. Hsu, H. K. Burke, S. M. Orloff, and C. A. Upham (2005), Examples of EO-1 Hyperion Data Analysis, *Lincoln Lab. J.*, **15**, 271–298, Available at: https://archive.ll.mit.edu/publications/journal/pdf/vol15_no2/15_2-08.pdf [accessed on April 30, 2020].

Hall, D. K., G. A. Riggs, and V. V. Salomonson (1995), Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data, *Remote Sens. Environ.*, **54**, 127–140, https://doi.org/10.1016/0034-4257(95)00137-P.

Ishida, H. and T. Y. Nakajima (2009), Development of an unbiased cloud detection algorithm for a spaceborne multispectral imager, *J. Geophys. Res.*, **114**, D07206, https://doi.org/10.1029/2008JD010710.

Ishida, H., T. Nakajima, and N. Kikuchi (2011a), Algorithm Theoretical Basis Document for GOSAT TANSO-CAI L2 cloud flag, Available at: https://data2.gosat.nies.go.jp/doc/documents/ATBD_CAIL2CLDFLAG_V1.0_en.pdf [accessed on April 30, 2020].

Ishida, H., T. Y. Nakajima, T. Yokota, N. Kikuchi, and H. Watanabe (2011b), Investigation of GOSAT TANSO-CAI Cloud Screening Ability through an Intersatellite Comparison. *J. Appl. Meteor. Climatol.*, **50**, 1571–1586, https://doi.org/10.1175/2011JAMC2672.1.

Ishida, H., Y. Oishi, K. Morita, K. Moriwaki, T. Y. Nakajima (2018), Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions, *Remote Sens. Environ.*, **205**, 390–407, https://doi.org/10.1016/j.rse.2017.11.003.

Kriebel, K. T, G. Gesell, M. Kästner, and H. Mannstein (2003), The cloud analysis tool APOLLO: Improvements and validations, *Int. J. Remote Sens.*, **24**, 2389–2408, https://doi.org/10.1080/01431160210163065.

O'Dell, C. W. and T. E. Taylor (2014), OCO-2 Algorithm Theoretical Basis Document Oxygen-A Band Cloud Screening Algorithm (ABO2), Available at: https://docserver.gesdisc.eosdis.nasa.gov/public/project/OCO/oco2_abo2_atbd_prelaunch_4.pdf [accessed on April 30, 2020].

Oishi, Y., A. Kamei, Y. Yokota, K. Hiraki, and T. Matsunaga (2014), Evaluation of the Accuracy of GOSAT TANSO-CAI L2 Cloud Flag Product by Visual Inspection in the Amazon and of the Impact of Changes in the IFOV Sizes of TANSO-FTS. *J. Remote Sens. Soc. Jpn.*, **34**, 153–165, https://doi.org/10.11440/rssj.34.153 [in Japanese with English abstract].

Oishi, Y., H. Ishida, T. Y. Nakajima, R. Nakamura, and T. Matsunaga (2018), Preliminary verification for application of a support vector machine-based cloud detection method to GOSAT-2 CAI-2, *Atmos. Meas. Tech.*, **11**, 2863–2878, https://doi.org/10.5194/amt-11-2863-2018.

Rossow, W. B. and L. C. Garder (1993), Cloud Detection Using Satellite Measurements of Infrared and Visible Radiances for ISCCP, *J. Clim.*, **6**, 2341–2369, https://doi.org/10.1175/1520-0442(1993)006<2341:CDUSMO>2.0.CO;2.

Stowe, L. L., P. A. Davis, and E. P. McClain (1999), Scientific Basis and Initial Evaluation of the CLAVR-1 Global Clear/Cloud Classification Algorithm for the Advanced Very High Resolution Radiometer. *J. Atmos. Oceanic Technol.*, **16**, 656–681, https://doi.org/10.1175/1520-0426(1999)016<0656:SBAIEO>2.0.CO;2.

Taylor, T. E., C. W. O'Dell, D. M. O'Brien, N. Kikuchi, T. Yokota, T. Y. Nakajima, H. Ishida, D. Crisp, and T. Nakajima (2012), Comparison of Cloud-Screening Methods Applied to GOSAT Near-Infrared Spectra, *IEEE Trans. Geosci. Remote Sens.*, **50**, 295–309, https://doi.org/10.1109/TGRS.2011.2160270.

Yoshida, Y. (2020), GOSAT-2 TANSO-FTS-2 SWIR L2 Processing Algorithm Theoretical Basis Document [in preparation].