Algorithm Theoretical Basis Document of GCOM-C/SGLI Ver.3 Cloud Flag Product (CLFG)

by

Kazuhisa Tanada

-

Japan Aerospace Exploration Agency
Earth Observation Research Center

8 October 2021

Contents

1 Introduction **2**
- 1.1 Objectives . . . 2
- 1.2 Development strategy . . . 2

2 Theoretical Description **3**
- 2.1 Processing flow . . . 3
- 2.2 Deep neural network method . . . 3
  - 2.2.1 Output labels and Input data selection . . . 3
  - 2.2.2 DNN model . . . 5
  - 2.2.3 Ground truth . . . 6
  - 2.2.4 Training and results . . . 6
- 2.3 Flag selection for the final output . . . 6

3 Validation **10**
- 3.1 Validation results . . . 10
- 3.2 Future work . . . 11

Bibliography **13**

Introduction

1.1 Objectives

This Ver.3 cloud flag algorithm (CLFG; A1ABC) is developed for improving Ver.2 CLFG algorithm, where the known problems of Ver.2 CLFG are as follows:

- Accuracy of discriminating snow and ice from clear sky in day time
- Accuracy of discriminating heavy aerosols and clouds in day time
- Accuracy of discriminating cloud and clear-sky at night time

1.2 Development strategy

To solve the problems above, we newly developed a deep neural network (DNN) method. This DNN method is processed at the same time as Ver.2 CLFG method (CLAUDIA) [2] processing. Comparing the results obtained from two different methods for each pixel, the more plausible result is chose to output the final cloud flag. The cloud flag for the areas that do not need any improvements would be the same as the Ver.2 method because the algorithm of Ver.3 basically developed to focus on the weak points of Ver.2.

In particular, at night time, it is difficult to classify because only 2 thermal channels (10.8 and 12.0 μm) can be used. In addition, preparing the ground truth data is almost impossible since there are no important information such as visible images. We therefore use latitude, elevation, day of year, and 2 thermal channels data at day-time as the night-time training and ground truth data to improve the accuracy of classification.

Theoretical Description

2.1 Processing flow

Fig. 2.1 indicates the flow of the Ver.3 CLFG processing. The new method "DNN method" is processing simultaneously with processing of CLAUDIA method. The input data for the DNN method are the satellite data (LTOA product which is radiance data at the top of atmosphere observed by GCOM-C/SGLI) and 1/24° spatial resolution digital elevation model (DEM) data. The resolution of the input satellite radiance product is depending on the resolution of the output CLFG (Q:250m, K:1km and F:1/24°). Then, comparing the both results obtained from CLAUDIA and the DNN methods, the final cloud flag is determined by "flag selection" process.

