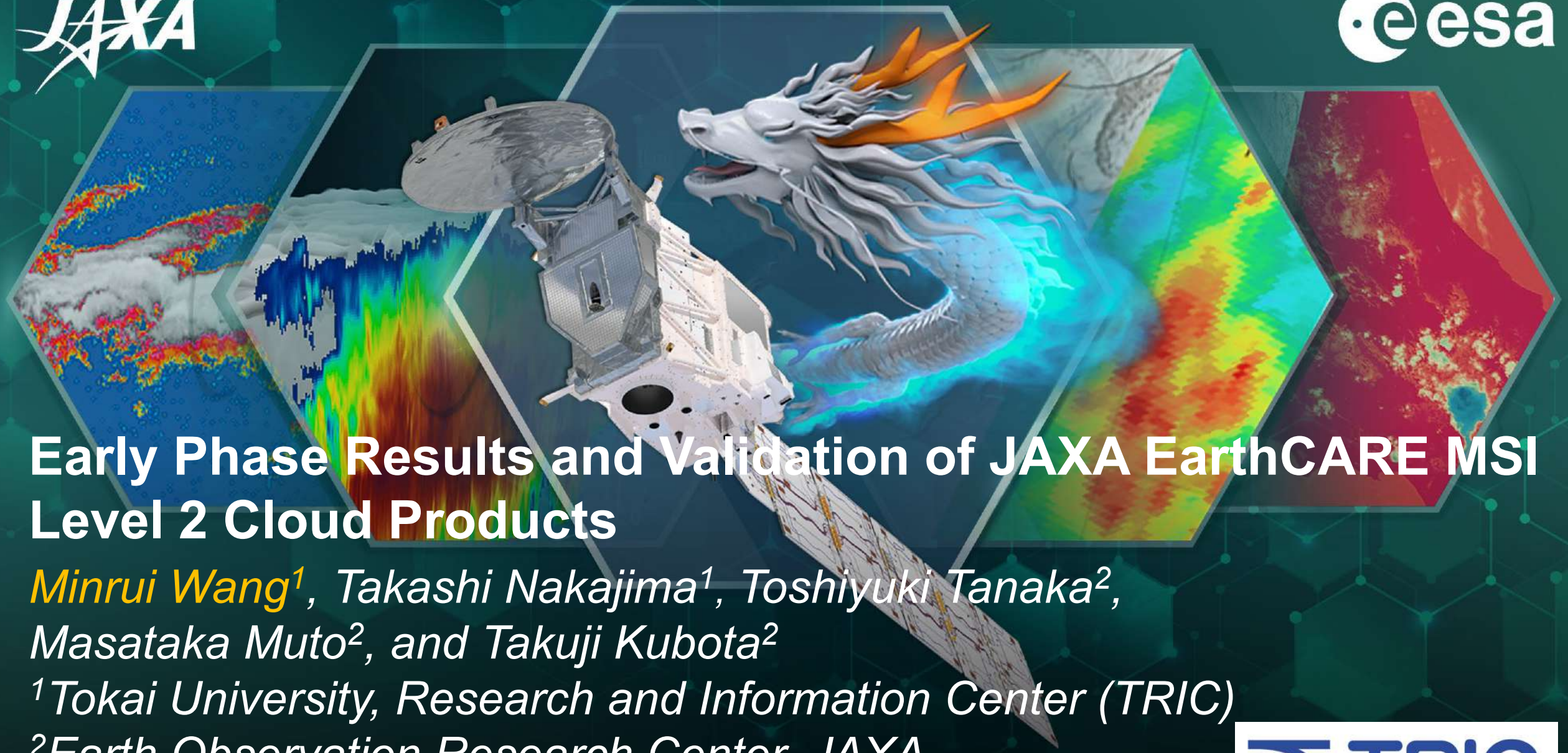


Early Phase Results and Validation of JAXA EarthCARE MSI Level 2 Cloud Products

Minrui Wang[1], Takashi Nakajima[1], Toshiyuki Tanaka[2], Masataka Muto[2], and Takuji Kubota[2]

[1]Tokai University, Research and Information Center (TRIC)

[2]Earth Observation Research Center, JAXA

2nd ESA-JAXA EarthCARE In-Orbit Validation Workshop

17 – 20 March 2025 | ESA-ESRIN | Frascati (Rome), Italy

Introduction

- The JAXA EarthCARE/MSI L2 cloud products (MSI_CLP) has been updated to v1.0 by early 2025.
  - *During the updates from v0.4 to v1.0, we fixed several bugs and applied new LUTs for both water and ice cloud retrievals.*

- After the launch of EarthCARE in May 2024, we started the validation of MSI L2 cloud flag products using our whole-sky camera system.
- On the other side, validation of cloud properties based on the newest version (vAc)※ of MSI L1 cloud data also stared from Jan 2025.

※ up to middle February of 2025

Output Chart of MSI_CLP

プロダクト出力項目表（Ver 1.0）

MSI単体　L2a　雲プロダクト　（標準プロダクト）

2025/1/14

構造定義ファイル設定値

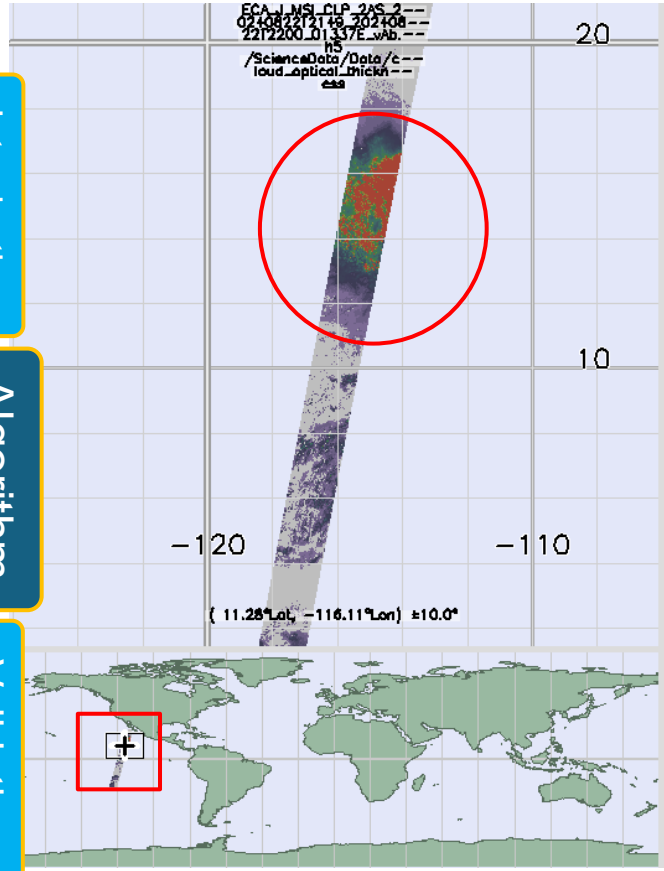
NC_FILL_BYTE　(-127)
NC_FILL_SHORT　(-32767)
NC_FILL_INT　(-2147483647)
NC_FILL_FLOAT　(9.9692099683868690e+36)
NC_FILL_DOUBLE　(9.9692099683868690e+36)

構造定義ファイル設定値

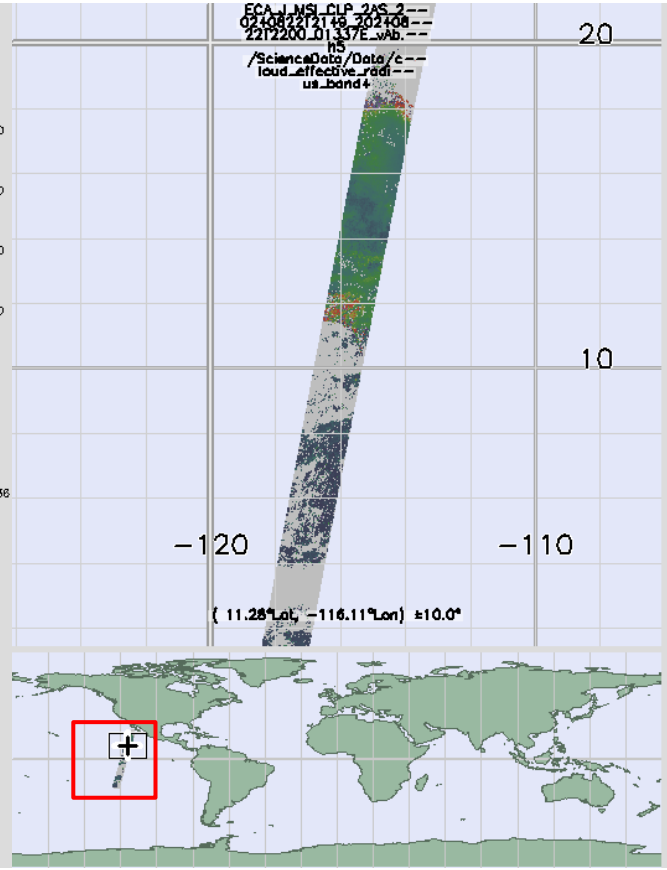
NC_FILL_BYTE	-127
NC_FILL_SHORT	-32767
NC_FILL_INT	-2147483647
NC_FILL_FLOAT	9.9692099683868690e+36
NC_FILL_DOUBLE	9.9692099683868690e+36
-	-

