Dear Dr. T. E. Taylor,

We are very grateful for your polite comments and suggestion to our manuscript. We have carefully revised the original submission of this manuscript in accordance with your suggestion.

It is right that CLAUDIA1 needs to set various algorithm thresholds by radiative transfer calculation results and fine tuning in some method. In this study we used thresholds in ATBD of the GOSAT CAI L1 cloud flag product. Thus, it is also right that visual inspection results can optimize thresholds of CLAUDIA1 to apply for locally cloud discrimination. On the other hand, CLAUDIA3 uses machine learning method, and so CLAUDIA3 needs only obvious clear-sky training data on various land cover types. Furthermore CLAUDIA3 has a multiplicity of uses; actually, Support Vectors and thresholds for CLAUDIA3 used in this study were generated from MODIS data.
In this time we want to check accuracy improvement of cloud discrimination results by CLAUDIA3 in comparison with CLAUDIA1 in tropical rainforest areas. To realize this, we have made validation dataset from 2014 like USGS Landsat 8 cloud cover assessment validation data (https://landsat.usgs.gov/landsat-8-cloud-cover-assessment-validation-data) with visual inspection. Then, we used our dataset in tropical rainforest areas not to optimize thresholds but only to evaluate CLAUDIA1 and 3.

Meanwhile, you suggested very useful additional analysis to compare directly CLAUDIA1 and 3 for various land cover types. Then, we performed the analysis and added its results.

We corrected our manuscript according to your minor and technical comments.
We are deeply grateful to you.

Yours sincerely,

Yu Oishi, PhD
Artificial Intelligence Research Center,
National Institute of Advanced Industrial Science and Technology, Japan

Dear Anonymous Referee #1,

We are very grateful for your polite comments and carefully indications to our manuscript. We have carefully revised our manuscript.
The answers to the comments are as follows.

1. The language throughout the paper needs to be improved, the topic of this study is mainly on the cloud detection, while in Introduction section, the discussion is mainly on the CO2 product. I suggest to reorganize the Introduction section and provided more background knowledge on the motivation of this work.

->
Thank you for your opinion. The role of GOSAT CAI L2 cloud flag products and GOSAT-2 CAI-2 L2 cloud flag products is not only cloud discrimination of satellite images but also cloud discrimination to increase accuracy of CO_2 concentration estimates. Greenhouse gases observing satellites have many targets and roles, such as to monitor CO_2 hot spots, to monitor climate change, to monitor the impacts of human activities, and to contribute to climate science and climate change related policies. Among them, we focused on the role as MRV for climate change related policies. If we focus on CO_2 hot spots monitoring, it may be better to remove all cloud and ambiguous pixels. However, to reduce the uncertainty of flux estimation in tropical rain forests, it is also necessary to increase the number of clear-sky FTS data. This means that both omission error and commission error need to be lower at the same time. Meanwhile, these products don't optimize thresholds to specific local regions because they are required to be calculated under the same condition all over the world. Under these situations, we need to compare between CLAUDIA1-CAI and CLAUDIA3-CAI in tropical rain forests. In other words, introduction section needs to describe not only cloud detection of satellite images but also CO_2 concentrate estimation and climate change related policies.
----

2. In the abstract, the authors referred the algorithm is developed in another paper (Page 1, Line 19-20), there is no detailed discussion on the difference between these two close related works.

->
Thank you for your question. We described the detail of two algorithms (CLAUDIA1 and CLAUDIA3) in section 2.2 and 2.3. CLAUDIA1 performs multiple thresholding tests, and then integrate the results. On the other hand CLAUDIA3 uses a machine learning method. Therefore

CLAUDIA3 can automatically identify the optimized thresholds using clear-sky training data, although CLAUDIA1 requires setting various thresholds by radiative transfer calculation results and fine tuning in some method.

----

3. The uncertainty of the CLAUDIA-CAI is not fully analyzed and discussed. Several cases are shown in Figure 9 and 10 to describe the performance of CLAUDIA-CAI over various land cover types, however, there is no quantitative results and discussion. There needs to be a better description in Section 3.1.

->

Thank you for your suggestion. I added data statistics in Section 3.1.

----

4. Figure 6 and 7 are very similar with each other in the upper panel. It would be more concise to combine these two figures.

->

Thank you for your opinion. Your indication “Figure 6 and 7 are very similar with each other in the upper panel” is right. On the other hand, we think that flow chart of different algorithm should be divided to facilitate coding and easy to understand.

----

5. Page 2 line 24-25: what is meant by “Accuracy 16 times higher than at present is required assuming that the MRV for REDD+ and JCM needs an accuracy of 10 %”?

->

Thank you for your indication. We revised it as follows.

“It is required to reduce the uncertainty of the L4A CO_2 product by a factor of 16 assuming that the MRV for REDD+ and JCM needs an accuracy of 10 %.”

----

6. Page 5 line 21-30: “GOSAT returns to a similar footprint after 44 orbits (44 CAI paths) in three days. The satellite ground path of one orbit is divided into 60 equidistant CAI frames. We used the GOSAT CAI L1B product, which general users could download from the GOSAT User Interface Gateway (GUIG, https://data.gosat.nies.go.jp), for various land cover types on

the beginning of the month from 2012 to 2014 in the same as the previous study (Oishi et al., 2017) (Table 2), and for rainforests (Table 3). Currently GUIG has been changed to GOSAT Data Archive Service (GDAS, https://data2.gosat.nies.go.jp/index_en.html). The spatial resolution of these products (pixel size at nadir) is 500 m, the image size is 2048 × 1355 pixels (approximately 1000 × 680 km). The CLAUDIA algorithm requires a land/sea mask and a surface albedo data. The CAI L1B product includes the Shuttle Radar Topography Mission's 15" land/sea mask. For areas with latitudes higher than ±60°, the USGS Global Land 1-KM AVHRR Project mask is used. Surface albedo data at 1/30° resolution was generated from the CAI L1B data from 10 recurrent cycles by separating the land and water regions."
This section describes the data involved in the validation of algorithm, while the spatial resolution is inconsistent, did you resample the data?

->
Thank you for your indication. Because our description was unclear, I modified it as follows.
"The CAI L1B product includes a land/sea mask with 500 m resolution which is generated from the Shuttle Radar Topography Mission's 15″ land/sea mask and the USGS Global Land 1-KM AVHRR Project mask for areas with latitudes higher than ±60°."
----

7. Page 7 table 2: the last line "1 April 2012–1 March" data lacks the information of the year.

->
Thank you for your carefully indication. I added the year.
----

8. Page 7 table 3: the GOSAT product applied to Borneo seems only half of that to Amazon, why?

->
Thank you for your question. We mainly used CAI Path 7 Frame 31 for Borneo and CAI Path 29_32 for the Amazon, and their surroundings CAI data. Borneo can be covered by 2 CAI scenes, but Amazon needs 9 scenes (see Figure 4). For this reason, the number of CAI scenes for the Amazon is larger than that for Borneo.
----

9. Page 8 figure 6: “Solar Zenith Angle >=85 (Night)” Could the authors provide some related references?

->
Thank you for your question. The algorithm theoretical basis document (ATBD) for GOSAT CAI L2 cloud flag product is open to the public.
https://data2.gosat.nies.go.jp/GosatDataArchiveService/doc/GU/ATBD_CAIL2CLDFLAG_V1.0_en.pdf
On the other hand, the ATBD for GOSAT-2 CAI-2 L2 cloud discrimination product has not been open to be the public yet.
----

10. Page 12 line 4-5: “clear despite cloudy” and “cloudy despite clear” are vague, and these terms should be replaced with already explained alphabets. Furthermore, it is better to provide references to the two formulas.

->
Thank you for your suggestion, but we couldn’t find out relevant items.
----

11. Page 14 figure 9, Page 17 figure 11, Page 18 figure 12, Page 21 figure13-14: the name of vertical axis just like “Percent” could be added.

->
Thank you for your indication. I changed “%” to “Percent” in Figures 9, 11-14.
----

12. Page 19 table 5: the second CLAUDIA3-CAI 0.56 should be corrected to 0.5.

->
Thank you for your carefully review comment. I corrected it.
----

13. Page 23 line 9-10: The authors wrote the results that “The averaged accuracy of CLAUDIA3 used with GOSAT CAI data (CLAUDIA3-CAI) was approximately 89.5 % in tropical rainforests, which was greater than that of CLAUDIA1-CAI (85.9 %) for the test cases

presented here.” But how to calculate 89.5% and 85.9% two values?

->

Thank you for your question. We think that accuracy of cloud discrimination is different between in Borneo and the Amazon. For this reason, we handled them as being independent, and then averaged 87.0 and 84.8 for CLAUDIA1 and averaged 92.0 and 86.9 for CLAUDI3.

----

14. Page 23 line 6: purposeâ˘A˘T> the purpose

->

Thank you for your carefully review comments. I corrected it.

----

15. Page 24 line 8: formâ˘A˘T> from

->

Thank you for your carefully review comments. I corrected it.

Preliminary verification for application of a support vector machine based cloud detection method to GOSAT-2 CAI-2

Yu Oishi[1], Haruma Ishida[2], Takashi Y. Nakajima[3], Ryosuke Nakamura[1], Tsuneo Matsunaga[4]

[1]National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi, Koto, Tokyo 135-0064, Japan
[2]Meteorological Research Institute, 1-1 Nagamine, Tsukuba, Ibaraki 305-0052, Japan
[3]Reseach and Information Center, Tokai University, 2-28-4 Tomigaya, Shibuya, Tokyo 151-0063, Japan
[4]National Institute for Environmetal Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan

Correspondence to: Yu Oishi (oishi.yu@aist.go.jp)

Abstract. The Greenhouse Gases Observing Satellite (GOSAT) was launched in 2009 to measure global atmospheric CO_2 and CH_4 concentrations. GOSAT is equipped with two sensors: the thermal and near-infrared sensor for carbon observation (TANSO)-Fourier transform spectrometer (FTS) and TANSO-cloud and aerosol imager (CAI). The presence of clouds in the instantaneous field of view of the FTS leads to incorrect estimates of the concentrations. Thus, the FTS data suspected to have cloud contamination must be identified by a CAI cloud discrimination algorithm and rejected. Conversely, overestimating clouds reduces the amount of FTS data that can be used to estimate greenhouse gases concentrations. This is a serious problem in tropical rainforest regions, such as the Amazon, where the amount of useable FTS data is small because of cloud cover. Preparations are continuing for the launch of the GOSAT-2 in fiscal year 2018. To improve the accuracy of the estimates of greenhouse gases concentrations, we need to refine the existing CAI cloud discrimination algorithm: Cloud and Aerosol Unbiased Decision Intellectual Algorithm (CLAUDIA1). A new cloud discrimination algorithm using a support vector machine (CLAUDIA3) was developed and presented in another paper. Although the use of visual inspection of clouds as a standard for judging is not practical for screening a full satellite data set, it has the advantage of allowing for locally optimized thresholds, while CLAUDIA1+3 use common global thresholds. Thus, the accuracy of visual inspection is better than that of these algorithms in most regions, with the exception of snow and ice covered surfaces, where there is not enough spectral contrast to distinguish cloud. For this reason visual inspection can be used for the truth metric for the cloud discrimination verification exercise. In this study, we compared between CLAUDIA1-CAI and CLAUDIA3-CAI for various land cover types, and evaluated the accuracy of CLAUDIA3-CAI by comparing the both of CLAUDIA1-CAI and CLAUDIA3-CAI against visual inspection of the same CAI images in tropical rainforests. Comparative results between CLAUDIA1-CAI and CLAUDIA3-CAI for various land cover types indicated that CLAUDIA3-CAI had tendency to identify bright surface and optically thin clouds, however, misjudge the edges of clouds as compared with CLAUDIA1-CAI. The accuracy of CLAUDIA3-CAI was approximately 89.5 % in tropical rainforests, which is greater than that of CLAUDIA1-CAI (85.9 %) for the test cases presented here.