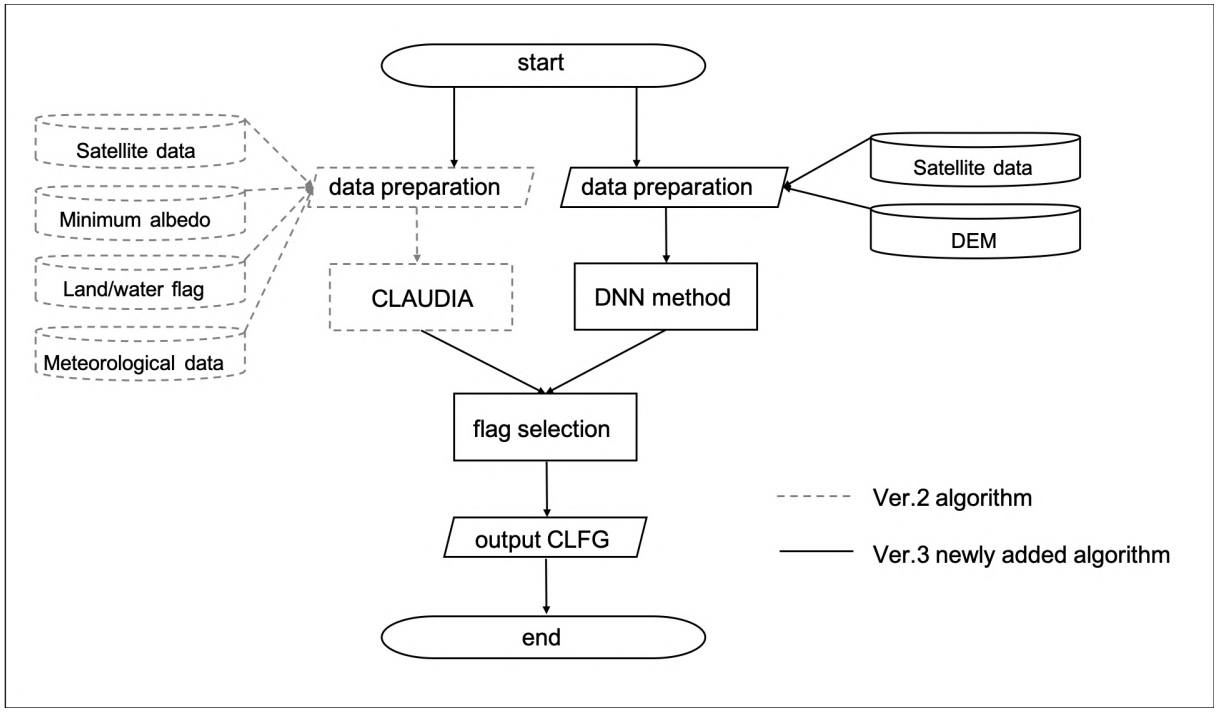

Figure 2.1

2.2 Deep neural network method

2.2.1 Output labels and Input data selection

The purpose of the DNN method is to obtain the classification such as clear sky, cloudy, snow/ice cover for each pixel of satellite images. First, we defined the output labels of the DNN model as 5 categories, "land", "ocean", "cloudy", "heavy aerosol", and "ice/snow" for the day-time process. Since the number of available satellite input data for the night (ascending) process is less than that of the day (descending) process, we reduce the number of categories to 4 to simplify the classification, including "land/ocean", "cloudy", "heavy aerosol", and "ice/snow" for the night-

time process. The difference between the number of categories for day and night is not a problem for the output cloud flag because both the categories of "land" and "ocean" have the same meaning of clear sky pixels.

In order to select the effective input data for the DNN model, we calculate the standard partial regression coefficient which shows the relative importance of different variables with the output categories. Fig. 2.2 shows the value distribution of the input data candidates for 5 categories (the list of the input data candidates is shown in Table 2.1). This distribution is obtained by collecting the SGLI's LTOA data and DEM globally for various day of year (the used data detail will be described in Section 2.2.3). The calculated standard partial regression coefficients for the input data candidates are shown in Fig. 2.3. We selected the 9 SGLI channels from 380 nm to 12.0 μm that have high coefficients, in addition to latitude data, elevation data (DEM) and day of year for the day-time process. For the night-time process, we used 2 SGLI thermal channels (10.8 and 12.0 μm), latitude data, DEM and day of year.

