DEVELOPMENT AND VALIDATION OF CLOUD SCREENING ALGORITHM FOR GOSAT/CAI

H. Ishida [a,*], T. Y. Nakjima [a], T. Yokota [b], N. Kikuchi [b], H. Watanabe [b]

[a] Research and Information Center, Tokai University, 2-28-4 Tomigaya, Shibuya-ku, Tokyo 151-0063, Japan - ishidah@yamaguchi-u.ac.jp, nkjm@yoyogi.ycc.u-tokai.ac.jp
[b] National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan -

Commission VIII, WG VIII/3

KEY WORDS: Atmosphere, Meteorology, Development, Observations, Comparison, Satellite, Algorithms

ABSTRACT:

A new and versatile cloud screening algorithm (CLoud and Aerosol Unbiased Decision Intellectual Algorithm, CLAUIDA) bas been developed for multi-channel imagers on board satellites, and is applied to the operational procedure for GOSAT/CAI (a 4-channel imager) cloud screening. CLAUDIA consists of many threshold tests to deal with various cloud types on various surfaces, and is designed to aim at neutral, no-biased cloud screening. Considering that the cloud optical thickness in nature continuously changes and the border between clouds and clear sky is often indistinct, clear confidence level, which is a value to represent certainty of clear or cloud discrimination, is introduced to avoid distinct classification. In addition, threshold tests in the algorithm are categorized into two groups according to their characteristics. A threshold test in group 1 tends to mistake clear sky for cloud, whereas a threshold test in group 2 tends to incorrectly classify cloudy areas as clear. The representative clear confidence level for group 1 is calculated to be cloud conservative, and that for group 2 is to be clear conservative. Consequently, the overall clear confidence level results in neutral. Visual inspection for several examples of cloud screening for CAI L1B data proves that the results over ocean without sunglint and desert are roughly correct. For validation, the results of clear confidence level derived from GOSAT/CAI data are compared to that from Aqua/MODIS data, to which the same cloud screening algorithm is applied. The validation suggests that discrimination over ocean without sunglint is almost in good agreement with each other. Over land, there are some types of the surface where GOSAT/CAI tends to result in cloudy tendency compared to MODIS, because of the lack of effective threshold tests.

1. INTRODUCTION

Greenhouse gases Observing SATellite (GOSAT) is a satellite developed by National Institute for Environmental Studies (NIES) and Japan Aerospace eXploration Agency (JAXA), in order to globally measure precise amounts of greenhouse gases (CO_2 and CH_4) in the atmosphere from space. GOSAT contains two sensors: Fourier Transform Spectrometer (FTS) and Cloud Aerosol Imager (CAI). FTS is the main sensor to observe high-resolution spectra in wide ranges of near infrared and thermal infrared wavelength, whereas CAI is a narrow-band imager that consists of 4 bands at wavelengths from ultra-violet to near infrared (Table 1 summarizes the characteristics of CAI). FTS is designed to measure gaseous absorption of solar radiance and thermal emission. However, because clouds and aerosols also perturb radiation that reaches sensors on satellites, exclusion of cloud areas and correction of effects of aerosol are required to obtain precise amounts of greenhouse gases in the atmosphere. Therefore, GOSAT has been equipped with CAI to carry out cloud screening and to retrieve properties of cloud and aerosol in the field of view of FTS.
GOSAT has been launched in January 2009 and is accumulating observation data enough to use for researches and validations through more than 1 year. Currently, CAI cloud flag product, which is one of the Level-2 data, as well as Level-1B data (radiance data of CAI) are opened to the general public. A cloud screening algorithm, Cloud and Aerosol Unbiased Decision Intellectual Algorithm (CLAUDIA; Ishida et al., 2009) is applied to the procedure for cloud flag data, which are used in the retrieval procedure of gaseous amounts. CLAUDIA is a versatile algorithm to be applied to every multi-channel imager. In this study, we explain the outline of CLAUDIA and its application to CAI. Furthermore, we present validation results of the cloud screening algorithm using the accumulated CAI data.