Copyright statement. We confirm that this manuscript has not been published elsewhere and is not under consideration by another journal. All authors have approved the manuscript and agree with submission to AMT. The authors have no conflicts of interest to declare.

Funding. This research is supported by the GOSAT-2 Project at the National Institute for Environmental Studies (NIES), Japan (2015, 2016); and based on results obtained as part of a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO). NIES and NEDO had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

1 Introduction

The Greenhouse Gases Observing Satellite (GOSAT) was launched in 2009 to measure global atmospheric CO_2 and CH_4 concentrations. Preparations are continuing for the launch of its successor, GOSAT-2, in fiscal year 2018. The mission objectives of GOSAT-2 are as follows: to continue and improve the satellite measurements of major greenhouse gases performed by GOSAT; to monitor the effects of climate change and human activities on the carbon cycle; and to contribute to climate science and climate change related policies (NIES GOSAT-2 Project, 2014). These policies include Reducing Emissions from Deforestation and forest Degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks in developing countries (REDD+), and the Joint Crediting Mechanism (JCM), which was proposed by the Japanese government to facilitate the diffusion of leading low-carbon technologies, products, systems, services, and infrastructure in developing countries (Ministry of the Environment, Japan, 2015). Monthly regional CO_2 fluxes are estimated from the column-averaged dry air mole fractions of CO_2 (XCO_2) retrieved from spectral observations made by GOSAT (Maksyutov et al., 2013). The results are publicly available as the L4A CO_2 product (Maksyutov et al., 2014). The expected role of the CO_2 fluxes estimated from the GOSAT data is the system for measurement, reporting and verification (MRV) of CO_2 fluxes estimated from forest inventory data. Currently, the uncertainty of the L4A CO_2 product is about 0.9 Gt-C/region/year in the Amazon (L4A CO_2 product V02.03 in region ln 09-12, 2009-2012). Thus, total net CO_2 flux from deforestation for the period 2000–2010 in tropical America was estimated to be 0.56 Gt-C/year (Baccini et al., 2012). It is required to reduce the uncertainty of the L4A CO_2 product by a factor of 16 assuming that the MRV for REDD+ and JCM needs an accuracy of 10 %.

GOSAT is equipped with two sensors: the Thermal and Near-infrared Sensor for Carbon Observation (TANSO)-Fourier Transform Spectrometer (FTS) and TANSO-Cloud and Aerosol Imager (CAI) (Table 1). The presence of clouds in the instantaneous field of view of the FTS leads to incorrect estimates of greenhouse gas concentrations (Uchino et al., 2012). To solve the problem, the FTS data suspected to have cloud contamination must be identified by the Cloud and Aerosol Unbiased Decision Intellectual Algorithm used with CAI (CLAUDIA1-CAI) (Ishida and Nakajima, 2009) and rejected. The cloud information is publicly available as the CAI L2 cloud flag product. However, CAI does not have a thermal infrared band. In general, cirrus cloud is identified by using multiple thermal infrared bands, which include water vapour absorption bands (Ishida et al., 2011a). Meanwhile, the FTS has a 2 μm band that contains many strong water vapour absorption bands. Moreover, the CAI L2 cloud flag product may not be sensitive enough to detect clouds of sub-pixel size in ocean observations. To cope with these difficulties, the FTS data suspected to have cloud contamination are identified by two additional tests: the 2 μm band test and the CAI coherent test (Yoshida et al., 2010). Conversely, overestimation of clouds reduces the amount of the FTS data that can be used to estimate greenhouse gas concentrations. This is a serious problem in tropical rainforest regions, such as the Amazon, where there is a small amount of suitable FTS data (approximately 3 % of the number of observations) because of cloud cover (Figs. 1, 2). For the reason we need to optimize thresholds between cloud and clear-sky because there are tradeoffs in maximizing cloud detection accuracy while minimizing false detection. To solve the problem, a new cloud discrimination algorithm (CLAUDIA3) using a support vector machine (SVM) (Vapnik and Lerner, 1963) was developed (Ishida et al., 2018). CLAUDIA3 can automatically identify the optimized thresholds using clear-sky training data, although CLAUDIA1 requires setting various thresholds by radiative transfer calculation results and fine tuning in some method. Verification was also performed by comparing with the MODIS cloud mask algorithm (Ackerman et al., 2010) and ceilometer data provided by Atmospheric Radiation Measurement (ARM) (Mather and Voyles, 2013) in the paper (Ishida et al., 2018). Furthermore the impact of different Support Vector generation procedures on cloud discrimination using CLAUDIA3 has also been evaluated in a previous study (Oishi et al., 2017).

Table 1: Specifications of CAI.

	Band 1	Band 2	Band 3	Band 4
Spectral coverage (μm)	NUV 0.370-0.390	Red 0.664-0.684	NIR 0.860-0.880	SWIR 1.56-1.65
Swath (km)	1000	1000	1000	750
Spatial resolution At nadir (m)	500	500	500	1500

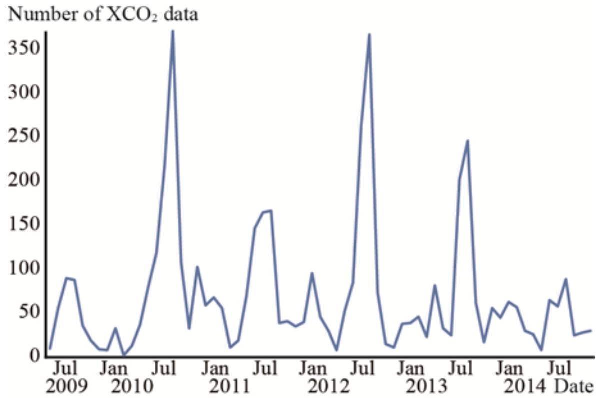

Figure 2: Monthly changes in the number of FTS L2 XCO_2 data in the Amazon. The five-point cross track scan mode was used until 1 August 2010, when it was replaced with the three-point cross track scan mode. Therefore the numbers themselves before and after 1 August 2010 cannot be compared. (single column)

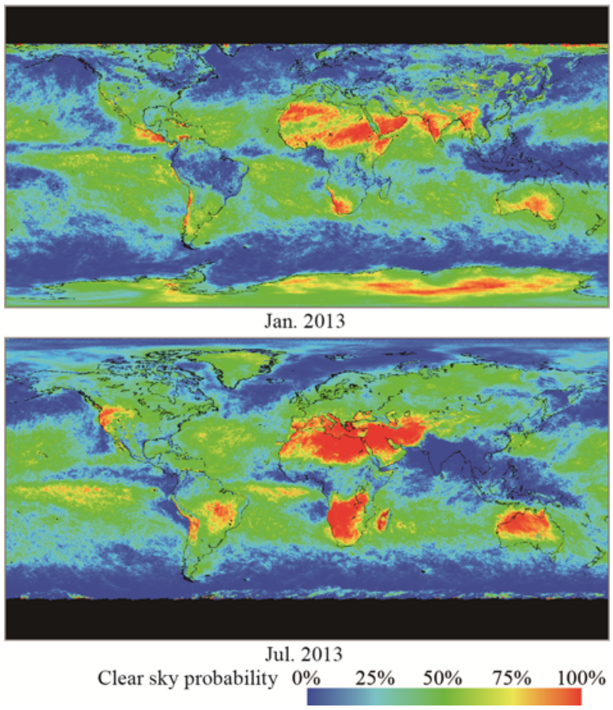

Figure 2: Clear-sky probability at 0.1° × 0.1° calculated with MYD35_L2. There are low clear-sky probabilities over most tropical rainforests because the moisture helps create clouds. (single column)

The accuracy of CLAUDIA1-CAI was evaluated by comparing it with the MODIS/Aqua cloud mask data product (MYD35) (Ackerman et al., 2010) because the MODIS cloud mask algorithm uses a larger number of bands for cloud discrimination than CLAUDIA1-CAI, and CLAUDIA1 was developed based on the MODIS cloud mask algorithm (Taylor et al., 2012; Ishida et al, 2011b). However, these comparisons cannot identify common weak points in the algorithms and another verification method is required. Although the use of visual inspection of clouds as a standard for judging is not practical for screening a full satellite data set, it has the advantage of allowing for locally optimized thresholds, while CLAUDIA1+3 use common global thresholds. Thus, the accuracy of visual inspection is better than that of these algorithms in most regions, with the exception of snow and ice covered surfaces, where there is not enough spectral contrast to distinguish cloud. For this reason visual inspection can be used for the truth metric for the verification exercise. Therefore, the accuracy of CLAUDIA1-CAI also has been evaluated by visual inspection in tropical rain forests (Oishi et al., 2014). In this study, we deal with the application of the CLAUDIA3 to GOSAT CAI data. And then, we compare between CLAUDIA1-CAI and CLAUDIA3-CAI for various land cover types, and evaluate the accuracy by comparing both against visual inspection of the same CAI images in tropical rainforests.

2 Materials and Methods

2.1 Study area and data

The study area for directly comparing CLAUDIA1-CAI and CLAUDIA3-CAI for various land cover types is the same as the previous study (Oishi et al., 2017) (Fig. 3), and for evaluation of the accuracy by comparing both against visual inspection is Borneo and the Amazon (Fig. 4).