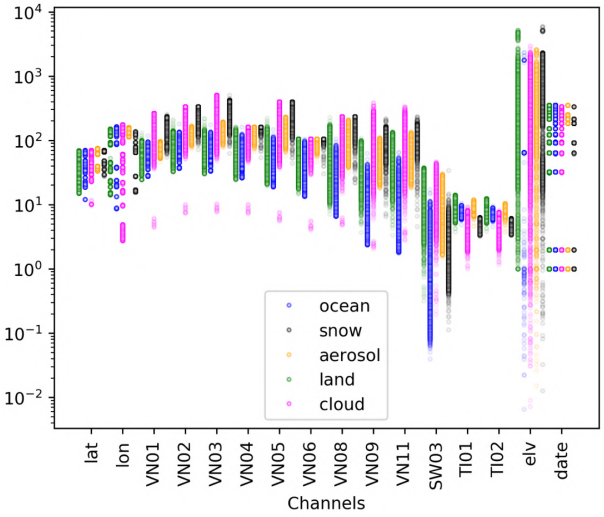

Figure 2.2

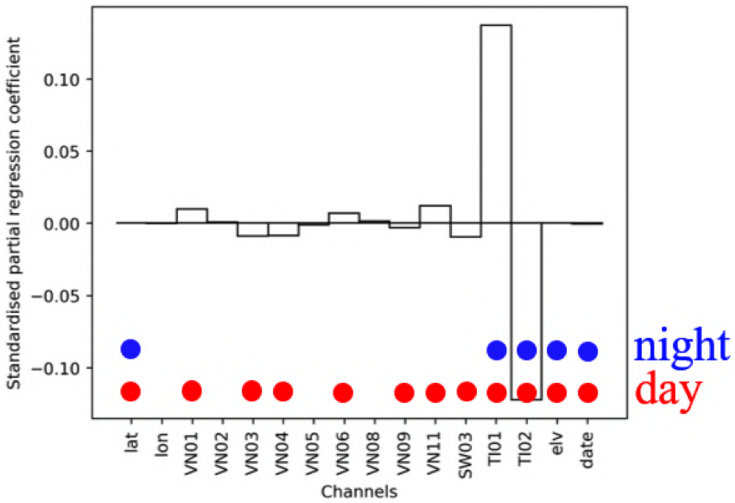

Figure 2.3

Then, the data as shown in Table 2.1 are used to input the DNN model.

Table 2.1: Input data for the day- and night-time processes

candidates	spatial resolution	day	night
latitude	250 m	✓	✓
longitude	250 m		
VN01 (380 nm)	250 m	✓	
VN02 (412 nm)	250 m		
VN03 (443 nm)	250 m	✓	
VN04 (490 nm)	250 m	✓	
VN05 (530 nm)	250 m		
VN06 (565 nm)	250 m	✓	
VN07 (673.5 nm)	250 m		
VN08 (673.5 nm)	250 m		
VN09 (763 nm)	250 m	✓	
VN10 (868.5 nm)	250 m		
VN11 (868.5 nm)	250 m	✓	
SW01 (1050 nm)	1000 m		
SW02 (1380 nm)	1000 m		
SW03 (1630 nm)	250 m	✓	
SW04 (2210 nm)	1000 m		
TI01 (10.8 μm)	250 m	✓	✓
TI02 (12.0 μm)	250 m	✓	✓
DEM	1/24°	✓	✓
date (day of year)	250 m	✓	✓

Note: Different input data are used for day (descending) and night (ascending) processing.

2.2.2 DNN model

We use a 3-layer DNN architecture in this method as described below (with Keras model [1]):

```
#Input layer:
model.add(Dense(200, input_dim=dim))
model.add(Activation("relu"))
model.add(Dropout(0.2))
#1-hidden layer
model.add(Dense(100))
model.add(Activation("relu"))
model.add(Dropout(0.2))
#2-hidden layer
model.add(Dense(8))
model.add(Activation("relu"))
model.add(Dropout(0.2))
#Output layer
model.add(Dense(final_dim))
model.add(Activation("softmax"))
```

The dimension of the input data "input_dim" is 12 for the day-time process, and 5 for the night-time process as shown in Table 2.1. The number of the output categories "final_dim" is 5 for the day-time process, and 4 for the night-time process. The training parameters are shown in Table 2.2.

Table 2.2: The training parameters of the DNN model

parameter	day	night
input dimension	12	5
output dimension	5	4
epochs	20	20
batch size	280	280
validation split ratio	0.1	0.1
optimizer	Adam(lr=0.001)	Adam(lr=0.001)

2.2.3 Ground truth

We made the training input dataset and the ground truth data at the various regions in various days of year. The locations of the training data (upper left location of the SGLI tiles) are shown in Fig. 2.4. At each location, we extracted the values of the selected input data as shown in Table 2.1, and labeled the category for each pixel considering the RGB true color image and QA flag of SIPR (Snow Ice PRoperties) prooduct with eyes. The examples of the way of categorizing are shown in Fig. 2.5 and Fig. 2.6. The total number of the training dataset is more than 500,000.