No.	出力	グループ Group	項目名 Parameters	日本語名称	英語名称 long_name	単位 units	出力最小値 Min of	出力最大値 Max of	欠損値 _FillValue	グリッドスペーシング（ピクセル間隔） Grid spacing 軌道方向	観測幅方向	鉛直
1	○	Geo	latitude	観測緯度	Latitude	degree_north	-90.0	90.0	NC_FILL_DOUBLE	0.5 km	0.5 km	-
2	○	Geo	longitude	観測経度	Longitude	degree_east	-180.0	180.0	NC_FILL_DOUBLE	0.5 km	0.5 km	-
3	○	Geo	day_night_flag	昼夜フラグ	Day night flag	-	0	1	NC_FILL_BYTE	0.5 km	0.5 km	-
4	○	Geo	land_water_flag	陸水フラグ	Land water flag	-	0	1	NC_FILL_BYTE	0.5 km	0.5 km	-
5	○	Geo	time	観測時刻	Time	seconds since 2000-1-1 00:00:00.0 0:00	-	-	NC_FILL_DOUBLE	0.5 km	0.5 km	-
6	○	Geo	relative_azimuth_angle	相対方位角	Relative Azimuth Angle	degree	0	180.0	NC_FILL_FLOAT	0.5 km	0.5 km	-
7	○	Geo	solar_zenith_angle	太陽天頂角	Solar Zenith Angle	degree	0	180.0	NC_FILL_FLOAT	0.5 km	0.5 km	-
8	○	Geo	satellite_zenith_angle	衛星天頂角	Satellite Zenith Angle	degree	0	90.0	NC_FILL_FLOAT	0.5 km	0.5 km	-
9	○	Data	cloud_optical_thickness	雲光学的厚さ	Cloud Optical Thickness	-	0	150	NC_FILL_FLOAT	0.5 km	0.5 km	-
10	○	Data	cloud_effective_radius_band3	雲粒有効半径(1.6μm)	Cloud Effective Radius at 1.6μm	micro m	0	100	NC_FILL_FLOAT	0.5 km	0.5 km	-
11	○	Data	cloud_effective_radius_band4	雲粒有効半径(2.2μm)	Cloud Effective Radius at 2.2μm	micro m	0	100	NC_FILL_FLOAT	0.5 km	0.5 km	-
12	○	Data	cloud_top_temperature	雲頂温度	Cloud Top Temperature	K	150.0	350.0	NC_FILL_FLOAT	0.5 km	0.5 km	-
13	○	Data	cloud_top_height	雲頂高度	Cloud Top Height	m	0.0	20000.0	NC_FILL_FLOAT	0.5 km	0.5 km	-
14	○	Data	cloud_top_pressure	雲頂気圧	Cloud Top Pressure	hPa	1.0	1100.0	NC_FILL_FLOAT	0.5 km	0.5 km	-
15	○	Data	quality_flag	QAフラグ	Quality Assurance Flag	-	0	255	-	0.5 km	0.5 km	-
16	○	Data	cloud_flag_bitfield	雲フラグビットフィールド	Cloud Flag Bit Field	-	0	4294967295	-	0.5 km	0.5 km	-

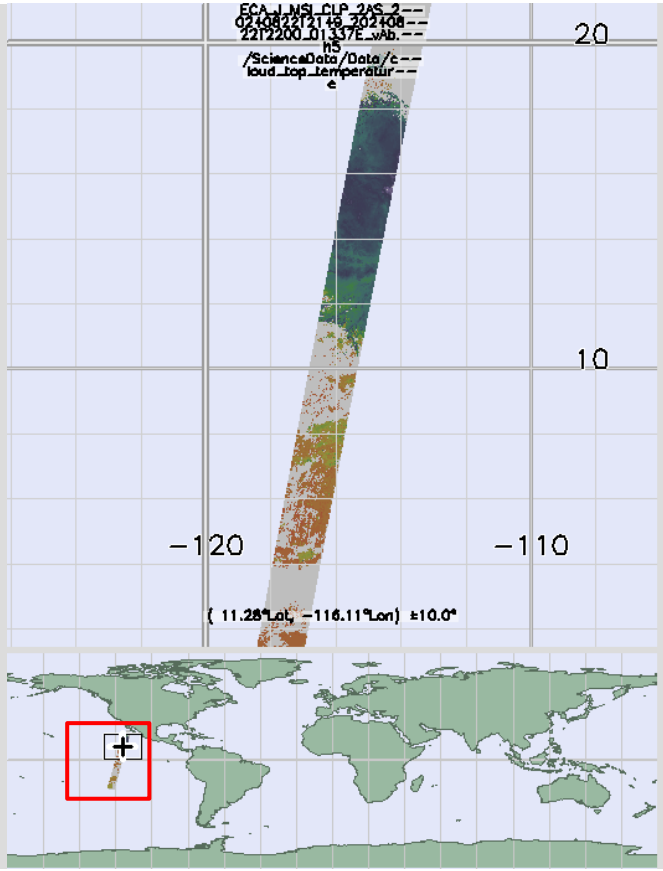
Sample of MSI_CLP



Cloud Optical Thickness (COT)



Cloud Droplet Radius (CDR) (band4=2.2μm)



Cloud Top Temperature (CTT)