The total forest area in the Amazon, Congo, and Southeast Asia rainforest basins is over 13 million km^2, which corresponds to one-third of the total global forest area (FAO and ITTO, 2011). The three most forest-rich countries (Brazil, Democratic Republic of Congo, and Indonesia) account for 57 % of the total global forest area (FAO and ITTO, 2011). However, the total net emissions of carbon from tropical deforestation and land use were estimated to be 1.0 Pg-C/yr in the three rainforest basins (Baccini et al., 2012). In particular, Brazil and Indonesia have by far the highest and second highest deforestation rates, respectively (Fig. 5). Therefore, the study area for rainforests is Borneo and the Amazon (Fig. 4).

GOSAT returns to a similar footprint after 44 orbits (44 CAI paths) in three days. The satellite ground path of one orbit is divided into 60 equidistant CAI frames. We used the GOSAT CAI L1B product, which general users could download from the GOSAT User Interface Gateway (GUIG, https://data.gosat.nies.go.jp), for various land cover types on the beginning of the month from 2012 to 2014 as was done in the previous study (Oishi et al., 2017) (Table 2), and for rainforests (Table 3). Recently the GUIG has been changed to GOSAT Data Archive Service (GDAS, https://data2.gosat.nies.go.jp/index_en.html). The spatial resolution of these products (pixel size at nadir) is 500 m, the image size is 2048 × 1355 pixels (approximately 1000 × 680 km). The CLAUDIA algorithm requires a land/sea mask and surface albedo data. The CAI L1B product includes a land/sea mask with 500 m resolution which is generated from the Shuttle Radar Topography Mission's 15″ land/sea mask and the USGS Global Land 1-KM AVHRR Project mask for areas with latitudes higher than ±60°. Surface albedo data at 1/30° resolution was generated from the CAI L1B data from 10 recurrent cycles by separating the land and water regions.

This processing consists of three steps (Ishihara and Nobuta, 2013): (1) calculate the minimum reflectance to remove cloud-contaminated pixels; (2) cloud shadow correction (Fukuda et al., 2013); and (3) atmospheric correction.

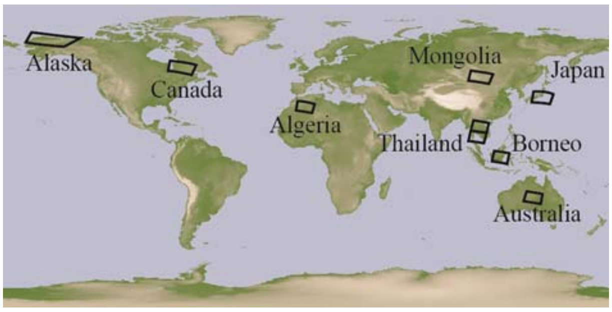

Figure 3: Study areas for various land cover types. Black rectangles indicate the location of CAI frames. (single column)

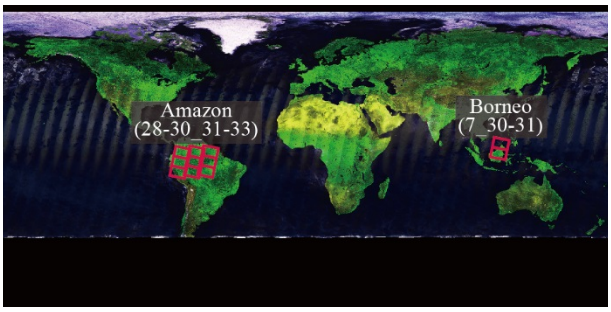

Figure 4: Study areas in Borneo and the Amazon. CAI path and frame system: XX_YY (XX indicates CAI path number and YY indicates CAI frame number). Red rectangles indicate the locations of CAI frames. The background image was generated from the CAI L3 global reflectance distribution product (15 June 2013 to 14 July 2013). (single column)

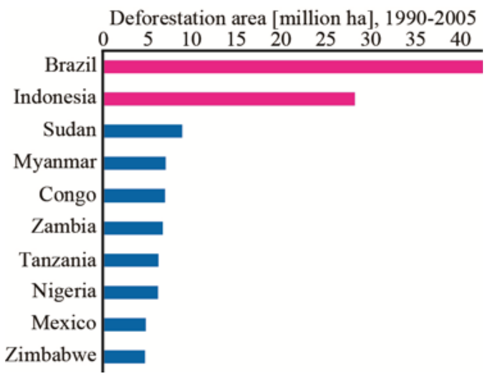

Figure 5: List of top 10 countries for changes in deforestation area (million ha) from 1990 to 2005. These were calculated with data from the Global Forest Resources Assessment 2005 (FAO, 2005). (single column)

Table 2: GOSAT CAI L1B product and CAI L2 cloud flag product used for various land cover types in this study. Land cover was derived from the MODIS land cover type product (MCD12). Japan scenes include urban areas.

Location (CAI Path_Frame)	Data Period	Land Cover
Australia (4_35)	3 April 2012–3 March 2014	Open shrublands
Japan (5_25)	1 April 2012–1 March 2014	Mixed forests
Borneo (7_31)	3 April 2012–3 March 2014	Evergreen broadleaf forest
Thailand 1 (9_28)	2 April 2012–2 March 2014	Cropland/natural vegetation
Thailand 2 (9_29)	2 April 2012–2 March 2014	Cropland/natural vegetation
Mongolia (10_23)	3 April 2012–3 March 2014	Grasslands
Algeria (22_26)	3 April 2012–3 March 2014	Barren or sparsely vegetated
Canada (32_22)	1 April 2012–1 March 2014	Evergreen needleleaf forest
Alaska (43_19)	1 April 2012–1 March 2014	Open shrublands

Table 3: GOSAT CAI L1B product and CAI L2 cloud flag product used for rainforests in this study.

Borneo		Amazon	
Date (yy/mm/dd)	Location (CAI Path_Frame)	Date (yy/mm/dd)	Location (CAI Path_Frame)
10/04/02	7_30	11/08/28	28_31
10/01/02	7_31	11/08/28	28_32
10/04/02	7_31	11/08/28	28_33
10/07/01	7_31	11/08/29	29_31
10/07/07	7_31	10/08/28	29_32
10/07/13	7_31	11/02/03	29_32
10/07/19	7_31	11/04/01	29_32
10/07/28	7_31	11/06/03	29_32
10/09/02	7_31	11/08/02	29_32
10/11/01	7_31	11/08/08	29_32
		11/08/14	29_32
		11/08/23	29_32

11/08/29	29_32
11/10/01	29_32
11/12/03	29_32
11/08/29	29_33
11/08/30	30_31
11/08/30	30_32
11/08/30	30_33

2.2 CLAUDIA1

CLAUDIA1-CAI calculates the clear-sky confidence levels (CCL) for every threshold test and their comprehensive integration (Ishida and Nakajima, 2009). Integrated-CCL of 0 means that the pixel is cloudy and 1 means that the pixel is cloud-free. Ambiguous pixels between cloudy and cloud-free are described by numerical values from 0 to 1. The threshold below which the integrated-CCL counts the pixel as cloud-free for GOSAT FTS L2 is 0.33, otherwise the pixel is regarded as cloudy (Yoshida et al., 2010). The flow of the algorithm is shown in Fig. 6.

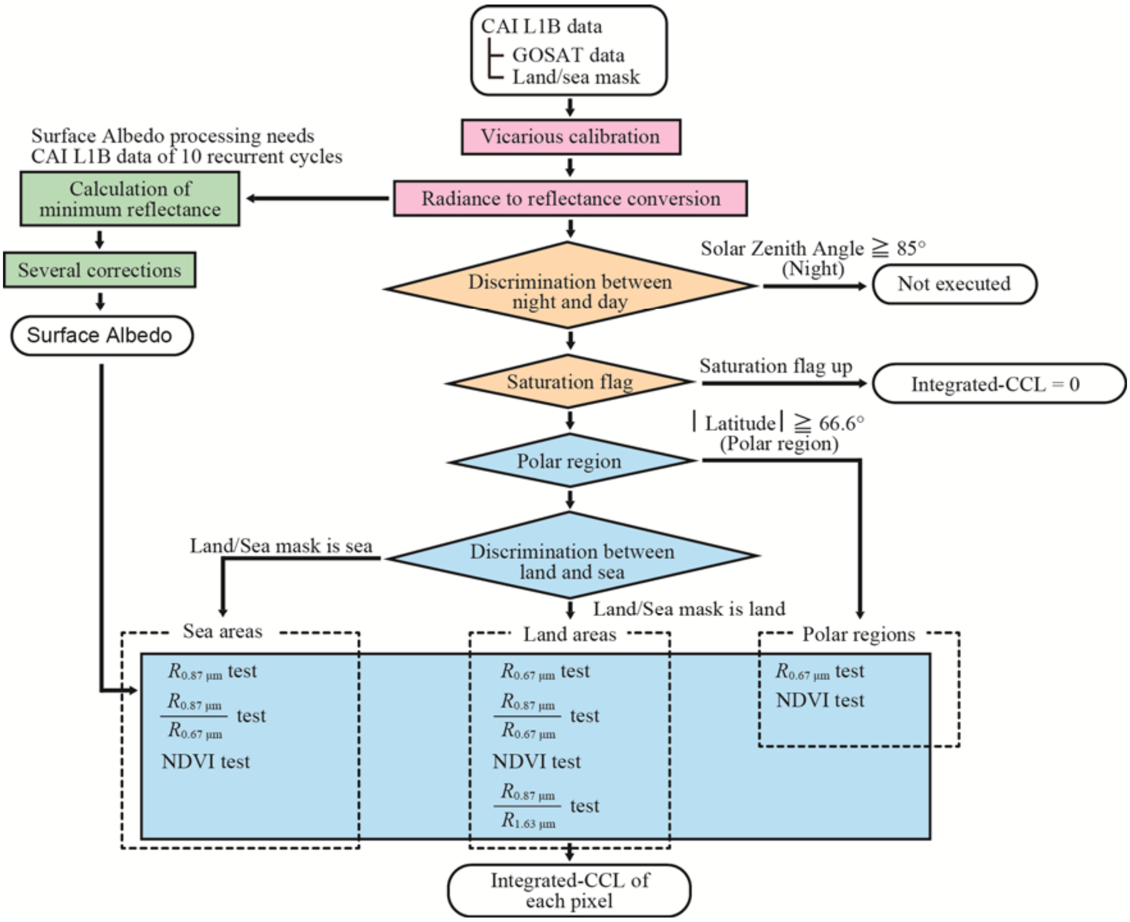

Figure 6: Flow chart for CLAUDIA1-CAI. For sun-glint areas, the thresholds are further increased based on the $R_{0.87\,\mu m}$ test. CCL: confidence level; $R_{wavelength}$: reflectance; NDVI: normalized difference vegetation index. (2 column)