Band	Center wavelength [μm]	Resolution [m]	Swath [km]
1	0.38	500 [m]	1500
2	0.68	500 [m]	1500
3	0.87	500 [m]	1500
4	1.6	1500 [m]	750

Table 1. The characteristics of CAI.

2. CLOUD SCREENING ALGORITHM

2.1 About CLAUDIA

CLAUDIA consists of several threshold tests, which mean a comparison of a value derived from radiance (for example, the reflectance of a certain wavelength) to the threshold value that determines the border between clouds and clear sky areas. In general, a threshold test is appropriate to a certain type of cloud or Earth's surface but may be not appropriate to other types. Therefore, it is necessary to prepare various types of threshold tests, in order to deal with various cloud types on various surfaces.
CLAUDIA is designed to carry out neutral, no-biased cloud screening. In nature, the border between clouds and clear sky is expected to be indistinct and there are ambiguous areas, because

* Corresponding author. ishidah@yamaguchi-u.ac.jp.

the optical thickness of cloud continuously varies. Furthermore, the definition of "cloud" is indeed often subjective. However, a usual threshold test distinctly discriminates between cloud and clear areas with one threshold value. Several cloud screening algorithms employ the fail-safe concept to have bias in determination, by classifying indistinct areas as cloudy (referred to as "clear conservative") or clear ("cloud conservative"), according to the purpose of observations. On the other hand, CLAUDIA employs the neutral concept with introducing "clear confidence level" and a categorization of threshold tests.

Clear confidence level (originally introduced in MOD35, which is the standard cloud screening algorithm for MODIS data) is an index to represent certainty of clear sky or cloud determination and is applied to avoid distinct discrimination. As illustrated in Figure 1, we define two threshold values, "upper limit" and "lower limit", rather than only one value for a threshold test (here this test is assumed to have a larger value for clear-sky areas than that for cloudy). If an observed value for the threshold test at a pixel is larger (smaller) than the upper (lower) limit, clear confidence level of this pixel is determined as 1 (0). If the observed value is between the upper and lower limit, clear confidence level is determined with linear interpolation as a value between 0 and 1, which indicates that this area is indistinct.

The overall clear confidence level is estimated from the results of all the threshold tests, with categorizing them into two groups. A threshold test in group 1 tends to mistake clear sky for clouds. On the other hand, a threshold test in group 2 tends to incorrectly classify cloudy areas as clear. Considering the characteristics of each group, the representative clear confidence level for group 1, G_1, is determined by

$$G_1 = 1 - \sqrt[n]{(1-F_1)\cdot(1-F_2)\cdots(1-F_k)\cdots(1-F_n)} \tag{1}$$

where F_k is the clear confidence level of the k-th threshold test of group 1, and n is the number of tests in group 1. Eq.(1) means cloud conservative, because G_1 = 1 (clear) even if only one threshold test results in the clear confidence level of 1. The representative clear confidence level for group 2, G_2, is determined by

$$G_2 = \sqrt[n]{F_1\cdot F_2\cdots F_k\cdots F_n} \tag{2}$$

where F_k is the clear confidence level of the k-th threshold test of group 2. Eq.(2) means clear conservative, because G_2 = 0 (cloudy) even if only one threshold test results in the clear confidence level of 0. The overall clear confidence level (Q), which is the final result of cloud screening, is determined by

$$Q = \sqrt{G_1\cdot G_2}\ . \tag{3}$$

Eq.(3) implies that classification as cloud at each group has priority. Q=1 (0) means clear sky (cloudy) and $0<Q<1$ indicates indistinct, ensuring neutral cloud screening.