01337E, 21:49-22:00 (UTC), 22 Aug 2024

ECA_J_MSI_CLP_2AS_20240822T2149_20240822T2200_01337E_vAb.h5

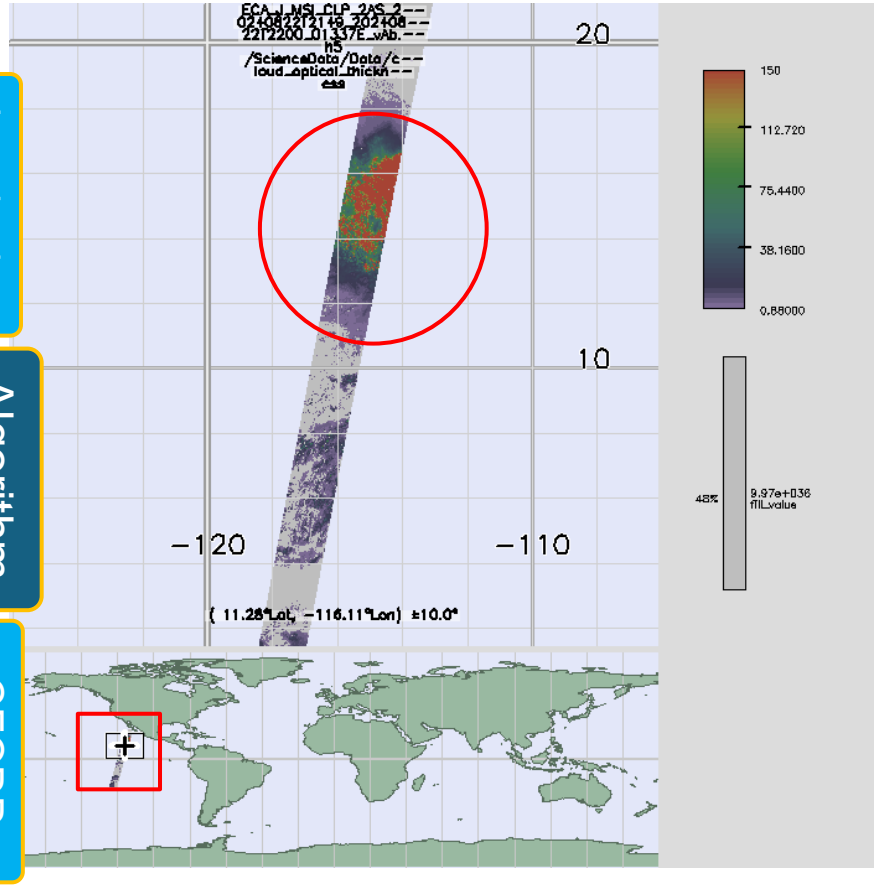
Measuring a hurricane

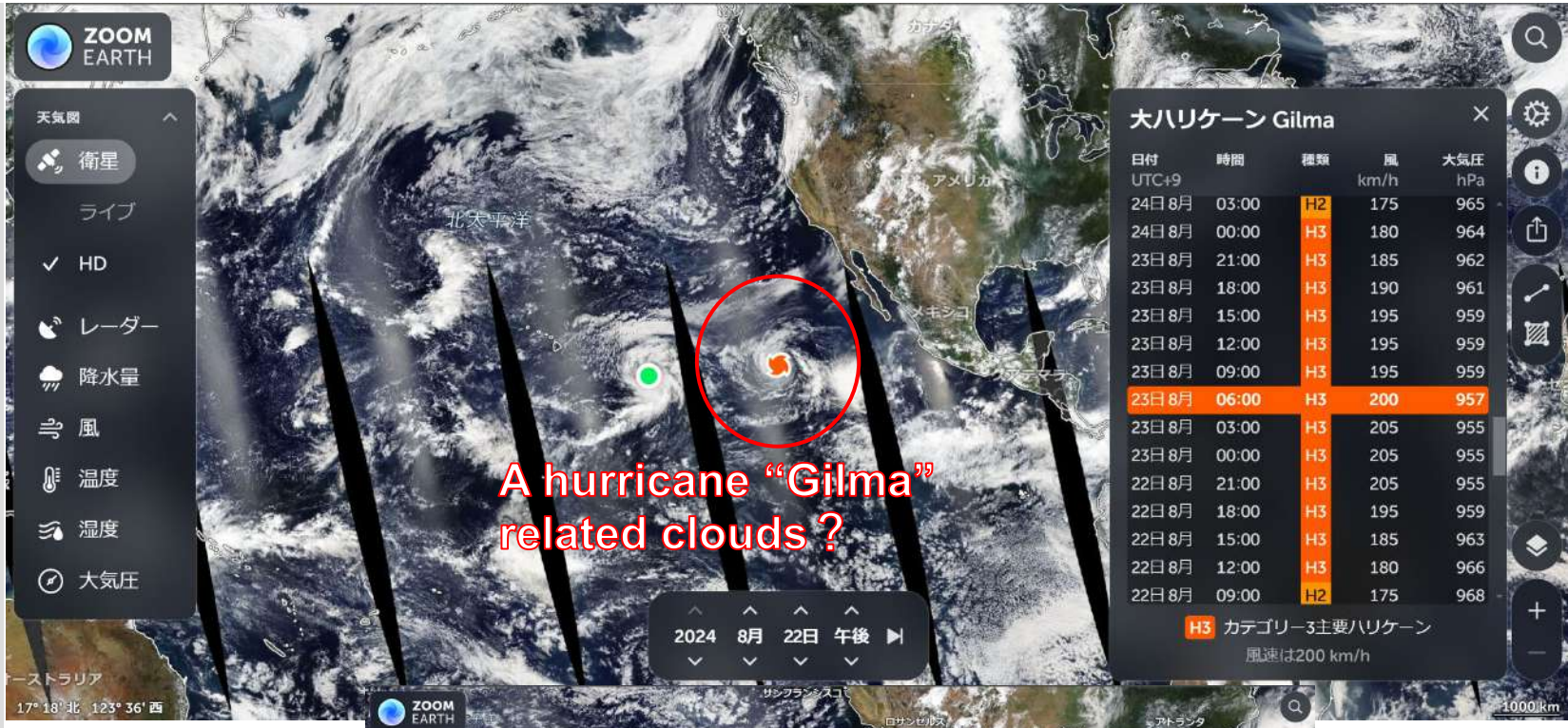
Introduction

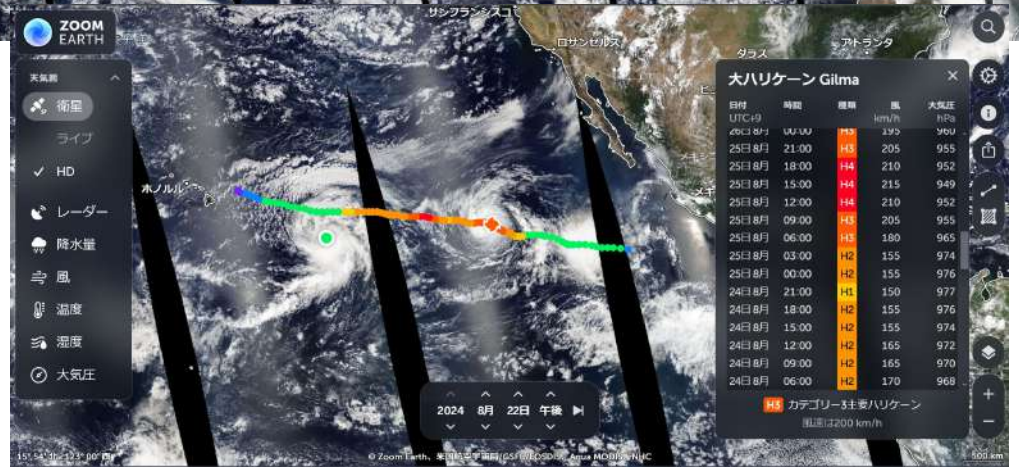
Algorithm

CFODD

Summary





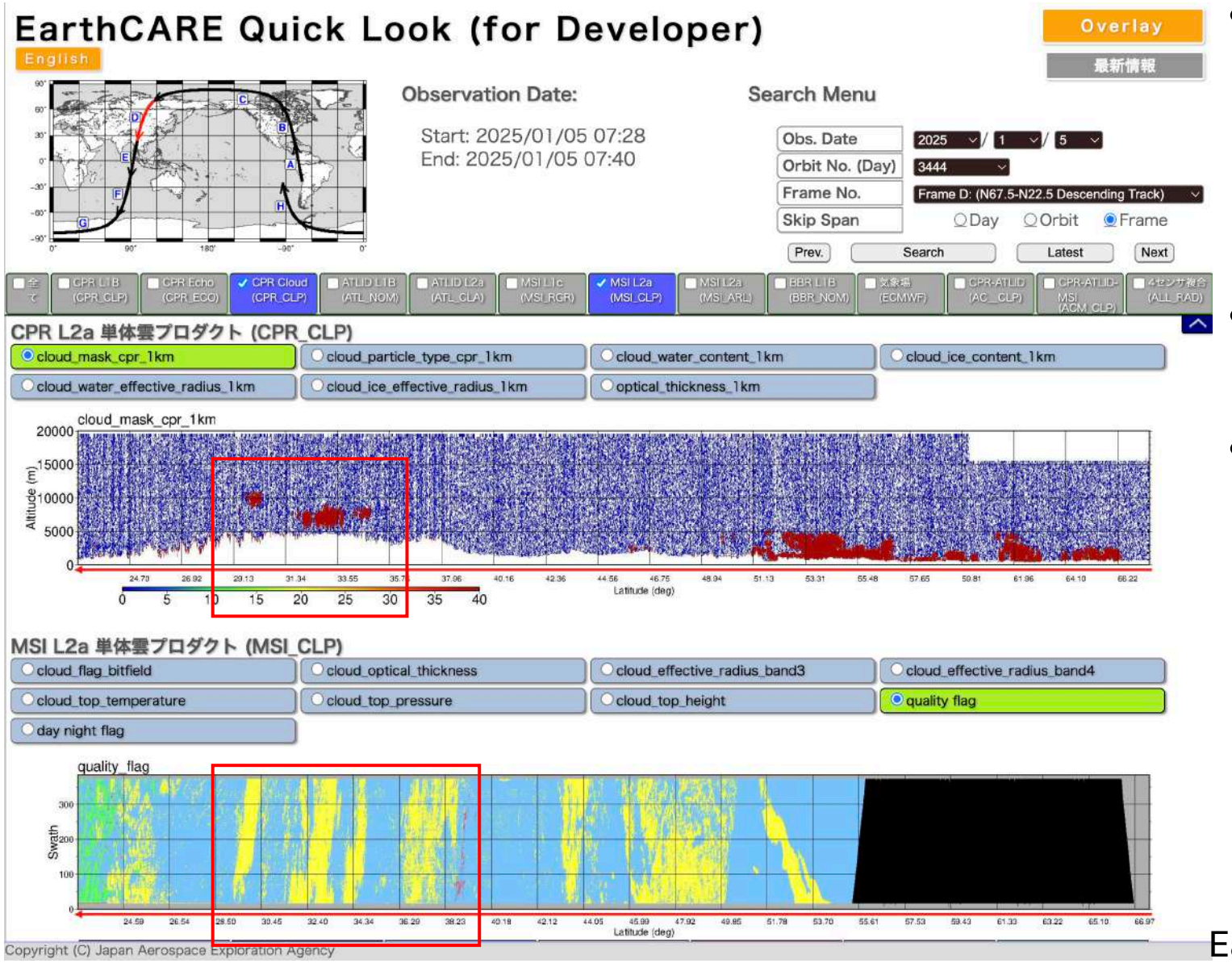


Cloud Optical Thickness (COT)

21:49-22:00 (UTC), 22 Aug 2024

https://zoom.earth/storms/gilma-2024/

Still some issues…



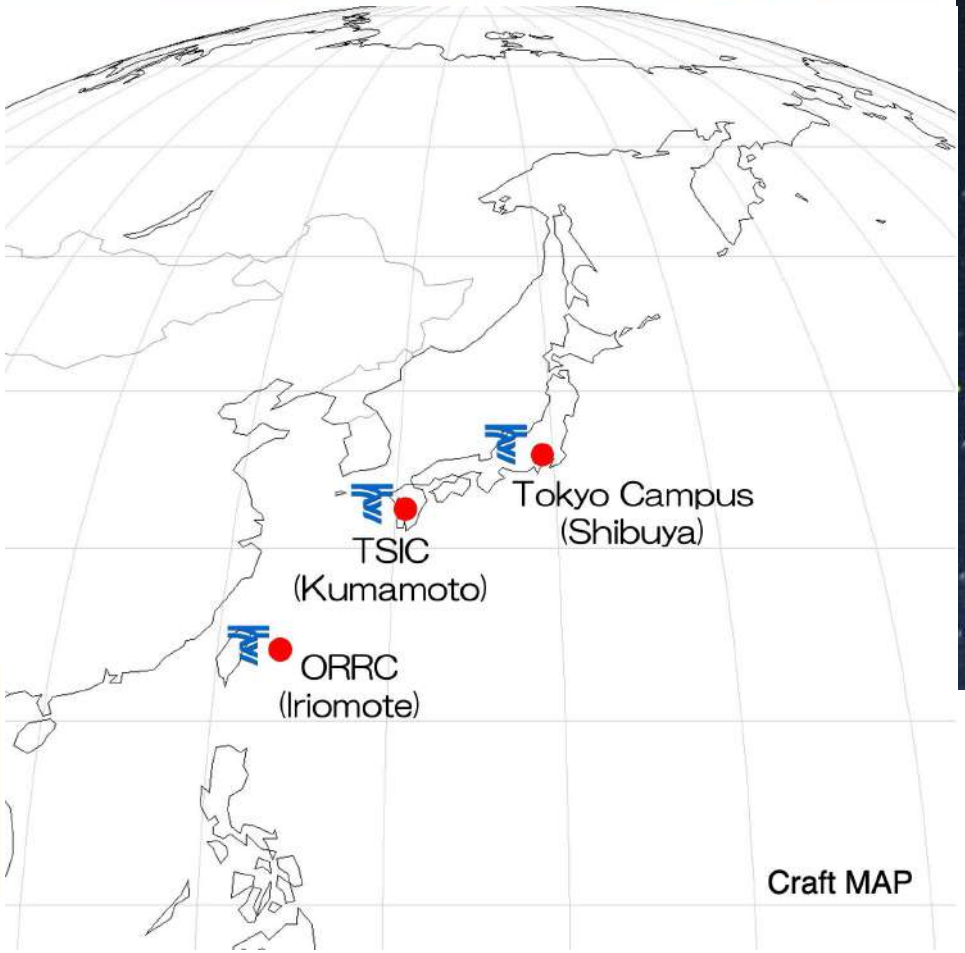
- "Unsatisfied condition" in MSI_CLP quality flag is frequently seen upon high elevation regions (Himalayas, Rockies etc.) and polar regions.
- MSI cannot distinguish snow cover and cloud well by itself.
- NDSI (Normalized Difference Snow Index) used in the snow/ice flag cannot be calculated, since MSI does not have 530nm band.