2.3 New cloud discrimination algorithm (CLAUDIA3)

CLAUDIA1 performs cloud discrimination by using thresholds set based on experience. The new cloud discrimination algorithm (CLAUDIA3, Ishida et al., 2018) uses SVM to decide the thresholds objectively by using multivariate analysis. SVM is a supervised pattern recognition method. First, it determines the following items using training samples of typical clear and cloudy pixels: 1) a decision function to discriminate between two classifications (clear and cloudy), 2) the thresholds, and 3) the support vectors, which are training samples specified by the decision function. The support vectors are decided in a high-dimensional feature space of the training samples. Next, it performs cloud discrimination by using the decision function, thresholds, and support vectors it determined. CLAUDIA3 applies the kernel trick (Boser et al., 1992) to soft-margin SVM (Cortes and Vapnik, 1995). The kernel uses a second-order polynomial (Eq. (1))

$$K(x_i, x) = \frac{(\mathbf{x}_i \bullet \mathbf{x} + 1)^2}{2}, \quad (1)$$

where K is the kernel function, $\mathbf{x}_i$ is the support vectors, and $\mathbf{x}$ is input data. The flow of CLAUDIA3-CAI is explained in Fig. 7. For CLAUDIA3-CAI, an integrated-CCL of 0.5 corresponds to the separating hyperplane of clear support vectors and cloudy support vectors. In this study, we used two kinds of support vector: (1) support vectors generated by using MODIS data in February for cloud discrimination between November and April, and (2) support vectors generated by using MODIS data in August for cloud discrimination between May and October based on a previous study (Oishi et al., 2017).

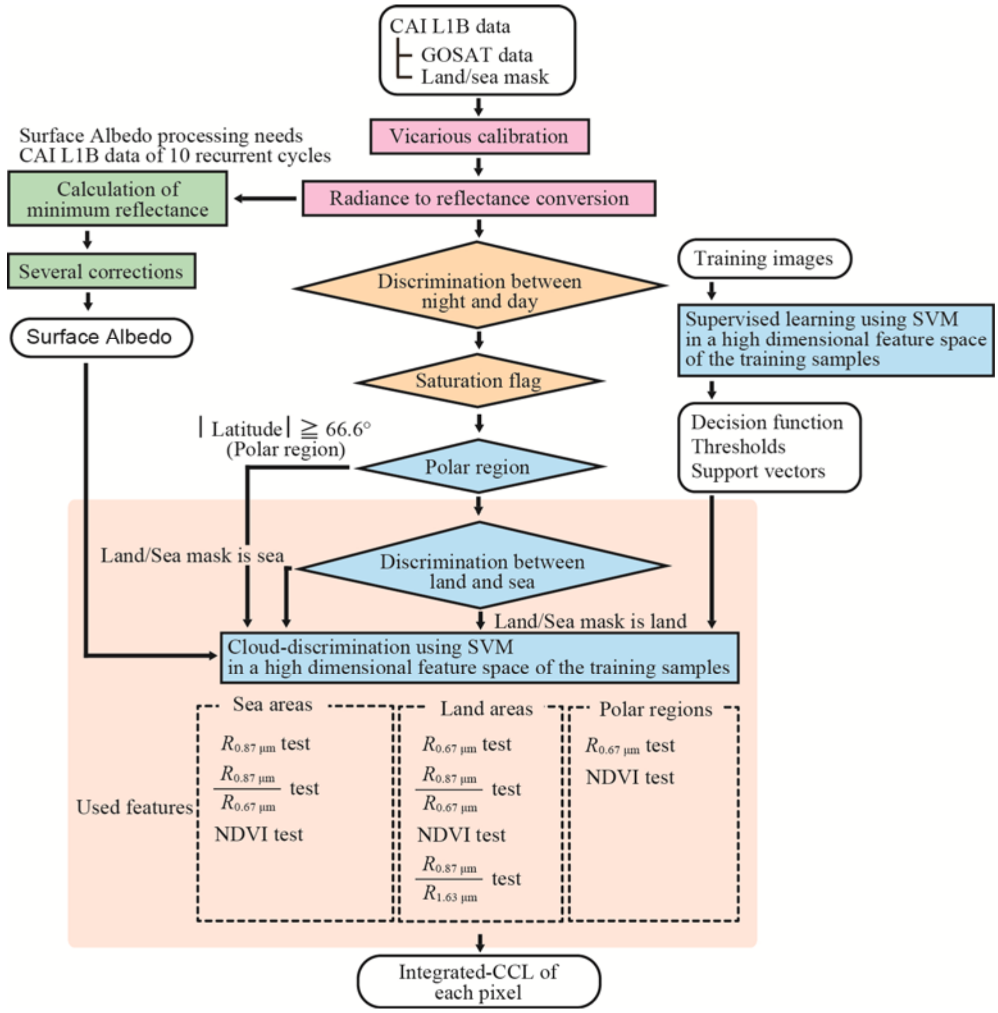

Figure 7: Flow chart for CLAUDIA3-CAI. CCL: clear-sky confidence level; Rwavelength: reflectance; NDVI: normalized difference vegetation index. (2 column)

2.4 Analysis procedure for rainforests

The analysis procedure consists of the following steps (Fig. 8).

1) Cut 400 × 400 pixels around the centre of CAI L1B images.

2) Perform visual inspection of the pixels cut from the CAI L1B images.

We performed a visual inspection of the presence or absence of clouds in every pixel.

3) Perform cloud discrimination by using CLAUDIA1-CAI and CLAUDIA3-CAI.

For CLAUDIA1-CAI, we produced output images setting the integrated-CCL threshold to 0.33. For CLAUDIA3-CAI, we produced output images setting the integrated-CCL threshold to 0.5.

4) Compare output with visual inspection.

We coloured the images by comparing the visual inspection images with the output images pixel-by-pixel.

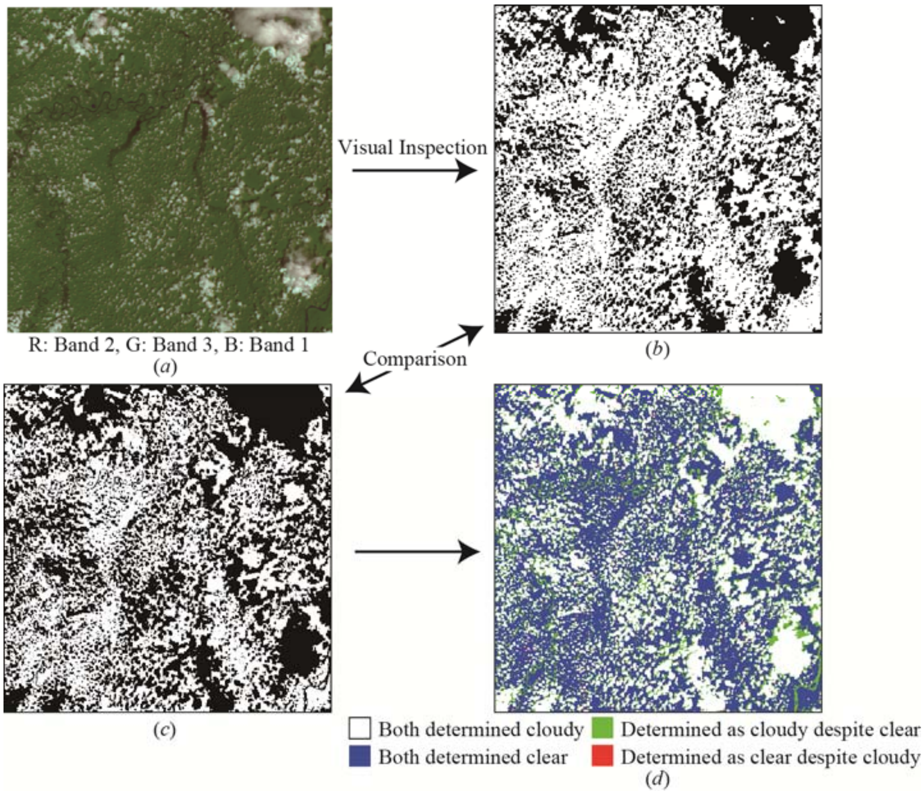

Figure 8: Analysis procedure. (a) CAI L1B image. (b) Visual inspection image of CAI L1B. (c) Output image from CLAUDIA1-CAI (CAI L2 cloud flag product) or CLAUDIA3-CAI. Pixels that are determined as cloudy are black. (d) Comparison of the visual inspection image and the output image. Pixels that are determined as cloudy in both are white. Pixels that are determined as clear in both are blue. Pixels that are determined as cloudy in the output image

and clear in the visual inspection image are green. Unusual pixels that are determined as clear in the output image and cloudy in the visual inspection image are red. (2 column)

3 Results

In this study, "accuracy" is defined as the ratio of the number of pixels for which the standard image and output from the cloud discrimination algorithm agree to the total number of pixels in the input image. "Overlook" is defined as the ratio of the number of pixels judged clear in the output and cloudy in the standard image to the number of pixels that were judged cloudy in the standard image. "Overestimate" is defined as the ratio of the number of pixels judged cloudy in the output and clear in the standard image to the number of pixels judged clear in the standard image. These definitions are written as follows.

$$\text{Accuracy} = \frac{\text{Both cloudy} + \text{Both clear}}{\text{Total number of pixels}}, \quad (2)$$

$$\text{Overlook} = \frac{\text{Clear despite cloudy}}{\text{Both cloudy} + \text{Clear despite cloudy}}, \quad (3)$$

$$\text{Overestimate} = \frac{\text{Cloudy despite clear}}{\text{Both clear} + \text{Cloudy despite clear}}. \quad (4)$$

3.1 Results for various land cover types

Figure 9 shows the monthly average accuracy, overlook, and overestimate for an integrated-CCL threshold of 0.33 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI. We used the CLAUDIA1-CAI result as the standard image.

In Australia and Algeria, Overlook was greater than Overestimate; the average of Overlook was 44.2 % (the lowest Overlook was 20.5 % in December in Australia), against the average of Overestimate was 0.4 % (the highest Overestimate was 3.5 % in January in Algeria). These mean that there was tendency that CLAUDIA3-CAI judged clear, despite CLAUDIA1-CAI judged cloudy in Australia and Algeria.