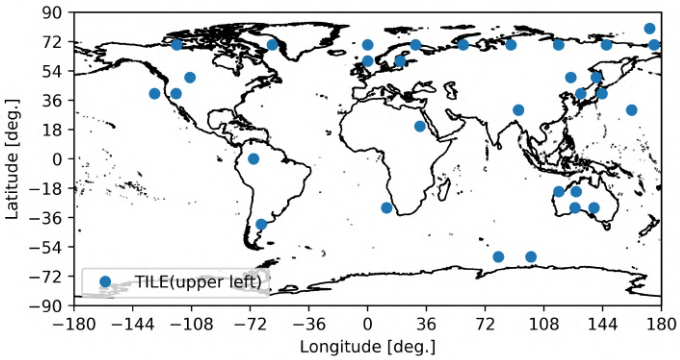

Figure 2.4: Training data set locations. The upper left locations of the SGLI tiles are plotted as blue filled circles.

2.2.4 Training and results

Fig. 2.7 and Fig. 2.8 show the training histories of the model loss and model accuracy for the day- and night-time processes using the DNN model as defined in Section 2.2.2. The accuracy of the trained model is shown in Table 2.3 and Table 2.4. Although the pixels used for verification were in the same tile as the area containing the training pixels (exactly the same pixels were not used to validate), the high accuracy was achieved for both day- and night-time trained models.

2.3 Flag selection for the final output

After processing both CLAUDIA and DNN methods, the final output cloud flag is determined by some flag selection process defined as in Table 2.5 for day-time process and Table 2.6 for night-time process.

After that selection process, the additional data screening will be applied in the cases below:

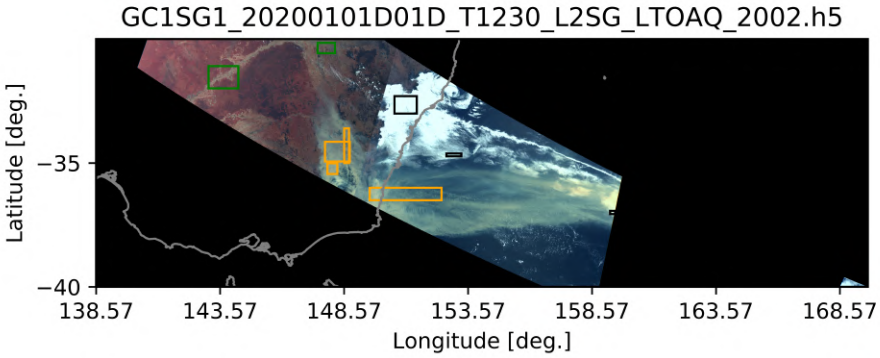

Figure 2.5: The example of the labeled regions of "land", "cloudy" and "heavy aerosol" are represented by the green squares, black squares and orange squares over SGLI's RGB true color image of the east area in Austalia, respectively.

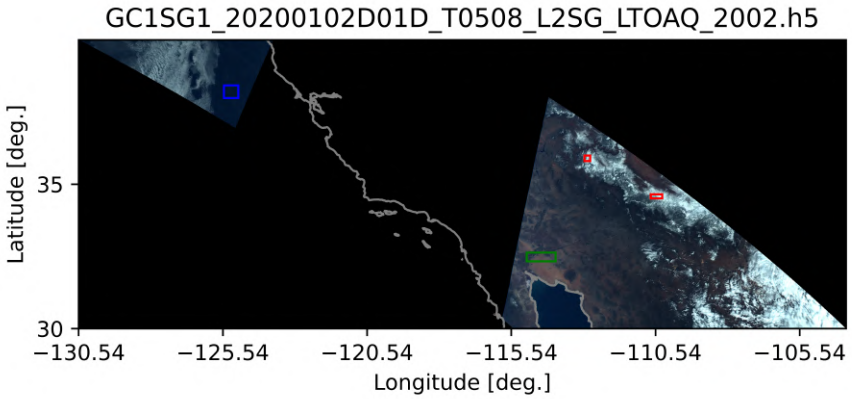

Figure 2.6: The example of the labeled regions of "ocean", "land" and "ice/snow" are represented by the blue square, green square and red squares over SGLI's RGB true color image of the west region in America, respectively.