$$\mathrm{NDSI} = \frac{R_f(530\ \mathrm{nm}) - R_f(1630\ \mathrm{nm})}{R_f(530\ \mathrm{nm}) + R_f(1630\ \mathrm{nm})}$$

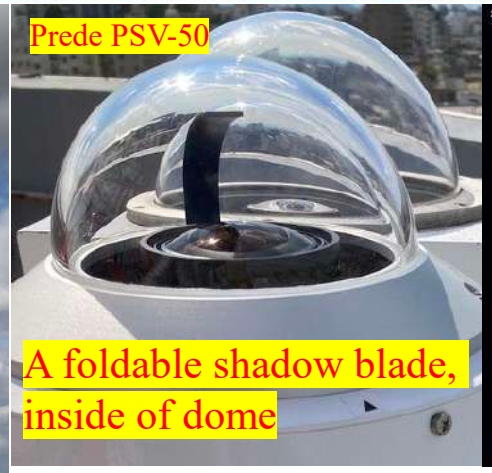
EarthCARE quick look site for developers by JAXA

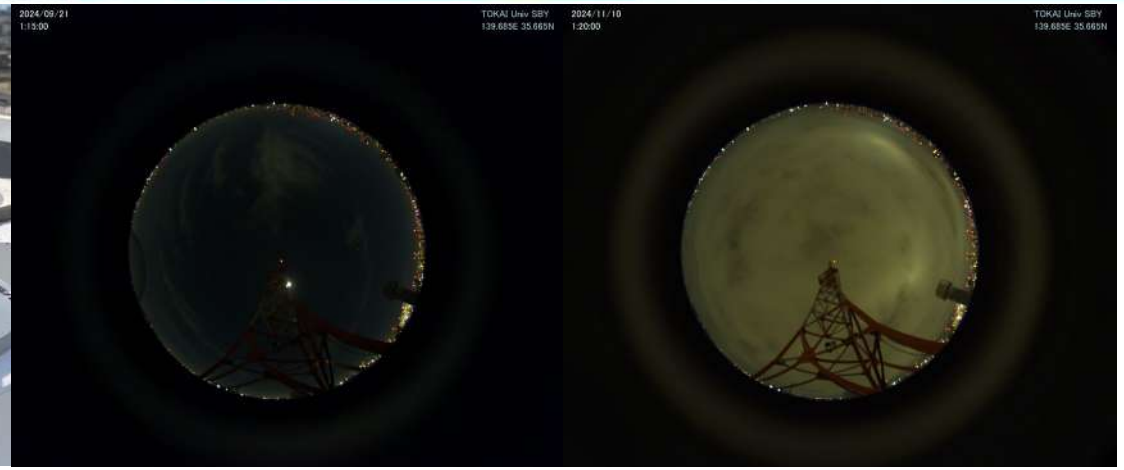
Early Phase Validation Results for L2 Cloud flag

Whole Sky Camera system in Tokai University



A New whole sky Camera system at Shibuya





Old Camera: Nikon D610

New Camera: Prede PSV-50

Usual method of cloud cover determination

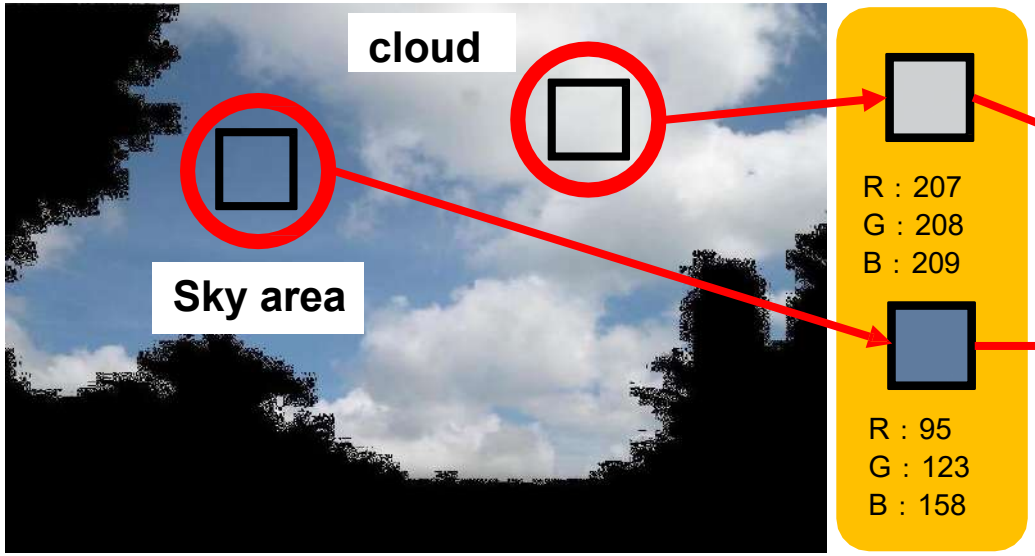
- Determines cloud area from the RGB information of each pixel in the camera image
- Using SI-BI method from Yoshimura and Yamashita(2013)
  - SI(Sky Index) : Indicator of the blue-whiteness of the sky

$$SI = (B - R) / (B + R)$$

  - BI(Brightness Index) : Indicator of sky brightness

$$BI = (R + G + B) / (255 * 3)$$

① apply the background mask for non-analysis areas
② calculate SI and BI for every pixel
③ make 2D histogram for SI, BI, setting the threshold manually
④ mark cloud pixels which under the threshold



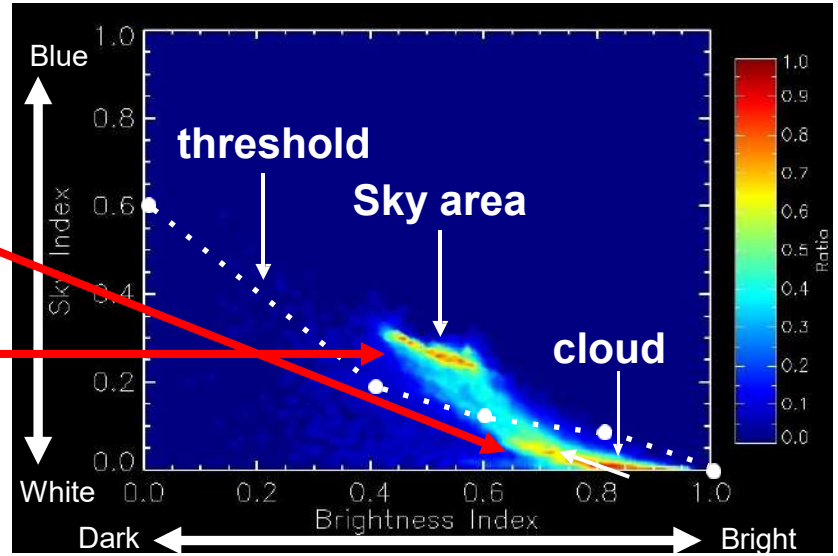
R : 207
G : 208
B : 209

R : 95
G : 123
B : 158

process

Determine the cloud area

2-dimensional Histogram



Eliminate the effect of subjectivity

Introduction

Algorithm

Validation

Summary

■Current situations

- Ground-based cloud cover observations made by the JMA were mostly visual inspection by human eyes.

（Subjectivity of the observer is likely to influence the results）

- Attempts have been made to determine cloud cover using camera images, but the accuracy was currently low.

■Objectives

- Improvement of accuracy, automation, and practical application of cloud coverage observation work from the ground using machines
  - ➢ Improvement of the analysis algorithm for fixed-point camera images
- Objective observation and quantification of cloud cover ＝Ensure that the observer's subjectivity does not influence the results of the observation

Supervised machine learning algorithm

➢Explanatory Variable

Input data used by machine learning models to make predictions

➢Objective Variable

- Data to be predicted by the model, in this case, 0 or 1(0 = Sky, 1 = Cloud)

➢Logistic regression(LR)

Find a **linear relationship** between the explanatory variable and the objective variable

➢XG-Boost(XG)

Objective variable is obtained by capturing the interaction of explanatory variables in a **nonlinear manner**

➢Why we employed machine learning algorithms with different approaches?