2.2 Application to CAI

CAI is able to carry out the threshold tests listed in Table 2. The reflectance of band 3 (0.87 µm) is efficient to detect relatively thick clouds over water, whereas the reflectance of band 2 (0.68 µm) is appropriate over land because of large reflectance of leaves in the near infrared region. We apply the "minimum albedo" maps, which are comprised of the minimum reflectance at each band for a month before the date of data for cloud screening, to the threshold test with reflectance. This scheme is based on the assumption that at least one time for a month is clear and the minimum value must correspond to the reflectance of the surface. Pixels that have the cone angle less than 36° over water are expected to be in sunglint areas, where we vary the threshold value for the tests with reflectance according to the cone angle. The ratio of reflectance is also efficient to detect optically thick clouds, because the reflectance of cloud in the solar radiation region is almost independent of the wavelength, whereas the reflectance of the Earth's surfaces usually varies with wavelength. The radio of band 2 to band 3 is appropriate over water and forest, whereas the radio of band 3 to band 4 is quite efficient over bright sand deserts. NDVI that is defined from the reflectance (R) at band 2 and band 3 by

$$NDVI = \frac{R(\text{band 2}) - R(\text{band 3})}{R(\text{band 2}) + R(\text{band 3})} \tag{4}$$

is effective to detect clouds over deep forests. Based on the characteristics of each test, the composition of threshold tests and their threshold values are changed with the surface type, water, land, and the polar region that is defined as the areas with latitude higher than 66.6° or lower than -66.6° (shown in Table 2). All the threshold tests prepared for CAI tend to mistake clear sky for clouds, and then should be categorized in group 1. Thus, the overall clear confidence level is calculated from Eq.(1). Furthermore, possibility of snow surface is determined from NDSI that is given by

$$NDVI = \frac{R(\text{band 2}) - R(\text{band 4})}{R(\text{band 2}) + R(\text{band 4})}\ . \tag{5}$$

Actually, snow/ice surfaces cannot be discriminated from clouds with only NDVI, because some types of cloud also have the same habit. We apply NDVI only as a guide. The flow of CLAUDIA for CAI data is depicted in Figure 2.

(a) water

Threshold test		Lower limit	Upper limit
R(band3) [1]		Rm+0.195	Rm+0.045
R(band3)/R(band2)	smaller	0.9	0.66
	larger	1.15	1.35
NDVI	smaller	-0.1	-0.22
	larger	0.22	0.46

1) The upper limit and the lower limit are increased in sunglint areas.

(b) land

Threshold test		Lower limit	Upper limit
R(band2)		Rm+0.195	Rm+0.045
R(band3)/R(band2)	smaller	0.9	0.66
	larger	1.1	1.7
NDVI	smaller	-0.1	-0.22
	larger	0.22	0.46
R(band3)/R(band4)		1.06	0.86

(c)polar

Threshold test		Lower limit	Upper limit
R(band2)		Rm+0.14	Rm+0.06
R(band3)/R(band2)	smaller	0.9	0.66
	larger	1.1	1.7
NDVI	smaller	-0.13	-0.23
	larger	0.35	0.45

Table 2. The threshold tests in CLAUDIA for CAI and their threshold values. (a) for over water, (b) for over land, and (c) for the polar region. Rm is the minimum albedo.

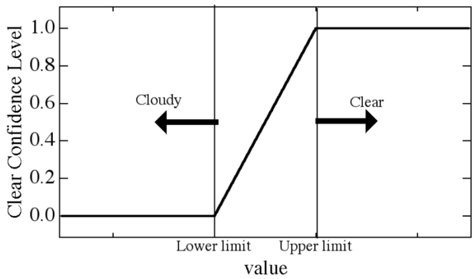

Figure 1. Definition of clear confidence level.