In Japan, Borneo, Canada, and Alaska, Overestimate was greater than Overlook; the average of Overlook was 1.6 % (the highest Overlook was 3.2 % in August), against the average of Overestimate was 13.7 % (the lowest Overestimate was 7.2 % in May) in Japan; the average of Overlook was 1.0 % (the highest Overlook was 2.3 % in July), against the average of Overestimate was 39.2 % (the lowest Overestimate was 24.2 % in April) in Borneo; the average of Overlook was 2.9 % (the highest Overlook was 8.8 % in July), against the average of Overestimate was 51.8 % (the lowest Overestimate was 23.2 % in July) in Canada; the average of Overlook was 11.9 % (the highest Overlook was 27.5 % in August), against the average of Overestimate was 50.3 % (the lowest Overestimate was 20.3 % in July) in Alaska. These mean that there was tendency that CLAUDIA3-CAI judged cloudy, despite CLAUIDA1-CAI judged clear in Japan, Borneo, Canada, and Alaska.

In Thailand and Mongolia, there was seasonal variation. In Thailand, Overlook was greater than Overestimate from March to May, and Overestimate was greater than Overlook from June to February; the average of Overlook was 12.7% (the lowest Overlook was 9.2% in May), against the average of Overestimate was 6.3% (the highest Overestimate was 7.7% in April) from March to May; the average of Overlook was 2.2% (the highest Overlook was 7.1% in February), against the average of Overestimate was 25.8% (the lowest Overestimate was 10.0% in January) from June to February. In Mongolia, Overestimate was greater than Overlook from February to March, and Overlook was greater than Overestimate from April to January; the average of Overlook was 4.0% (the highest Overlook was 4.1% in March), against the average of Overestimate was 40.1% (the lowest Overestimate was 37.8% in March) from February to March; the average of Overlook was 20.4% (the lowest Overlook was 11.9% in July and August), against the average of Overestimate was 6.1% (the highest Overestimate was 14.5% in December) from April to January.

Figure 10 compares the output images of CLAUDIA1-CAI and CLAUDIA3-CAI for select cases in each region.

In Australia and Algeria, CLAUDIA3-CAI could identify bright surface, however, there were a few oversights of the edges of clouds.

In Japan, CLAUDIA3-CAI misjudged vegetation areas as clouds.

In Borneo, CLAUDIA3-CAI could identify optically thin clouds.

In Canada and Alaska, they were snow or ice covered scenes. Since the CAI is not equipped with any thermal infrared bands, cloud discrimination based on the temperature at the top of clouds is not feasible. Accordingly, it is difficult to discriminate between ice or snow and clouds. The difference or coincidence between CLAUDIA1-CAI and CLAUIDA3-CAI was attributed to this source of error.

In Thailand, CLAUDIA3-CAI could judge smokes as non-clouds, despite CLAUDIA1-CAI judged clouds, however, there were oversights of optically thin clouds and the edges of clouds on 3 April 2013. Furthermore CLAUDIA3-CAI misjudged clear muddy rivers and boundaries between land and water as cloudy. This was also reported about CLAUDIA1-CAI in previous study (Oishi et al. 2014). Conversely, CLAUDIA3-CAI could identify optically thin clouds on 2 September 2012.

In Mongolia, it was snow covered scene on 3 February 2013 in the same as Canada and Alaska. On the other hand CLAUDIA3-CAI could identify bright surface, however, there were a few oversights of the edges of clouds on 2 June 2012.

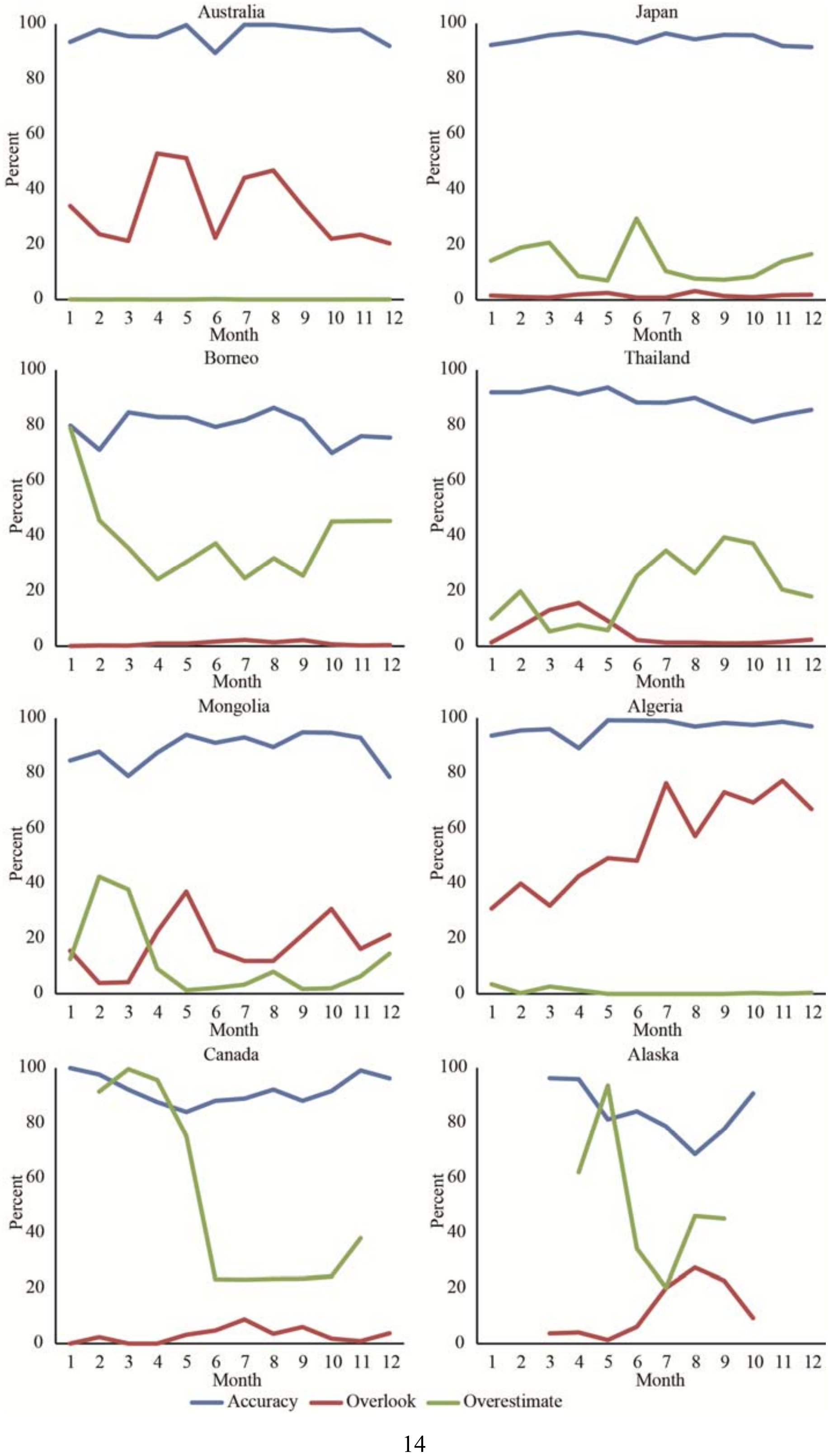
Australia
Japan
Borneo
Thailand
Mongolia
Algeria
Canada
Alaska
Percent
Month
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12
Accuracy
Overlook
Overestimate

Figure 9: Monthly average accuracy, overlook, and overestimate for various land cover types. Blue line indicates "Accuracy", red line indicates "Overlook, and green line indicates "Overestimate". (2 column)

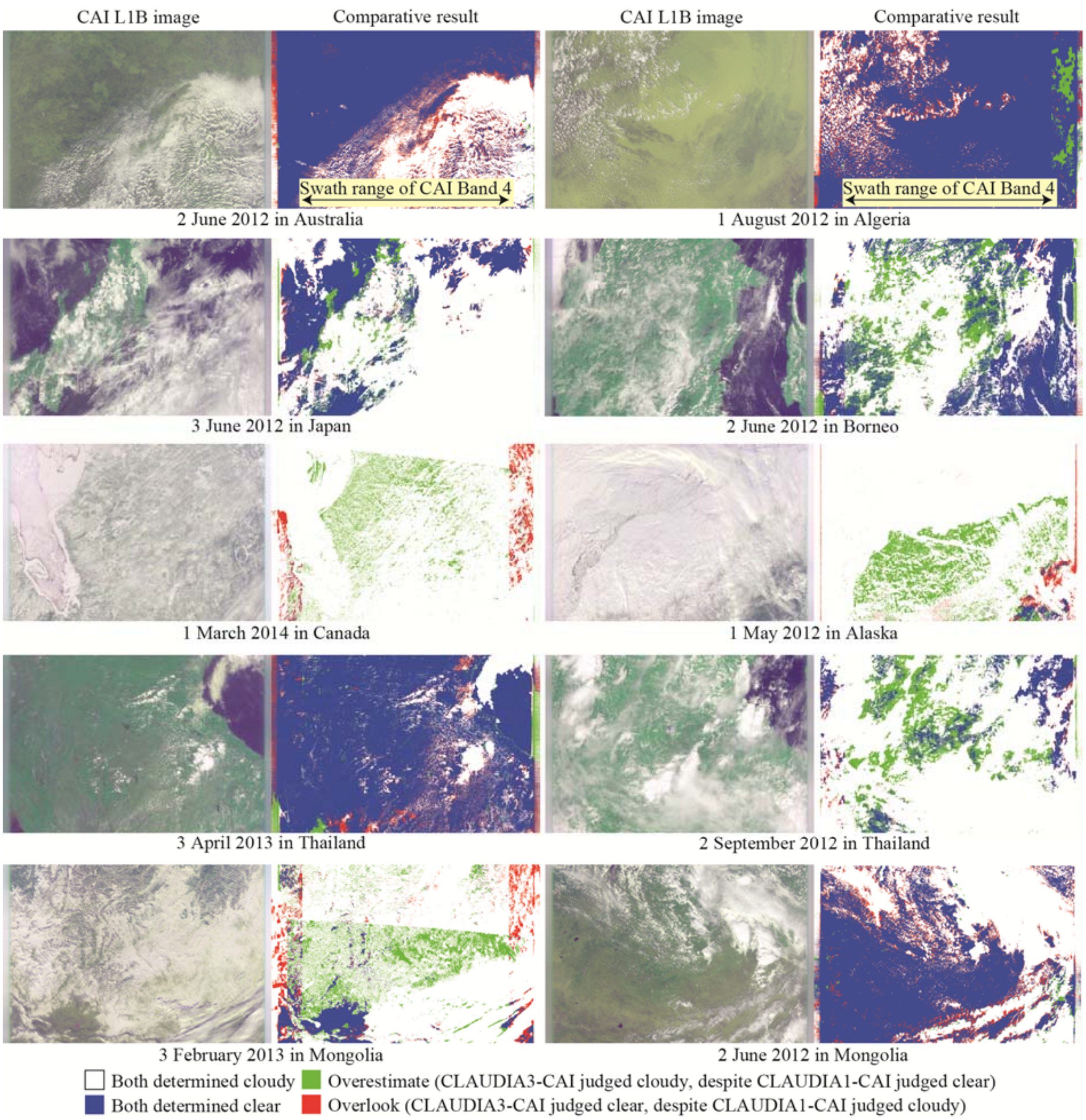

Figure 10: CAI L1B images (R: Band 2, G: Band 3, B: Band 1) and comparative results of CLAUDIA1-CAI and CLAUDIA3-CAI for various land cover types. (2 column)

3.2 Results in the Amazon

Figure 11 compares the visual inspection images and the output images for four select cases in the Amazon: low cloud cover, high cloud cover, small scattered clouds, and optically thin clouds. We used the visual inspection result as the standard image.