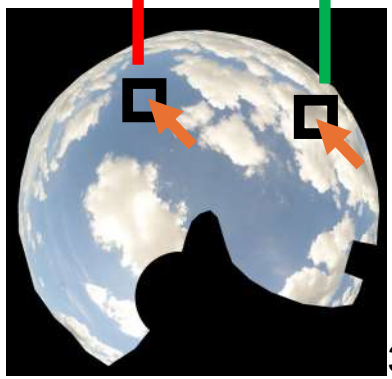
To capture the features of the created training model

To investigate the generalizability of the created training model

Creating Learning Model

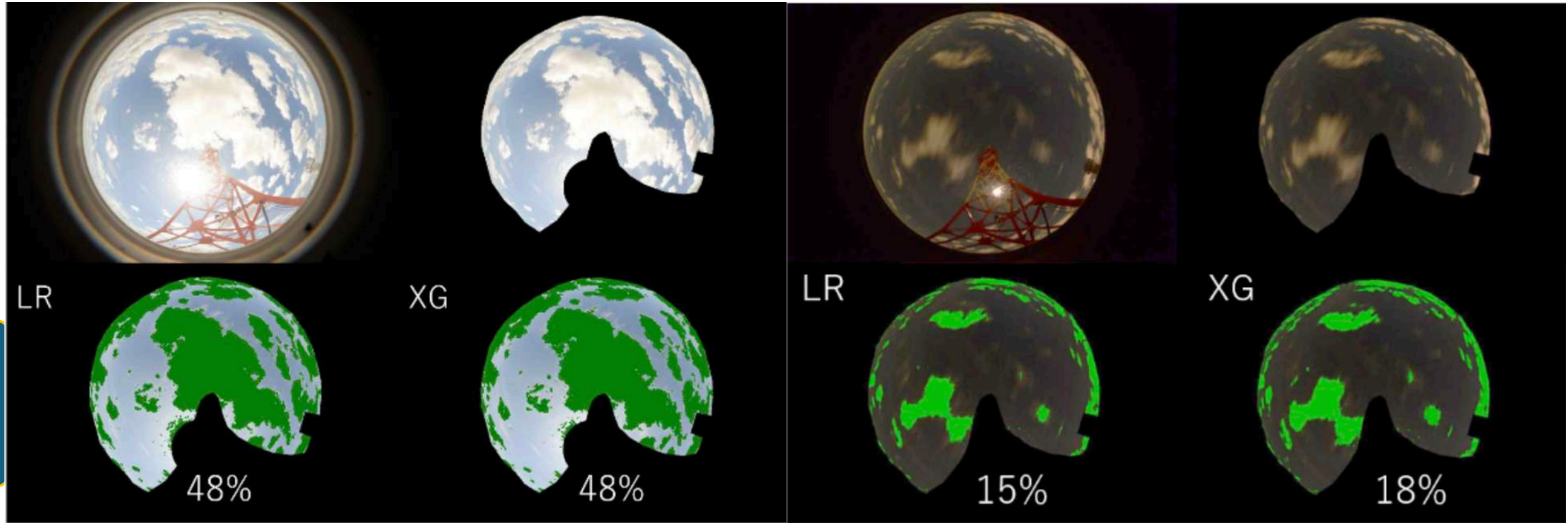
- Learning model = input data for the machine leaning algorithm
- extracts pixel-by-pixel data from images
  - 1000 images
  - 2500 pixel data
  - 1 model for 1 location

r_ave	g_ave	b_ave	R	G	B	SI_ave	BI_ave	SI	BI	Flag
108.0364	115.3431	129.5755	238	233	230	0.090648	117.6516	-0.01709	0.91634	1
108.0364	115.3431	129.5755	237	232	228	0.090648	117.6516	-0.01935	0.911111	1
121.6784	125.2098	133.8789	132	134	152	0.047741	126.9224	0.070423	0.546405	1
121.6784	125.2098	133.8789	153	153	161	0.047741	126.9224	0.025478	0.610458	1
66.65632	81.00546	104.8096	83	116	159	0.222512	84.15712	0.31405	0.467974	0
66.65632	81.00546	104.8096	72	103	149	0.222512	84.15712	0.348416	0.423529	0
85.56523	99.64785	123.3208	70	96	146	0.180747	102.8446	0.351852	0.407843	0
85.56523	99.64785	123.3208	73	101	149	0.180747	102.8446	0.342342	0.422222	0



- Explanatory variable
  - R, G, B, SI, BI
  - Averaged R, G, B, SI, BI (_ave) for target pixels
- Objective variable
  - 0 or 1(0 = Sky, 1 = Cloud)

Sample images of cloud fraction determination



Basically, the accuracy of cloud fraction determination is higher in daytime than in nighttime.