Table 2.3: The obtained accuracy of the trained model for the day-time process

Category	Accuracy [%]	Corrected pixels	Total pixels
Total	99.78	5226469	5238201
Land	99.67	1230057	1234083
Ocean	99.93	811784	812343
Cloudy	99.84	1362311	1364443
Ice/snow	99.75	864008	866134
Heavy aerosol	99.70	958309	961198

Table 2.4: The obtained accuracy of the trained model for the night-time process

Category	Accuracy [%]	Corrected pixels	Total pixels
Total	96.29	4981847	5173855
Land/ocean	98.46	1925196	1955228
Cloudy	97.00	1331517	1372685
Snow/ice	99.57	841730	845393
Heavy aerosol	88.29	883404	1000549

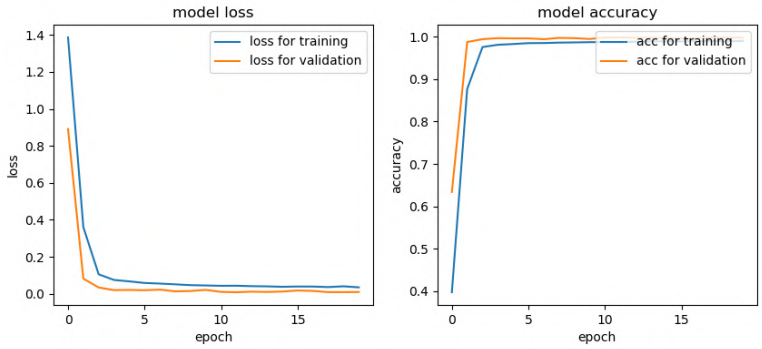

Figure 2.7

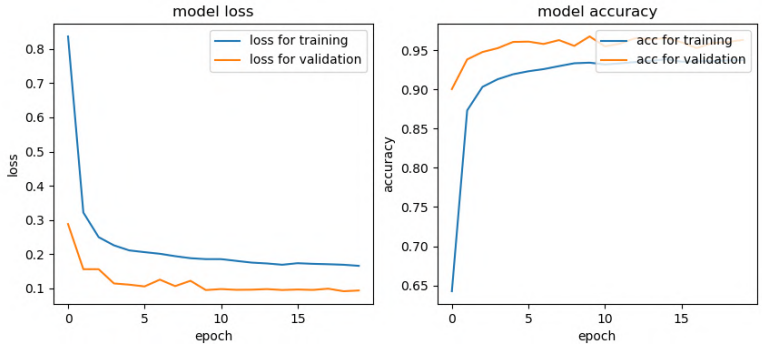

Figure 2.8

(1) Check the snow/ice flag in summer at low altitude area.
If:

- DNN flag is snow/ice &
- DOY (day of year) < 60 and DOY > 330 &
- DEM < 1500 m

use CLAUDIA flag.
else:
use DNN flag (snow/ice).

(2) Check the snow/ice flag at southern arctic area.
If:

- DNN flag is not cloudy &
- reflectance of VN09 (763 nm) > 0.5 &
- Land/Water flag $= 1$ (land) &
- $CCL_{\mathrm{CLDA}} \geq 67\%$

set the snow/ice flag. In addition, output CCL will be replaced to $CCL_{\mathrm{out}} = CCL_{\mathrm{CLDA}} + 17\%$.
else:
use Output flag as shown in Table 2.5, Table 2.6.

Table 2.5: Definition of the flag selection for day

CLAUDIA flag	DNN flag	Probability limit	Condition-1	CCL_{out}	Condition-2	Output flag
no snow/ice & no heavy aerosol	land/ocean	-	-	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	cloudy	$\geq 50\%$	if $CCL_{\mathrm{CLDA}} < 50\%$	$CCL_{\mathrm{CLDA}} - 17\%$	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	snow/ice	$\geq 60\%$	if $CCL_{\mathrm{CLDA}} \geq 67\%$	$CCL_{\mathrm{CLDA}} + 17\%$	-	snow/ice
			else	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	heavy aerosol	$\geq 99\%$	-	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
snow/ice	land/ocean	-	-	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	cloudy	$\geq 50\%$	if $CCL_{\mathrm{CLDA}} < 50\%$	$CCL_{\mathrm{CLDA}} - 17\%$	if $CCL_{out} \leq 33\%$	no snow/ice
	snow/ice	$\geq 60\%$	-	$CCL_{\mathrm{CLDA}} + 17\%$	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	heavy aerosol	$\geq 99\%$	-	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
heavy aerosol	land/ocean	-	-	CCL_{CLDA}	-	no heavy aerosol
	cloudy	$\geq 50\%$	-	CCL_{CLDA}	-	no heavy aerosol
	snow/ice	$\geq 60\%$	-	CCL_{CLDA}	-	no heavy aerosol
	heavy aerosol	$\geq 99\%$	-	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$