CLAUDIA3-CAI produced fewer overlooked clouds but slightly more overestimated clouds than CLAUDIA1-CAI did. CLAUDIA3-CAI misjudged clear muddy rivers on 23 August 2011 in CAI Path 29, Frame 32 and the surroundings of clouds on 1 April 2011 in CAI Path 29, Frame 32. The maximum accuracy values of CLAUDIA3-CAI and the CLAUDIA1-CAI occur at different integrated-CCL values with the thresholds for the Amazon. Fig. 12 shows the average accuracy, overlook, and overestimate of all the data in the Amazon for all 19 cases. These results indicate that the most suitable integrated-CCL thresholds are 0.75 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI in the Amazon. Since curved lines of overestimate and overlook intersect, CLAUDIA3-CAI can appropriately determine the boundary between cloud and clear-sky.

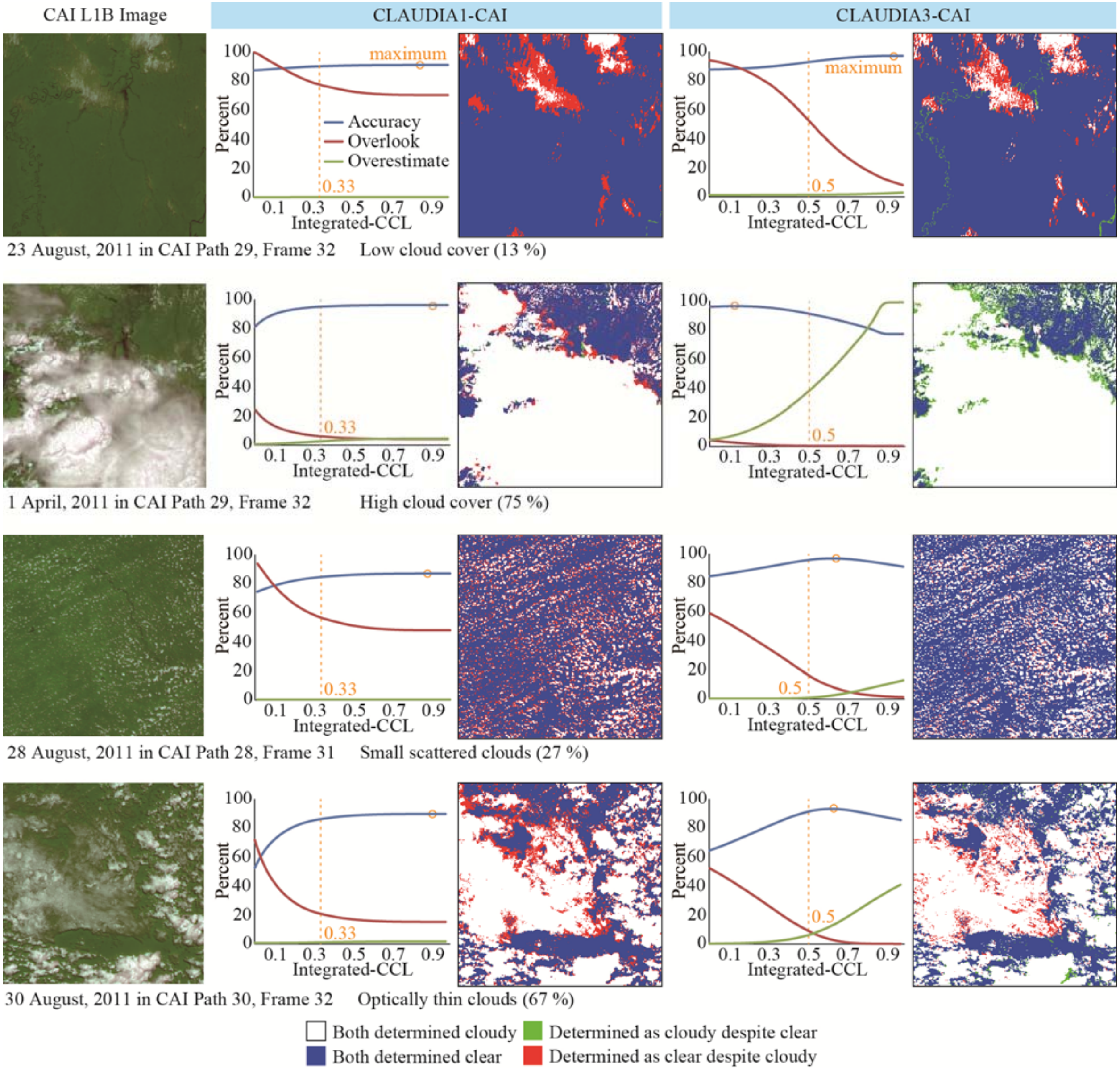

Figure 11: Comparison of the visual inspection images and the output images in the Amazon. Orange circles indicate the maximum accuracy values. Orange dotted lines indicate the integrated-CCL thresholds. Blue line indicates "Accuracy", red line indicates "Overlook, and green line indicates "Overestimate". (2 column)

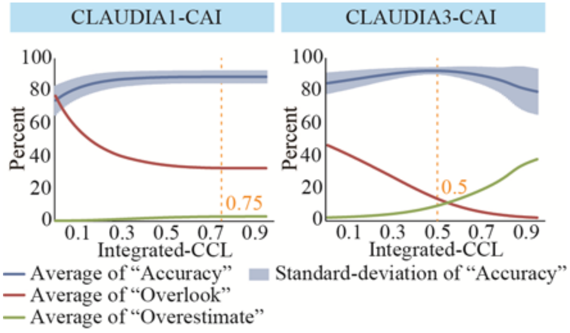

Figure 12: Average accuracy, overlook, and overestimate for all data for the Amazon. The most suitable integrated-CCL thresholds are 0.75 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI in the Amazon. (single column)

Table 4 shows the results for an integrated-CCL threshold of 0.33 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI, and Table 5 shows the results for an integrated-CCL threshold of the maximum accuracy values in Fig. 12 (CLAUDIA1-CAI: 0.75, CLAUDIA3-CAI: 0.5). There was no notable change in the accuracies with the season or location. When the integrated-CCL threshold was 0.33 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI, the accuracies were 87.0 % and 92.0 %, respectively. When the accuracy of CLAUDIA1-CAI was higher than that of CLAUDIA3-CAI, optically thick clouds covered a wide area of the input images. Furthermore, when the integrated-CCL threshold was 0.75 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI, the accuracy was the highest, at 88.3 % and 92.0 %, respectively. In the both cases, the accuracy of CLAUDIA3-CAI was higher than that of CLAUDIA1-CAI.

Table 4: Results for integrated-CCL thresholds of 0.33 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI in the Amazon.

		Accuracy (%)		Overlook (%)		Overestimate (%)	
Date (yy/mm/dd)	Location (CAI Path_Frame)	CLAUDIA1 (0.33)	CLAUDIA3 (0.5)	CLAUDIA1 (0.33)	CLAUDIA3 (0.5)	CLAUDIA1 (0.33)	CLAUDIA3 (0.5)
11/08/28	28_31	84.6	95.1	56.6	16.9	0.0	0.5
11/08/28	28_32	80.6	92.9	49.7	7.5	0.1	6.9
11/08/28	28_33	92.0	95.9	11.6	13.4	7.4	2.4
11/08/29	29_31	87.6	93.8	27.2	9.5	0.3	3.5
10/08/28	29_32	89.8	90.8	32.6	9.9	1.7	9.0
11/02/03	29_32	86.6	92.9	35.5	2.4	0.5	9.9

11/04/01	29_32	95.0	91.6	5.8	0.1	2.1	36.6
11/06/03	29_32	89.9	90.2	38.1	4.1	0.8	11.7
11/08/02	29_32	77.9	90.6	71.0	27.3	0.1	1.5
11/08/08	29_32	84.5	92.9	66.0	26.3	0.1	1.2
11/08/14	29_32	87.8	93.2	77.4	36.0	0.1	1.4
11/08/23	29_32	90.0	92.2	77.8	54.0	0.1	1.0
11/08/29	29_32	79.6	91.0	52.4	19.7	0.1	2.2
11/10/01	29_32	87.1	92.2	33.9	5.5	0.1	9.1
11/12/03	29_32	82.8	93.4	30.7	1.7	0.1	12.9
11/08/29	29_33	90.6	90.8	20.8	15.1	2.3	5.6
11/08/30	30_31	85.7	85.1	24.7	9.2	3.2	21.0
11/08/30	30_32	86.0	91.4	20.9	10.2	0.4	5.5
11/08/30	30_33	94.9	93.0	11.1	3.6	1.5	9.1
Average		87.0	92.0	39.1	14.3	1.1	7.9

Table 5: Results for integrated-CCL thresholds of the maximum accuracy values in Fig. 11 (CLAUDIA1-CAI: 0.75, CLAUDIA3-CAI: 0.5) in the Amazon.