Comparison with RGB composite images

Natural Color RGB composite image using band 3, 2, 1of MSI as R, G, B

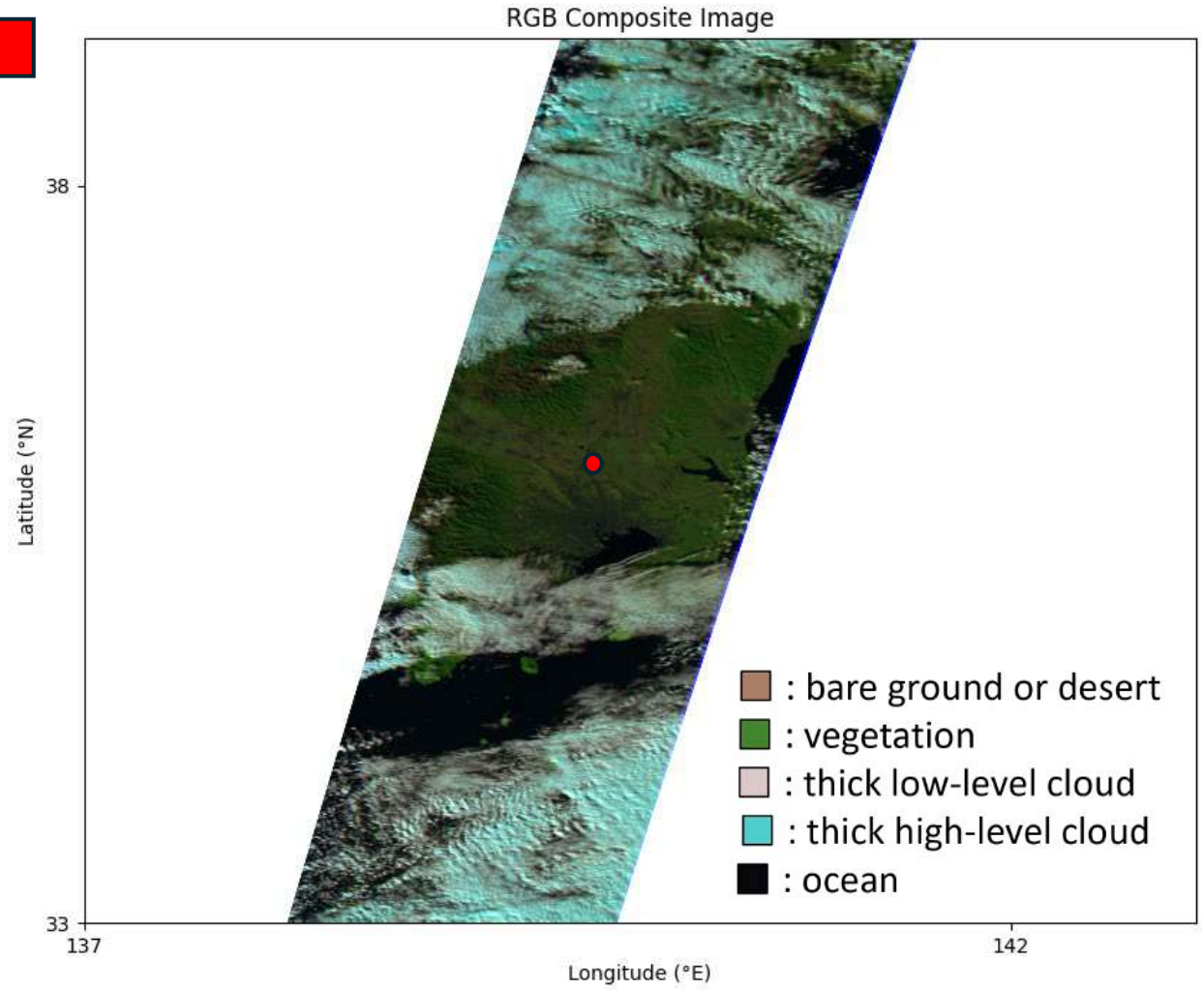
Introduction

Algorithm

Validation

Summary

Clear

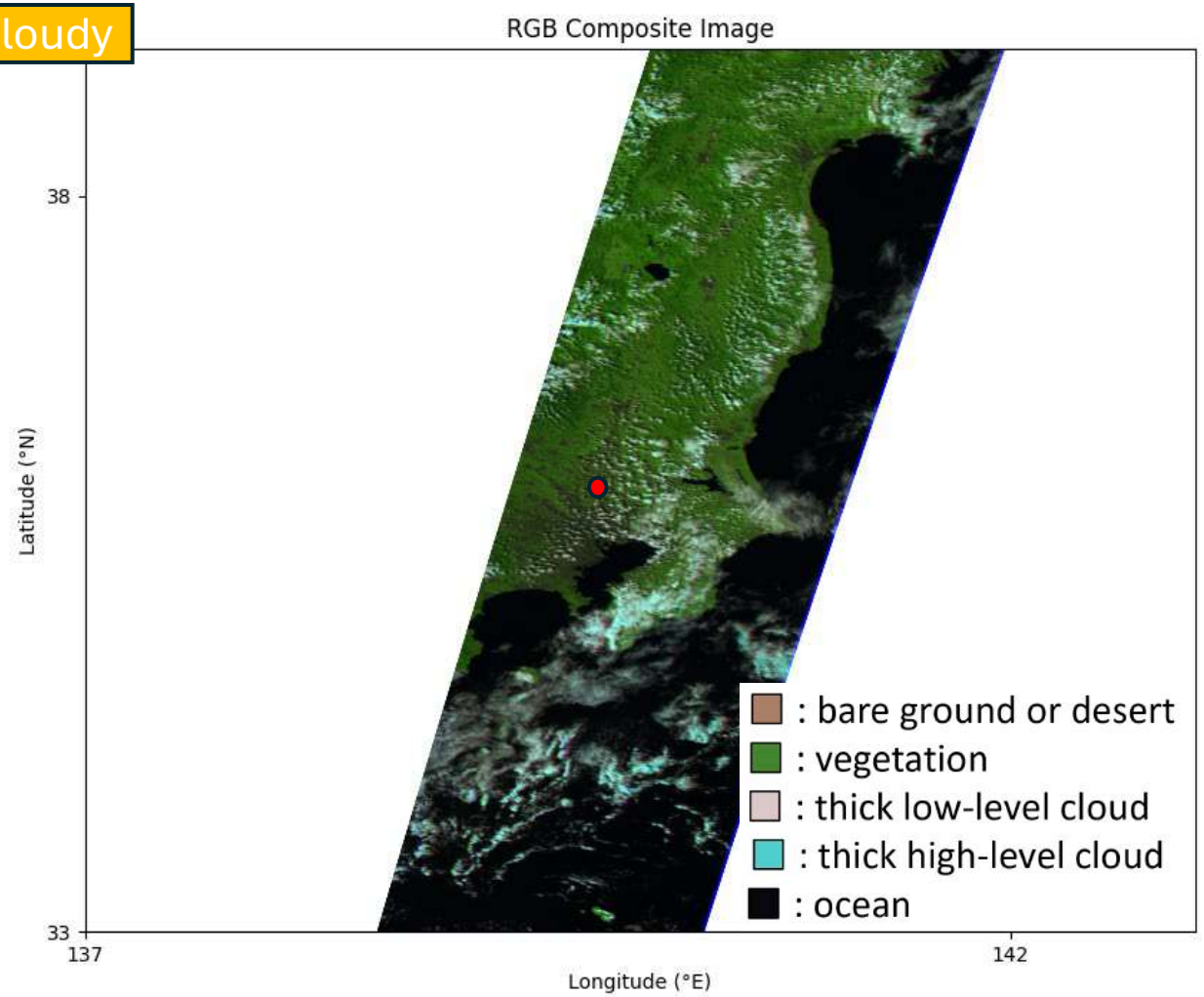


● Shibuya, Tokyo

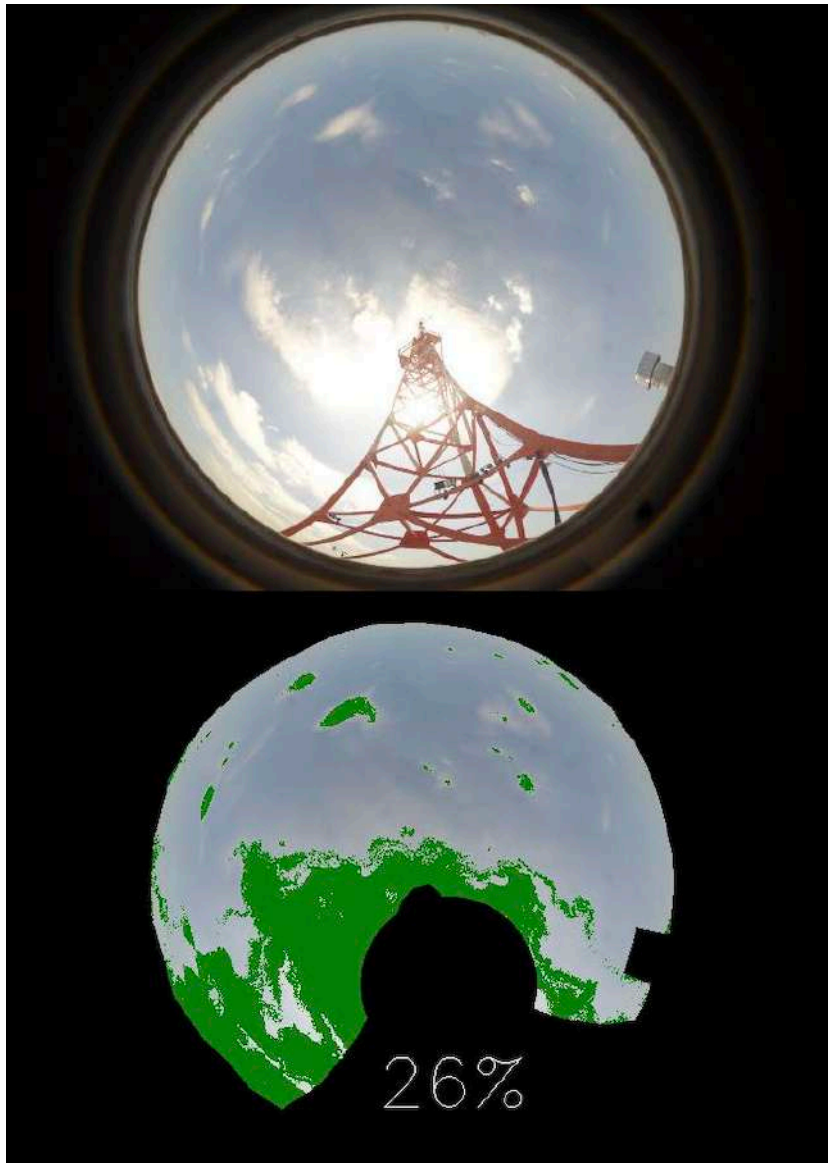
Comparison with RGB composite images

Natural Color RGB composite image using band 3, 2, 1of MSI as R, G, B

Partly Cloudy



● Shibuya, Tokyo



Comparison with RGB composite images

Natural Color RGB composite image using band 3, 2, 1of MSI as R, G, B

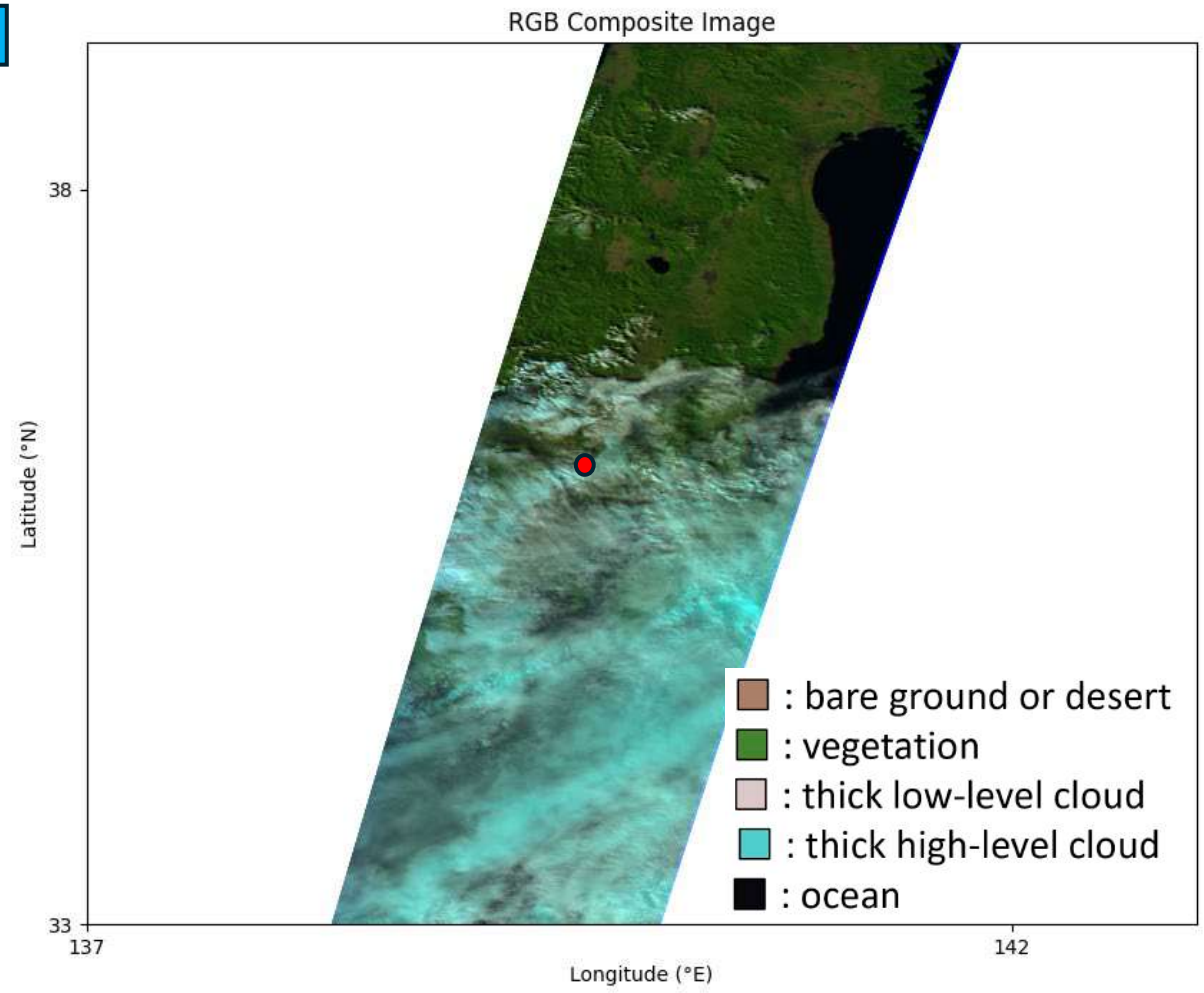
Introduction

Algorithm

Validation

Summary

Cloudy



● Shibuya, Tokyo

Data matchup between MSI and sky camera

Matchup Site List (SkyCamera)