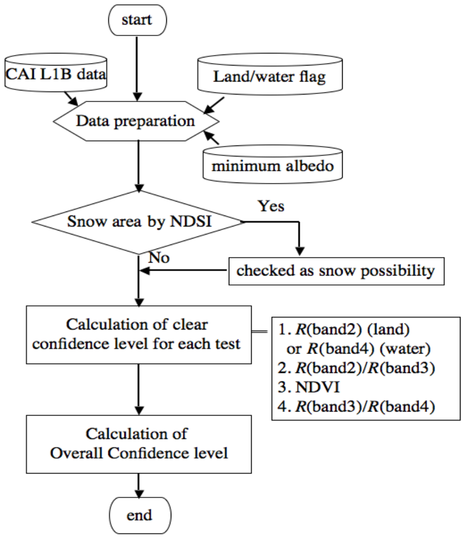

Figure 2. Flow chart of cloud screening procedure for CAI.

3. EXAMPLES OF CLOUD SCREENING

Here we present some examples of cloud screening for CAI L1B data. Before carrying out cloud screening, the L1B radiance data, which are calibrated from pre-launched experiments, are corrected with applying alternative conversion coefficients, because Kikuchi et al. (2010) indicated that the current radiance has some problems. The coefficients were estimated from a vicarious calibration by comparison to MODIS data (Ishida et al., 2010).

Figure 3 shows an example of cloud screening by CAI over ocean without sunglint. The visual inspection with the RGB image from MODIS data reveals that cloud screening by CAI is roughly adequate for scenes over ocean. The margins of clouds and the gaps tend to have clear confidence level of between 0 and 1, which suggests that CLAUDIA extracted indistinct areas as expected. Figure 4 illustrates an example over a land surface, which is mainly occupied by desert. From the visual inspection, broken clouds in the southern part of the scene seem to be correctly detected, and clear areas almost have clear confidence level of 1. However, some areas that seem to be clear in the northern part of the scene result in clear confidence level of about 0.5. This suggests that there are some types of Earth's surface where CAI tends to confuse clear areas as clouds.

4. VALIDATION

Cloud screening results by CAI are compared to those by Aqua/MODIS for validation. We apply the same cloud screening algorithm, *i.e.,* CLAUIDA, in order to eliminate

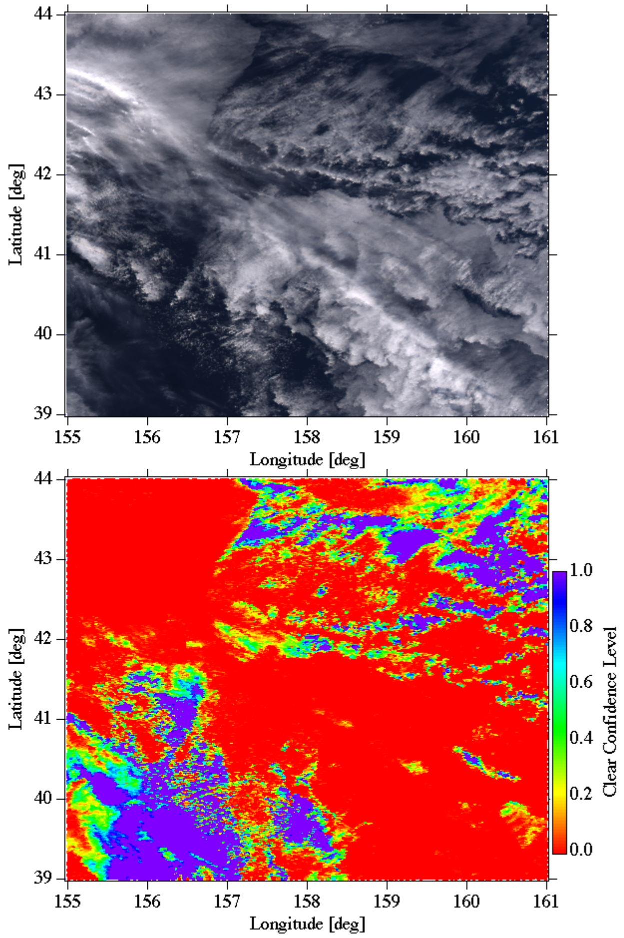

Figure 3. An example of cloud screening over ocean. (*upper*) The RGB composite from Aqua/MODIS data and (lower) distribution of clear confidence level estimated from CAI data at 2:41 (UTC) 14th Nov.,2009.