Note: "CLAUDIA flag" is the result flag obtained by the CLAUDIA method, where "DNN flag" is the predicted classification obtained by the DNN method. We use the results of the DNN method only when the prediction probability is more than the limit as shown in the column of "Probability limit". If the both results for CLAUDIA and DNN methods satisfy "condition-1", the output Clear Confidense Level (CCL_{out}) will be replaced to the value in the column of "CCL_{out}". When CCL_{out} is CCL_{CLDA}, it means there is no change from the CLAUDIA method. If CCL_{out} is more than 100% (less than 0%), CCL_{out} will be replaced to 100% (0%). After that, we apply the second check in the column of "condition-2", and if the condition is satisfied, the output flag will be replaced to the value in the column of "Output flag". When output flag is $\mathrm{flag}_{\mathrm{CLDA}}$, it means there is no change from the CLAUDIA method.

Table 2.6: Definition of the flag selection for night

CLAUDIA flag	DNN flag	Probability limit	Condition-1	CCL_{out}	Condition-2	Output flag
no snow/ice & no heavy aerosol	land/ocean	-	-	$CCL_{\mathrm{CLDA}} + 17\%$	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	cloudy	$\geq 50\%$	-	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	snow/ice	$\geq 60\%$	if $CCL_{\mathrm{CLDA}} \geq 17\%$	$CCL_{\mathrm{CLDA}} + 17\%$	-	snow/ice
			else	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	heavy aerosol	$\geq 99\%$	if $CCL_{\mathrm{CLDA}} = 100\%$	$CCL_{\mathrm{CLDA}} - 17\%$	-	$\mathrm{flag}_{\mathrm{CLDA}}$
			else	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
snow/ice	land/ocean	-	-	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	cloudy	$\geq 50\%$	if $CCL_{\mathrm{CLDA}} < 50\%$	$CCL_{\mathrm{CLDA}} - 17\%$	-	no snow/ice
	snow/ice	$\geq 60\%$	-	$CCL_{\mathrm{CLDA}} + 17\%$	-	$\mathrm{flag}_{\mathrm{CLDA}}$
	heavy aerosol	$\geq 99\%$	-	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$
heavy aerosol	land/ocean	-	-	CCL_{CLDA}	-	no heavy aerosol
	cloudy	$\geq 50\%$	-	CCL_{CLDA}	-	no heavy aerosol
	snow/ice	$\geq 60\%$	-	CCL_{CLDA}	-	no heavy aerosol
	heavy aerosol	$\geq 99\%$	-	CCL_{CLDA}	-	$\mathrm{flag}_{\mathrm{CLDA}}$

Note: the same description as Table 2.5.

Validation

3.1 Validation results

In this section, we show the validation results for the Ver.3 cloud flag. We focused on the 3 problems as described in Section 1.

Fig. 3.1 shows the example of the Ver.3 cloud flag result at the tile "T0426" in 1 January 2020 compared with the RGB true color image and Ver.2 cloud flag. Within the red circle area, the true classification is snow, however Ver.2 cloud flag includes many clear-sky pixels, which might cause the overestimation of the aerosol optical thickness product due to the high reflectance of ice/snow. The obtained Ver.3 cloud flag shows that the area is covered by snow. The number of the changed pixels from Ver.2 cloud flag is shown in the right panel of Fig.3.1.

Fig. 3.2 shows the example of the Ver.3 cloud flag result at the tile "T0508" in 9 September 2020 compared with the RGB true color image and Ver.2 cloud flag. Within the red circle area, the true classification is heavy aerosol, however Ver.2 cloud flag shows that the most of the pixels in the area are cloudy, which might cause the lack of the valid pixels for the aerosol products over heavy aerosols event. The obtained Ver.3 cloud flag correctly detected the heavy aerosol at the region. The number of the changed pixels from Ver.2 cloud flag is shown in the right panel of Fig.3.2.