> WINDAS > Radiometer > SkyCamera > BSRN > AD-Net > SAVER-Net > SKYNET > AERONET
> Arctic & Antarctic > Koganei > NIES > Italy

2024 | 9 | L2aMSI | < | > | 1km | Shibuya

https://www.eorc.jaxa.jp/EARTHCARE/EC_matchup2/matchlist/2024/09/muplist_202409_L2aMSI_SkyCamera_1

日時	orbitNo.	frame	地点名	lat (衛星)	lon (衛星)	matchup 距離(km)	仰角 (度)	方位角 (度)	GoogleMap
2024/09/05 05:02:56.585361	01544	D	Shibuya	35.600006	140.080521	36.5	84.46	101.32	■
2024/09/11 16:21:11.166635	01645	B	Shibuya	35.649311	139.585938	9.1	88.62	259.29	■
2024/09/20 16:18:54.815259	01785	B	Shibuya	35.734142	140.083649	36.9	84.41	77.87	■
2024/09/21 05:06:59.809743	01793	D	Shibuya	35.781239	139.011581	62.1	80.63	282.07	■

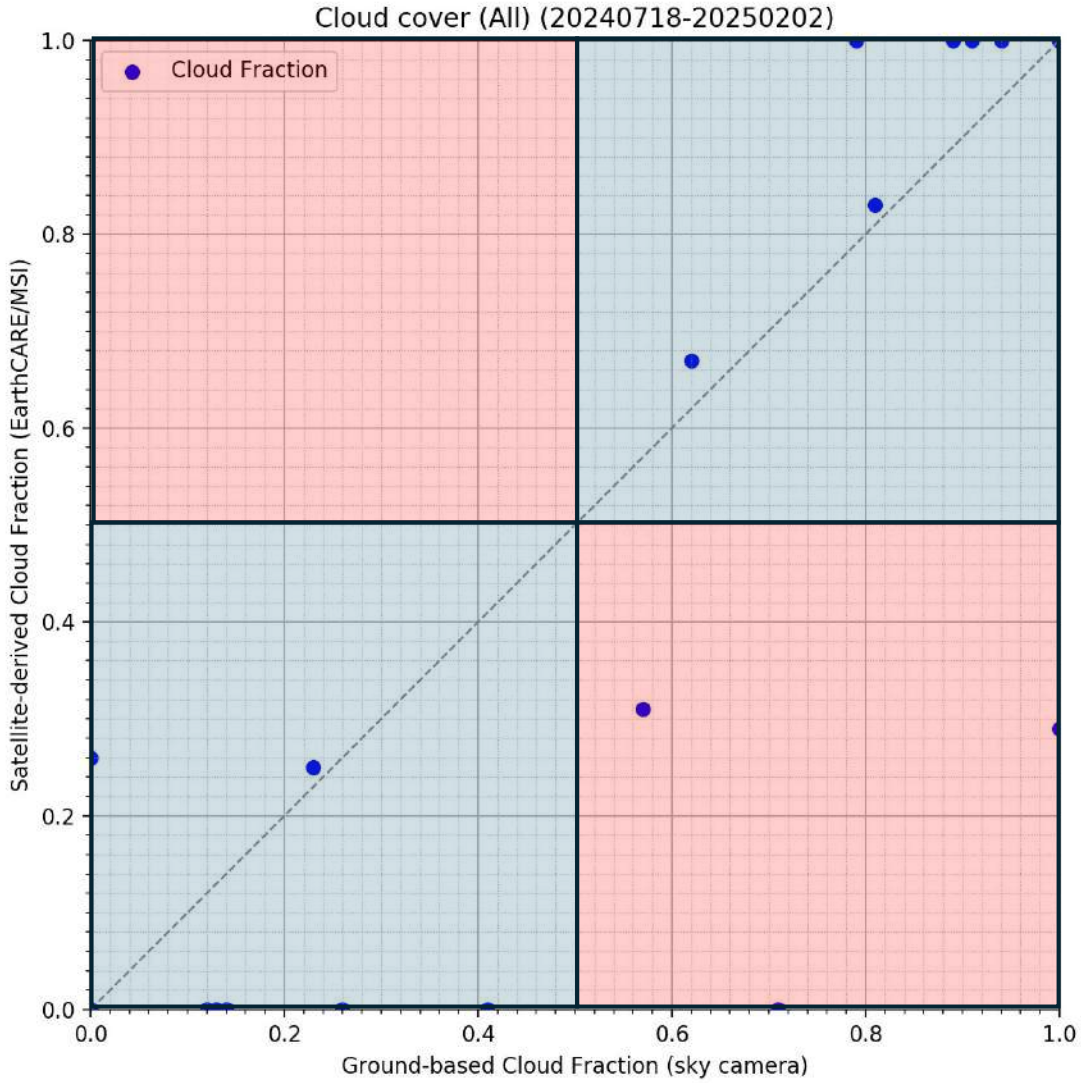
2024 | 10 | L2aMSI | < | > | 1km | Kumamoto

https://www.eorc.jaxa.jp/EARTHCARE/EC_matchup2/matchlist/2024/10/muplist_202410_L2aMSI_SkyCamera_1

日時	orbitNo.	frame	地点名	lat (衛星)	lon (衛星)	matchup 距離(km)	仰角 (度)	方位角 (度)	GoogleMap
2024/10/10 05:40:21.168073	02089	D	Kumamoto	32.979866	130.055405	77.7	78.33	281.88	■
2024/10/16 16:58:24.389506	02190	B	Kumamoto	32.790699	130.613434	24.5	86.28	258.12	■
2024/10/19 05:37:05.261343	02229	D	Kumamoto	32.846355	130.802094	6.4	89.02	280.33	■
2024/10/25 16:54:44.045284	02330	B	Kumamoto	32.936413	131.461273	56.4	81.48	78.57	■

Basically only 2-3 sets of matchup data from each location (Daytime only) per month

Cloud Fraction between MSI and sky camera



Date: 2024/07/18-2025/02/02
Daytime only
Sample from Shibuya: 11
Sample from Kumamoto: 13

Accurate
Inaccurate

		Cloud Fraction from sky images		
		Clear	Cloudy	Total
Cloud Fraction from MSI data	Clear	10	3	14
	Cloudy	0	11	11
	Total	10	14	24

Case All_day

Total Accuracy: 87.5%,
clear accuracy: 100%,
cloudy accuracy: 78.6%

Comparison with Himawari-8

Introduction
Algorithm
Validation
Summary

■ With snow screening

Ocean Total: 1553200		AHI	
		cloudy	clear
MSI	cloudy	717591	131494
		46.2%	8.46%
	clear	106744	597371
		6.87%	38.5%

Match : 84.7%, Mismatch : 15.3%

Land Total: 480512		AHI	
		cloudy	clear
MSI	cloudy	91562	106235
		19.1%	22.1%
	clear	8077	274638
		1.68%	57.2%

Match : 76.3%, Mismatch : 23.7%

- V1.0 MSI_CLP, From 28 Jan 2025 to 6 Feb 2025.
- The original pixels of MSI were averaged to match the spatial resolution: 500m → 5km
- Cloudy/Clear definition
  - MSI CCL: 0~3 → cloudy
  - MSI CCL: 4~7 → clear
  - AHI CM : 2~3 → cloudy
  - AHI CM : 0~1 → clear
- Very close to the JAXA release criteria (Ocean ±15%, Land ±20%)
- For more detailed information, see the presentation of Masataka Muto (next presentation).