effects of the difference of algorithm and investigate the difference of sensor itself. Aqua/MODIS is able to carry out not only all the threshold tests that CAI is able to do, but also do other threshold tests. Therefore, cloud screening by Aqua/MODIS is expected to be more correct than that by CAI. We extract the CAI and Aqua/MODIS pixels that are coincident within 0.005˚ of latitude and longitude, and within the interval of five minutes.

Figure 5 depicts the comparison of cloud screening results in the scene of Figure 3, red (blue) color meaning clear (cloudy) tendency by Aqua/MODIS. The result of CAI at both the apparent clear and cloudy areas discriminated with the visual inspection is in good agreement with the result of Aqua/MODIS. A few areas expected to be very thin clouds have clear tendency of MODIS. The margins of clouds tend to have the discrepancy, which is supposed to be due to cloud moving during the interval of the observation time or the difference of viewing angle of each sensor. In the scene, the pixels at which the cloud screening results of both satellites are completely opposite each other (*e.g.*, clear by CAI but cloudy by MODIS) are few. The validation indicates that the cloud screening by CAI over ocean without sunglint is generally reliable, even though the number of channels of CAI is less than that of MODIS.

Figure 6 is the same as Figure 5 but for the comparison in the scene of Figure 4.

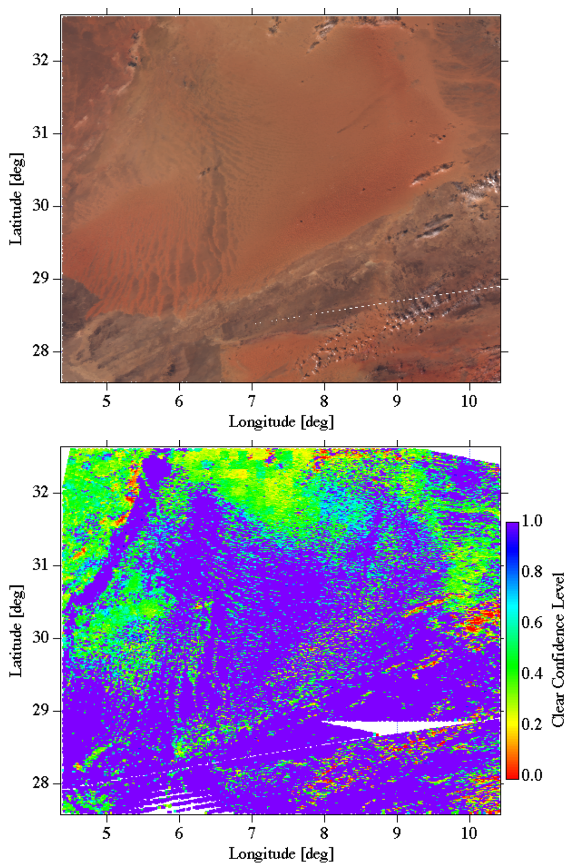

Figure 4. An example of cloud screening over land. (*upper*) The RGB composite from Aqua/MODIS data and (*lower*) distribution of clear confidence level estimated from CAI data at 12:35 (UTC) 5th Nov., 2009.

Figure 6 is the same as Figure 5 but for the comparison in the scene of Figure 4. The northern part of the scene has cloudy tendency by CAI. In general, clouds over brightly sand desert can be detected by the reflectance ratio of 0.87µm to 1.6µm, which tends to incorrectly identify some types of bare land and steppe as cloudy. MODIS identified this area as clear mainly by the threshold test with the reflectance ratio of 0.55µm to 1.24µm, which CAI does not contain. The comparison reveals that cloud screening by CAI over land sometimes has cloudy tendency and it is necessary to be careful to apply the results.