		Accuracy (%)		Overlook (%)		Overestimate (%)	
Date	Location	CLAUDIA1	CLAUDIA3	CLAUDIA1	CLAUDIA3	CLAUDIA1	CLAUDIA3
(yy/mm/dd)	(CAI Path_Frame)	(0.75)	(0.5)	(0.75)	(0.5)	(0.75)	(0.5)
11/08/28	28_31	86.9	95.1	47.9	16.9	0.0	0.5
11/08/28	28_32	84.2	92.9	40.2	7.5	0.2	6.9
11/08/28	28_33	83.6	95.9	7.1	13.4	18.1	2.4
11/08/29	29_31	89.6	93.8	21.8	9.5	1.2	3.5
10/08/28	29_32	90.6	90.8	23.5	9.9	4.0	9.0
11/02/03	29_32	88.9	92.9	27.8	2.4	1.4	9.9
11/04/01	29_32	96.2	91.6	3.7	0.1	4.1	36.6
11/06/03	29_32	90.9	90.2	29.3	4.1	2.4	11.7

11/08/02	29_32	80.1	90.6	63.6	27.3	0.3	1.5
11/08/08	29_32	85.9	92.9	59.4	26.3	0.2	1.2
11/08/14	29_32	88.8	93.2	70.1	36.0	0.2	1.4
11/08/23	29_32	90.9	92.2	70.3	54.0	0.1	1.0
11/08/29	29_32	82.2	91.0	45.5	19.7	0.2	2.2
11/10/01	29_32	89.7	92.2	26.6	5.5	0.4	9.1
11/12/03	29_32	86.7	93.4	23.3	1.7	0.5	12.9
11/08/29	29_33	90.9	90.8	13.5	15.1	6.4	5.6
11/08/30	30_31	87.1	85.1	20.4	9.2	4.9	21.0
11/08/30	30_32	89.9	91.4	14.7	10.2	1.0	5.5
11/08/30	30_33	95.1	93.0	7.0	3.6	3.6	9.1
Average		88.3	92.0	32.4	14.3	2.6	7.9

3.3 Results in Borneo

Figure 13 compares the results of the visual inspection images and the output images for two select cases in Borneo: small scattered clouds and optically thin clouds. We used the visual inspection result as the standard image. The comparison of the results for Borneo is similar to that for the Amazon. Figure 14 shows the average accuracy, overlook, and overestimate of all data for all cases in Borneo. These results indicate that the most suitable integrated-CCL thresholds are 0.85 for the CLAUDIA1-CAI and 0.35 for CLAUDIA3-CAI in Borneo. Since curved lines of overestimate and overlook intersect as same as the Amazon cases, CLAUDIA3-CAI can appropriately determine the boundary between cloud and clear-sky.

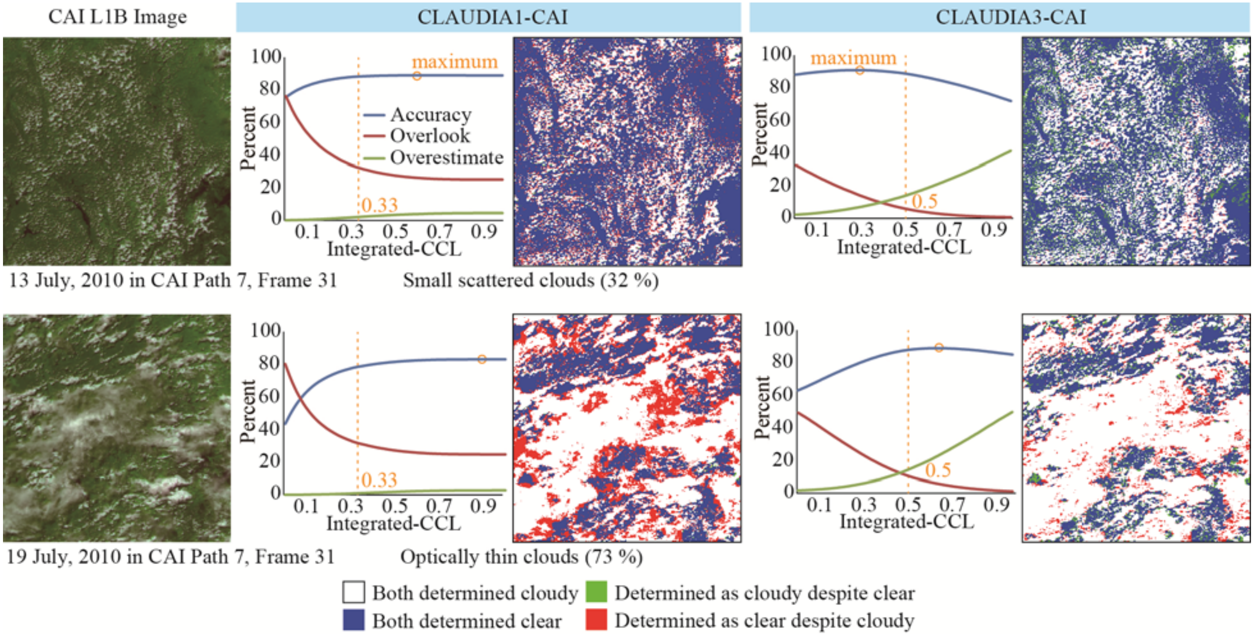

Figure 13: Comparison of the visual inspection images and the output images in Borneo. Orange circles indicate the maximum accuracy values. Orange dotted lines indicate the integrated-CCL thresholds. Blue line indicates "Accuracy", red line indicates "Overlook, and green line indicates "Overestimate". (2 column)

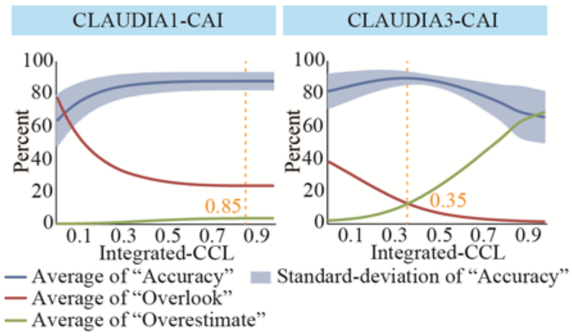

Figure 14: Average accuracy, overlook, and overestimate for all data for Borneo. The most suitable integrated-CCL thresholds are 0.85 for CLAUDIA1-CAI and 0.35 for CLAUDIA3-CAI in Borneo. (single column)

Table 6 shows the results for an integrated-CCL threshold of 0.33 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI, and Table 7 shows the results for an integrated-CCL threshold of the maximum accuracy values in Fig. 14 (CLAUDIA1-CAI: 0.85, CLAUDIA3-CAI: 0.35). There was no notable change in the accuracies with the season or location, similar to the results for the Amazon. For an integrated-CCL threshold of 0.33 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI, the accuracy was 84.8 % and 86.9 %, respectively. Furthermore, for an integrated-CCL threshold of 0.85 for CLAUDIA1-CAI and 0.35 for CLAUDIA3-CAI, the highest accuracies of 87.5 % and 88.8 %, respectively, were obtained. In both cases, the accuracy of CLAUDIA3-CAI was greater than that of CLAUDIA1-CAI.

Table 6: Results for integrated-CCL thresholds of 0.33 for CLAUDIA1-CAI and 0.5 for CLAUDIA3-CAI in Borneo.

		Accuracy (%)		Overlook (%)		Overestimate (%)	
Date (yy/mm/dd)	Location (CAI Path_Frame)	CLAUDIA1 (0.33)	CLAUDIA3 (0.5)	CLAUDIA1 (0.33)	CLAUDIA3 (0.5)	CLAUDIA1 (0.33)	CLAUDIA3 (0.5)
10/04/02	7_30	89.7	91.7	28.8	1.7	0.1	12.0
10/01/02	7_31	85.6	85.0	25.8	1.8	0.6	31.1
10/04/02	7_31	94.8	85.4	8.3	0.6	3.5	22.8
10/07/01	7_31	90.8	92.2	29.0	5.0	0.4	9.0
10/07/07	7_31	76.5	85.9	54.2	22.5	0.5	7.8
10/07/13	7_31	88.2	89.1	32.6	5.8	2.0	13.3
10/07/19	7_31	77.1	88.4	31.1	11.0	1.0	13.5
10/07/28	7_31	70.6	81.5	44.8	8.2	1.1	37.5
10/09/02	7_31	89.3	87.8	37.8	6.5	1.3	14.2
10/11/01	7_31	85.8	81.8	20.6	0.4	1.2	54.7
Average		84.8	86.9	31.3	6.3	1.2	21.6

Table 7: Results for integrated-CCL thresholds of the maximum accuracy values in Fig. 13 (CLAUDIA1-CAI: 0.85, CLAUDIA3-CAI: 0.35) in Borneo.

		Accuracy (%)		Overlook (%)		Overestimate (%)	
Date (yy/mm/dd)	Location (CAI Path_Frame)	CLAUDIA1 (0.85)	CLAUDIA3 (0.35)	CLAUDIA1 (0.85)	CLAUDIA3 (0.35)	CLAUDIA1 (0.85)	CLAUDIA3 (0.35)
10/04/02	7_30	91.9	94.6	22.3	8.5	0.3	3.8
10/01/02	7_31	89.2	90.7	16.8	8.0	3.6	10.9
10/04/02	7_31	93.8	91.5	4.6	2.3	7.2	12.2
10/07/01	7_31	92.1	93.2	21.5	10.3	1.9	5.3

10/07/07	7_31	79.4	83.5	46.1	33.0	1.6	4.2
10/07/13	7_31	88.9	90.9	25.1	11.4	4.4	7.9
10/07/19	7_31	81.7	83.4	24.1	20.1	2.7	7.1
10/07/28	7_31	77.3	80.7	33.2	18.9	3.2	20.0
10/09/02	7_31	90.3	90.6	29.0	12.3	3.0	8.3
10/11/01	7_31	90.8	89.4	10.9	3.3	5.8	25.5
Average		87.5	88.8	23.4	12.8	3.4	10.5

4 Discussions and conclusions

Comparative results between CLAUDIA1-CAI and CLAUDIA3-CAI for various land cover types indicated that CLAUDIA3-CAI had tendency to identify bright surface and optically thin clouds, however, misjudge the edges of clouds as compared with CLAUDIA1-CAI. There are tradeoffs in maximizing accuracy while minimizing overlook and overestimate. Thus, it is sufficient to change the integrated-CCL threshold according to the purpose. Furthermore, CLAUDIA3-CAI misjudged vegetation areas as clouds in Japan. It is necessary to add clear training data of Japanese vegetation areas for CLAUDIA3.