Fig. 3.3 shows the latitudinal distribution of the ratio of the clear-sky pixels at night in 4 seasons obtained from the Ver.3 cloud flag compared with that of Ver.2 cloud flag. The Ver.3 clear-sky pixel ratio modestly increased from Ver.2 cloud flag in every season (the maximum increase is $\sim$ 20%), although the visual changes are slight as shown in bottom panels of Fig. 3.3. The valid pixels for the other SGLI products such as LST (Land Surface Temperature) are expected to increase due to its improvements of cloud flag at night. However, the latitudinal distribution indicates that the problem still remains for improving the accuracy of clear-sky detection at high latitude area during night time (future work).

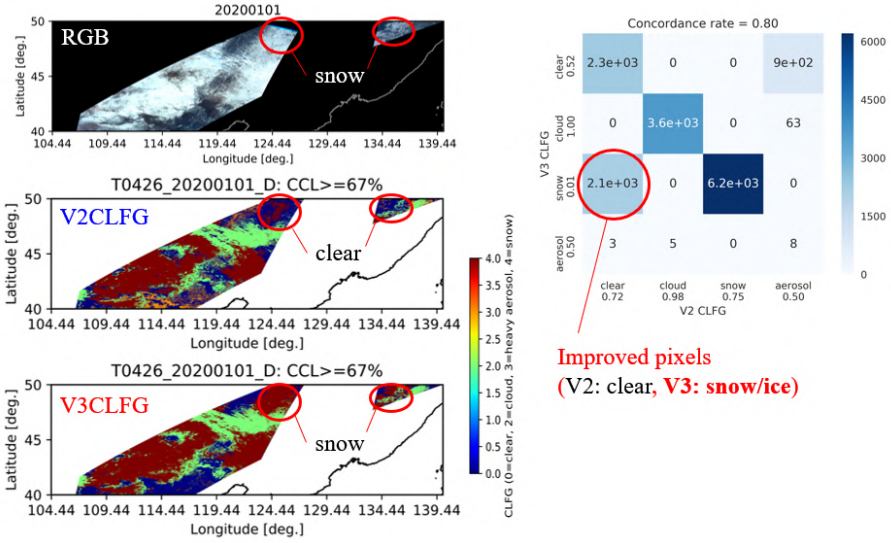

Figure 3.1: The example of the Ver.3 cloud flag result at the tile "T0426" in 1 January 2020. The Ver.3 algorithm correctly classified the ice/snow pixels.

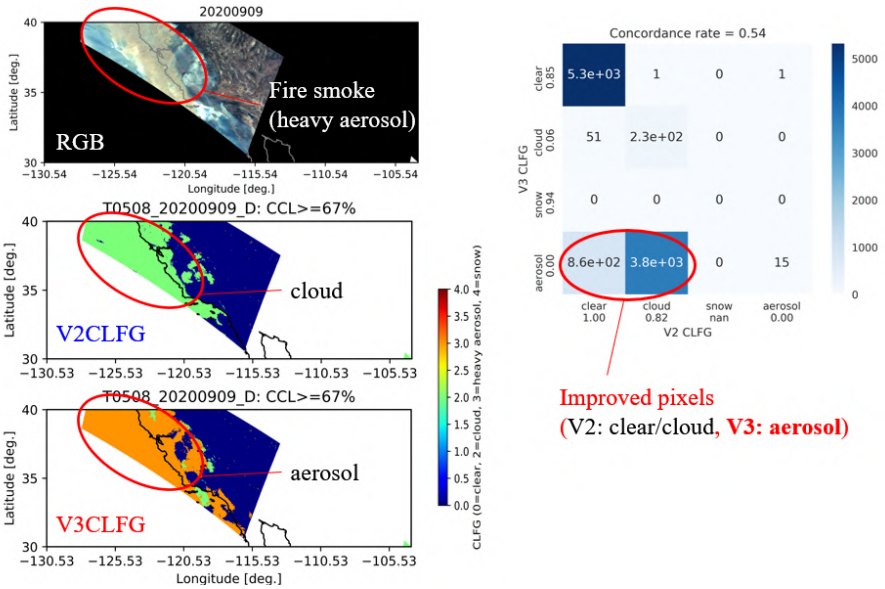

Figure 3.2: The example of the Ver.3 cloud flag result at the tile "T0508" in 9 September 2020. The Ver.3 algorithm correctly classified the heavy aerosol pixels.