5. SUMMARY

We have developed a new cloud screening algorithm, CLAUDIA, which is currently applied to operational cloud screening for GOSAT/CAI data. The most important characteristics of CLAUDIA is to carry out neutral, no-biased discrimination between clear and cloudy areas. To realize the neutral concept, clear confidence level is incorporated to avoid a distinct discrimination between clear sky and cloud. Furthermore, we categorize threshold tests into two groups, and the representative clear confidence level of each group is calculated to be cloud conservative or clear conservative according to their characteristics. We carried out cloud screening for CAI data with CLAUDIA and estimated adequacy of the results for several examples with visual inspection. It was suggested that cloud screening by CAI was roughly correct for scenes of ocean without sunglint and desert. The differences in clear confidence level between CAI and Aqua/MODIS were derived for validation. In scenes of ocean without sunglint, clear confidence level of both satellites is almost in good agreement with each other, especially for apparently thick clouds and clear areas. On the other hand, some types of the surface of land had a large discrepancy of clear confidence level. The validation suggests that cloud screening by CAI over land sometimes tend to have cloudy tendency because of the lack of effective threshold tests that MODIS can carry out.

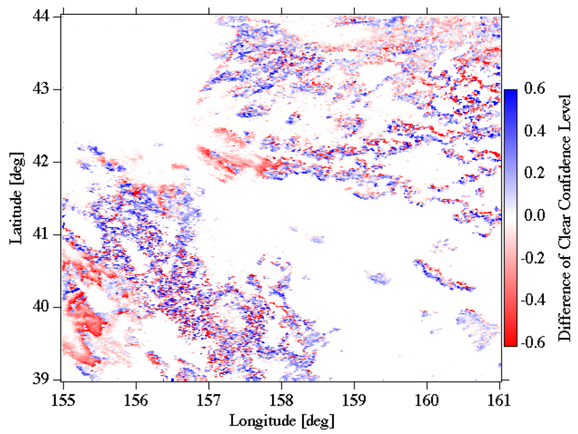

Figure 5. The difference of clear confidence level by Aqua/MODIS from that by CAI: Red (blue) color means clear (cloudy) tendency by Aqua/MODIS i.e., cloudy (clear) tendency by CAI at 2:41 (UTC) 14th Nov.,2009.

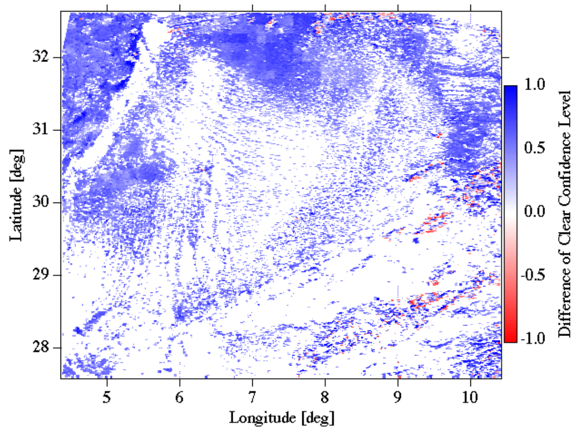

Figure 6. The difference of clear confidence level by Aqua/MODIS from that by CAI: Red (blue) color means clear (cloudy) tendency by Aqua/MODIS i.e., cloudy (clear) tendency by CAI at 12:35 (UTC) 5th Nov., 2009..

Acknowledgements
This work was supported by the Greenhouse Gases Observing Satellite (GOSAT) Science Project of the National Institute of Environmental Studies (NIES), Tsukuba, Japan (2006-2010

References

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

Ishida, H., Nkajima, T. Y., and Kikuchi, N., 2010. Algorithm theoretical basis document for processing GOSAT TANSO-CAI Level 2 cloud flag. National Institute for Environmental Studies.