The averaged accuracy of CLAUDIA3 used with GOSAT CAI data (CLAUDIA3-CAI) was approximately 89.5 % in tropical rainforests, which was greater than that of CLAUDIA1-CAI (85.9 %) for the test cases presented here. This is mainly because, in contrast to CLAUDIA1-CAI, CLAUDIA3-CAI can detect optically thin clouds and the edges of clouds, which prevents cloud-contaminated FTS-2 data from being processed as cloud-free FTS-2 data in the greenhouse gas concentration calculations. However, CLAUDIA3-CAI tends to overestimate the surroundings of clouds, which are judged to be cloudy despite being clear. Thus, CLAUDIA3-CAI is not expected to increase the amount of the FTS-2 data that can be used to estimate greenhouse gas concentrations in tropical rainforests. Conversely, CLAUDIA3-CAI may be able to detect optically thin clouds that cannot be detected by visual inspection.

CLAUDIA3-CAI misjudged clear muddy rivers and boundaries between land and water as cloudy in the same manner as CLAUDIA1-CAI. This has three possible causes: (1) insufficient training data for muddy rivers to distinguish the differences in the spectral reflectance properties of muddy water and other water; (2) deviation of the positions in each CAI band owing to the band-to-band registration error; and (3) insufficient resolution of the surface albedo data. The surface albedo data was generated at 1/8° resolution by separating the land and water region. If the border pixels between land and water regions were mixed pixels, the albedo data of 1/8° areas that include the mixed pixels would be included. To decrease this effect, higher resolution surface albedo data are needed. For boundaries between land and water, the resolution of surface albedo data is being investigated because it may be the main problem; the misjudged regions and grid pattern of albedo data match. CLAUDIA3-CAI is more sensitive to differences between land and water than CLAUDIA1-CAI because there is a large difference in the structure of support vectors between land and water. However, generating higher resolution surface albedo data from CAI L1B data for 10 recurrent cycles cannot completely remove clouds in the minimum reflectance calculation. To solve this, initially we need to confirm whether 500 m resolution albedo data should be used. If necessary, we will develop a new method for generating surface albedo data. For example, simple cloud discrimination could be added to calculate the minimum reflectance, and if it is a cloud-contaminated pixel then the pixel is replaced by a minimum reflectance pixel, which is calculated from the same month in several years.

Although we used MODIS data as training images to generate support vectors in this study, the MODIS data and CAI data depend on observation conditions. In future work, we will use CAI data as training images to perform cloud discrimination for CAI data. Furthermore, we will verify CLAUDIA3-CAI by using global CAI data with an alternative method. For instance, comparison with satellite LiDAR data, such as CALIPSO, because it is impossible to perform visual inspection of global data and visual inspection is also not itself perfect. Addressing these points will make CLAUDIA3-CAI more reliable for GOSAT-2 CAI-2 cloud discrimination.

Author Contributions. Yu Oishi, Haruma Ishida, Takashi Y. Nakajima, Ryosuke Nakamura, and Tsuneo Matsunaga conceived and designed the studies; Yu Oishi performed evaluations and analyzed the data; Haruma Ishida contributed analysis tools; Yu Oishi wrote the paper.

Competing interests. The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

Acknowledgements. This research is supported by the GOSAT-2 Project at the National Institute for Environmental Studies (NIES), Japan (2015, 2016); and the New Energy and Industrial Technology Development Organization (NEDO). The authors would like to thank the GOSAT Project, GOSAT-2 Project, and Dr. T. Endo for their helpful comments; Mr. T. Hirose for his assistance with visual inspection. We appreciate an anonymous reviewer who gave useful comments to the manuscript of the previous version.

References

Ackerman, S., Frey, R., Strabala, K., Liu, Y., Gumley, L., Baum, B., and Menzel, P.: Discriminating clear-sky from cloud with MODIS algorithm theoretical basis document (MOD35), Available at: http://modis-atmos.gsfc.nasa.gov/_docs/MOD35_ATBD_Collection6.pdf (last access: 8 December 2017), 2010.

Baccini, A., Goetz, S.J., Walker, W.S., Laporte, N.T., Sun, M., Sulla-Menache, D., Hackler, J., Beck, P.S.A., Dubayah, R., Friedl, M.A., Samanta, S., and Houghton, R.A.: Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps, Nature Clim. Change, 2, 182–185, doi:10.1038/nclimate1354, 2012.

Boser, B., Guyon, I., and Vapnik, V.: A training algorithm for optimal margin classifiers, COLT '92 Proc. 5th Worksh. on Computat. Learning Theory, 144–152, doi:10.1145/130385.130401, 1992.

Cortes, C. and Vapnik, V.: Support-vector networks, Mach. Learn, 20, 273–297, doi:10.1023/A:1022627411411, 1995.

FAO: Global Forest Resources Assessment 2005, Available at: http://www.fao.org/docrep/008/a0400e/a0400e00.htm (last access: 8 December 2017), 2005.

FAO and ITTO: The state of forests in the Amazon Basin, Congo Basin and Southeast Asia, available at: www.fao.org/docrep/014/i2247e/i2247e00.pdf (last access: 8 December 2017), 2011.

Fukuda, S., Nakajima, T., Takenaka, H., Higurashi, A., Kikuchi, N., Nakajima, T.Y., and Ishida, H.: New approaches to removing cloud shadows and evaluating the 380 nm surface reflectance for improved aerosol optical thickness retrievals from the GOSAT/TANSO-Cloud and Aerosol Imager, J. Geophys. Res., 118, 13520–13531, doi: 10.1002/2013JD020090 2013.

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

Ishida, H., Nakajima, T.Y., and Kikuchi, N.: Algorithm Theoretical Basis Document for GOSAT TANSO-CAI L2 cloud flag, Available at: https://data2.gosat.nies.go.jp/GosatDataArchiveService/doc/GU/ATBD_CAIL2CLDFLAG_V1.0_en.pdf (last access: 8 December 2017), 2011a.

Ishida, H., Nakajima, T.Y., Yokota, T., Kikuchi, N., and Watanabe, H.: Investigation of GOSAT TANSO-CAI cloud screening ability through an intersatellite comparison, J. Appl. Meteorol. Climatol., 50, 1571–1586, doi: 10.1175/2011JAMC2672.1, 2011b.

Ishihara, H. and Nobuta, K.: Algorithm Theoretical Basis Document (ATBD) on the processing of GOSAT TANSO-CAI L3 Global Reflectance Products, Available at: https://data2.gosat.nies.go.jp/GosatDataArchiveService/doc/GU/ATBD_CAIL3REF_V1.0_en.pdf (last access: 8 December 2017), 2013.

Ishida, H., Oishi, Y., Morita, K., Moriwaki, K., and Nakajima, T.Y.: Development of a support vector machine based cloud detection method for MODIS with the adjustability to various conditions, Rem. Sen. Env., 205, 390–407, doi: 10.1016/j.rse.2017.11.003, 2018.

Mather J. H. and Voyles J. W.: The ARM Climate Research Facility: A review of structure and capabilities, B. AM. Meteorol. Soc., 94, 3, 377–392, doi: 10.1175/BAMS-D-11-00218.1, 2013.

Maksyutov, S., Takagi, H., Valsala, V. K., Saito, M. Oda, T., Saeki T., Belikov, D.A., Saito, T., Ito, A., Yoshida, Y., Morino, I., Uchino, O., Andres, R.J., and Yokota, T.: Regional CO2 flux estimates for 2009–2010 based on GOSAT and ground-based CO2 observations, Atmos. Chem. Phys., 13, 9351–9373, doi: 10.5194/acp-13-9351-2013, 2013.

Maksyutov, S., Takagi, H., Belikov, D.A., Saito, M., Oda, T., Saeki, T., Valsala, V. K., Saito, R., Ito, A., Yoshida, Y., Morino, I., Uchino, O., and Yokota, T.: Algorithm Theoretical Basis Document (ATBD) for the estimation of CO2 fluxes and concentration distributions from GOSAT and surface-based CO2 data, Available at: https://data2.gosat.nies.go.jp/GosatDataArchiveService/doc/GU/ATBD_L4CO2_V1.0_en.pdf (last access: 8 December 2017), 2014.

Ministry of the Environment, Japan: New mechanisms information platform, Joint Crediting Mechanism (JCM), Available at: https://www.carbon-markets.go.jp/eng/jcm/index.html (last access: 8 December 2017), 2015.

NIES GOSAT-2 Project: GOSAT-2 Project at the National Institute for Environmental Studies, about GOSAT-2, Available at: www.gosat-2.nies.go.jp (last access: 8 December 2017), 2014.

Oishi, Y., Kamei, A., Yokota, Y., Hiraki, K., and Matsunaga, T.: 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, 3, 153–165, doi: 10.11440/rssj.34.153, 2014 (in Japanese with English Abstract).

Oishi, Y., Ishida, H., Nakajima, T. Y., Nakamura, R., and Matsunaga, T.: The impact of different support vectors on GOSAT-2 CAI-2 L2 cloud discrimination, Remote Sens., 9, 1236, doi:10.3390/rs9121236, 2017.

Taylor, T.E., O'Dell, C.W., O'Brien, D.M., Kikuchi, N., Yokota, T., Nakajima, T.Y., Ishida, H., Crisp, D., and Nakajima, T.: Comparison of cloud-screening methods applied to GOSAT near-infrared spectra, IEEE TGRS., 50, 295–309, doi: 10.1109/TGRS.2011.2160270, 2012.

Uchino, O., Kikuchi, N., Sakai, T., Morino, I., Yoshida, Y., Nagai, T., Shimizu, A., Shibata, T., Yamazaki, A., Uchiyama, A., Kikuchi, N., Oshchepkov, S., Bril, A., and Yokota, T.: Influence of aerosols and thin cirrus clouds on the GOSAT-observed CO2: A case study over Tsukuba, Atmos. Chem. Phys., 12, 3393–3404, doi: 10.5194/acp-12-3393-2012, 2012.

Vapnik, V. and Lerner, A.: Pattern recognition using generalized portrait method, Automat. Rem. Contr., 24, 6, 774–780, 1963.

Yoshida, Y., Eguchi, N., Ota, Y., Kikuchi, N., Nobuta, K., Aoki, T., and Yokota, T.: Algorithm Theoretical Basis Document (ATBD) for CO2 and CH4 column amounts retrieval from GOSAT TANSO-FTS SWIR, Available at: https://data2.gosat.nies.go.jp/GosatDataArchiveService/doc/GU/ATBD_FTSSWIRL2_V1.1_en.pdf (last access: 8 December 2017), 2010.