As a result, the main improvements of the Ver.3 cloud flag algorithm are as below:

- [∼Ver.2 problem] Accuracy of discriminating snow and ice from clear sky in day time
  → Misclassification pixels decreased.
- [∼Ver.2 problem] Accuracy of discriminating heavy aerosols and clouds in day time
  → Heavy aerosol pixels can be classified at day time.
- [∼Ver.2 problem] Accuracy of discriminating cloud and clear-sky at night time
  → Clear-sky pixels increased at night time. However, there are still not enough pixels for high latitude regions at night time.

3.2 Future work

- Improve the accuracy of the night-time classification among cloudy, clear-sky land and ice/snow at high latitude area.
- Detect the heavy aerosols at night time.

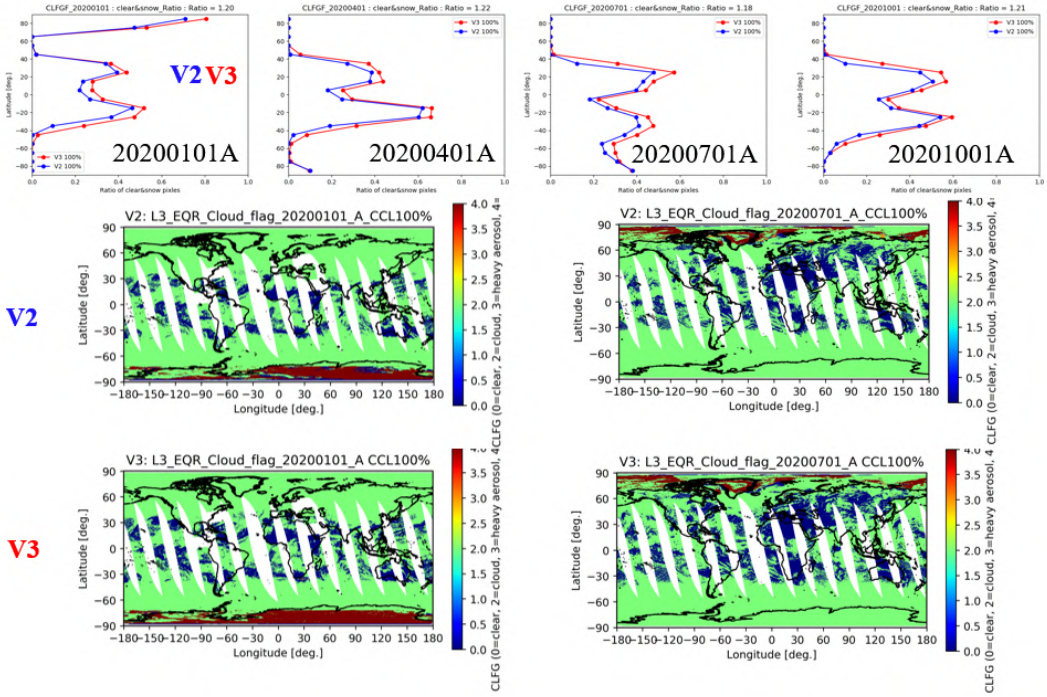

Figure 3.3: The comparison of Ver.2 and Ver.3's the latitudinal distribution of the clear-sky pixel ratio on 4 dates, 1 January 2020, 1 April 2020, 1 July 2020 and 1 October 2020 at night time (upper 4 panels). The bottom 4 panels show the examples of the global cloud flag images in 1 January 2020 and 1 July 2020 for Ver.2 and Ver.3.

Bibliography

[1] F. Chollet et al. Keras. https://keras.io, 2015. 5

[2] T. Y. Nakajima, H. Ishida, T. M. Nagao, M. Hori, H. Letu, R. Higuchi, N. Tamaru, N. Imoto, and A. Yamazaki. Theoretical basis of the algorithms and early phase results of the gcom-c (shikisai) sgli cloud products. *Progress in Earth and Planetary Science*, 6(1):52, 2019. 2