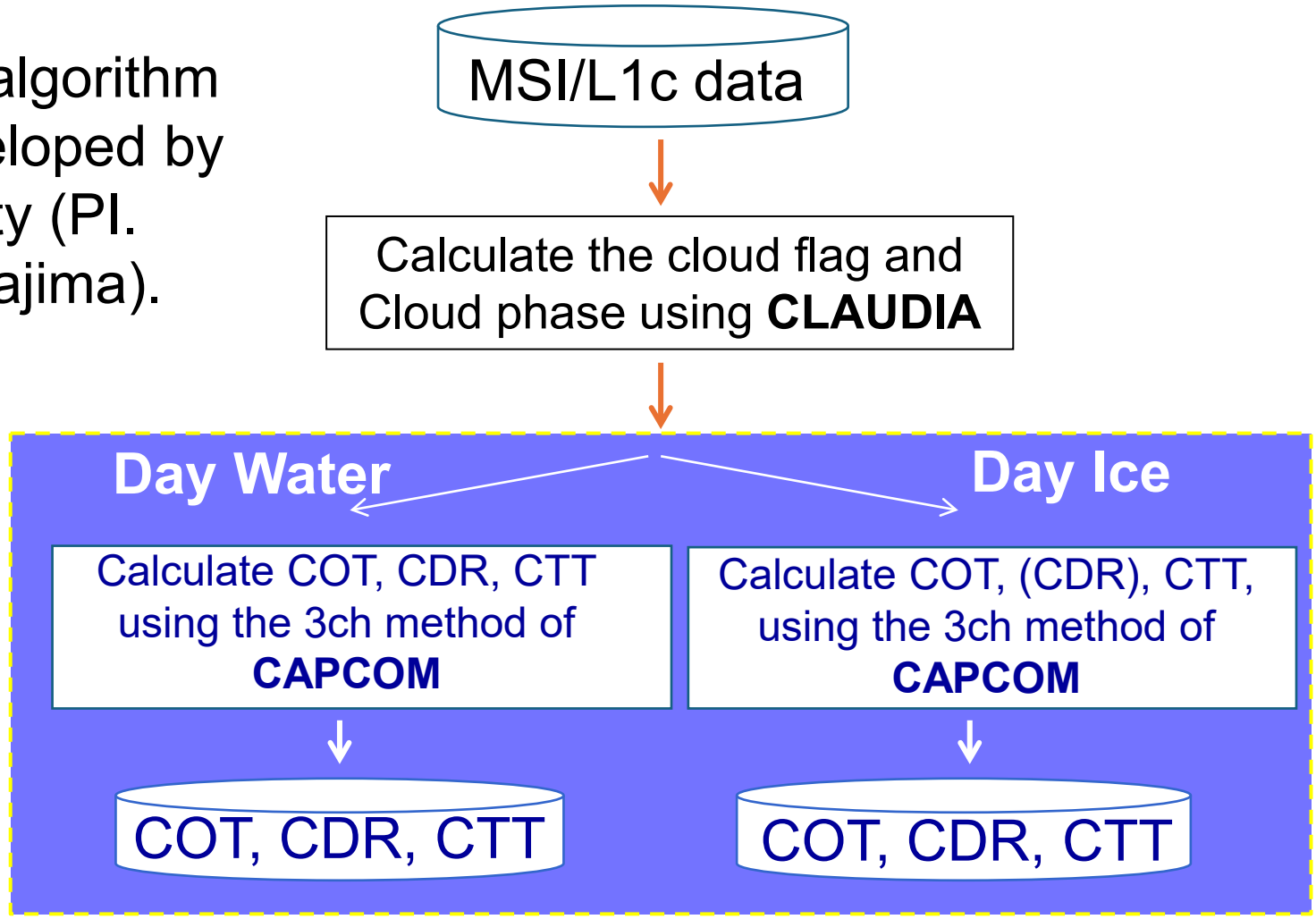
Key points

- MSI_CLP provides high-quality results of clouds (tropical cyclones, hurricanes, typhoons etc.) in low/middle latitude regions, especially when combined with other sensors in EarthCARE.
- Snow cover in high elevation regions and polar regions have a negative effect on the quality of MSI_CLP cloud products.
- Supervised machine learning can be used to determine cloud regions
  - Compared to conventional visual cloud cover determination and threshold method, it can eliminate the subjective and sensory aspects.
- The accuracy of cloud fraction was high in daytime, achieving an overall accuracy of over 85% and very high agreement with MSI's RGB composite images. In addition, the results were close to those obtained with Himawari.

Thanks for your attention!

JAXA EarthCARE/MSI cloud product process

JAXA MSI L2 algorithm has been developed by Tokai University (PI. Prof. T. Y. Nakajima).



Cloud Optical Thickness (COT), Cloud Droplet effective Radius (CDR), Cloud Top Temperature (CTT)

Logistic Regression (LR) method

- A supervised multivariate analysis learning method
- Predicting probabilities from the relationship between explanatory and objective variables
  - Objective variable (value to be predicted) is obtained with probability
  - Example
    - Outbreak of an illness for a patient (yes/no)
    - A custom can repay the loan (yes/no)
    - **The target pixel is cloud (yes/no)**
  - Create an equation to predict the probability (P) of the occurrence of the event set as the objective variable, and calculate P

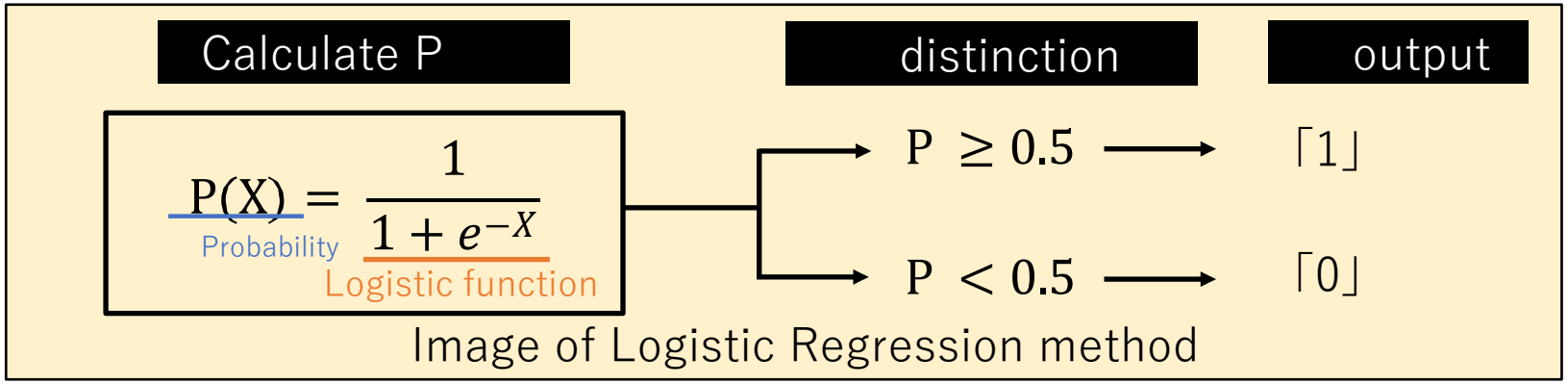
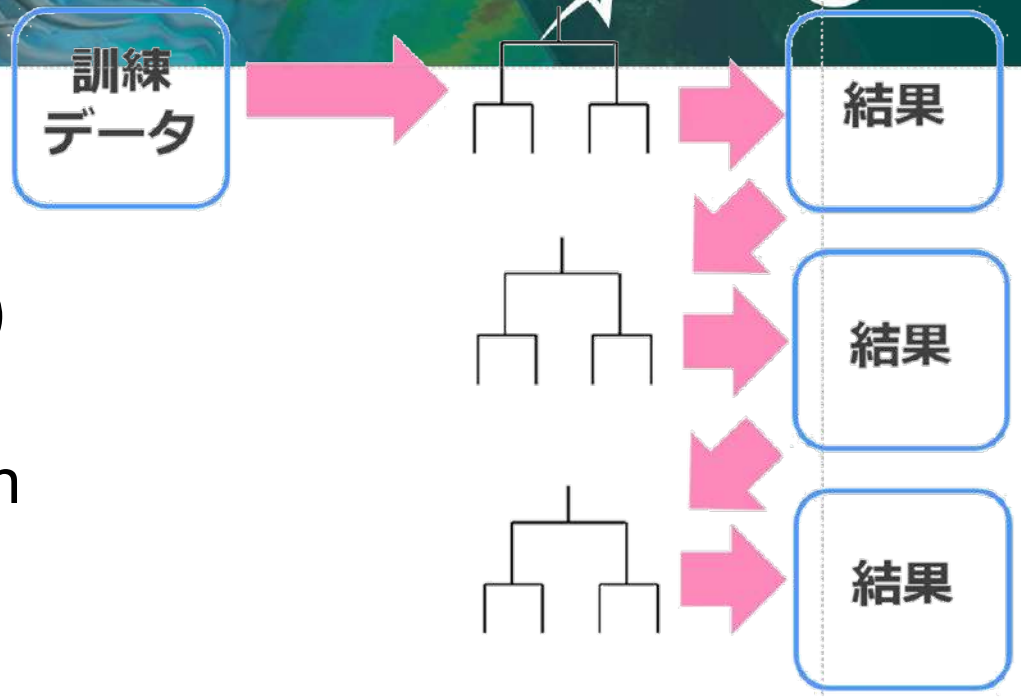
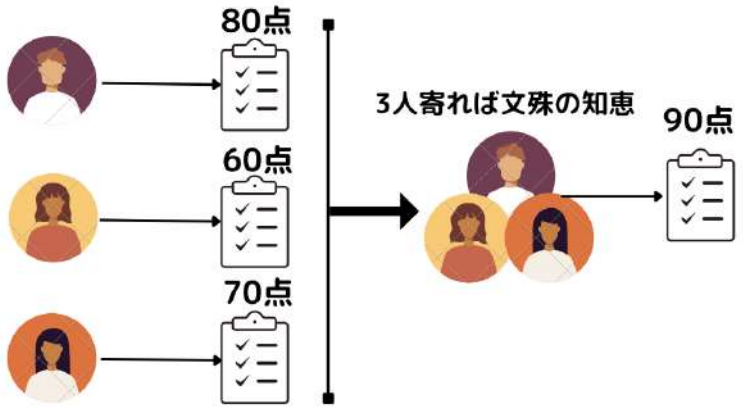


Image of Logistic Regression method

XG-Boost (XG) method



- XG-Boost(eXtreme Gradient Boosting)
  - Nonlinear machine learning algorithm
  - Combines ensemble learning and decision tree methods
- Ensemble learning
  - A method to increase accuracy by using a large number of weak learners that are not enough accurate on their own

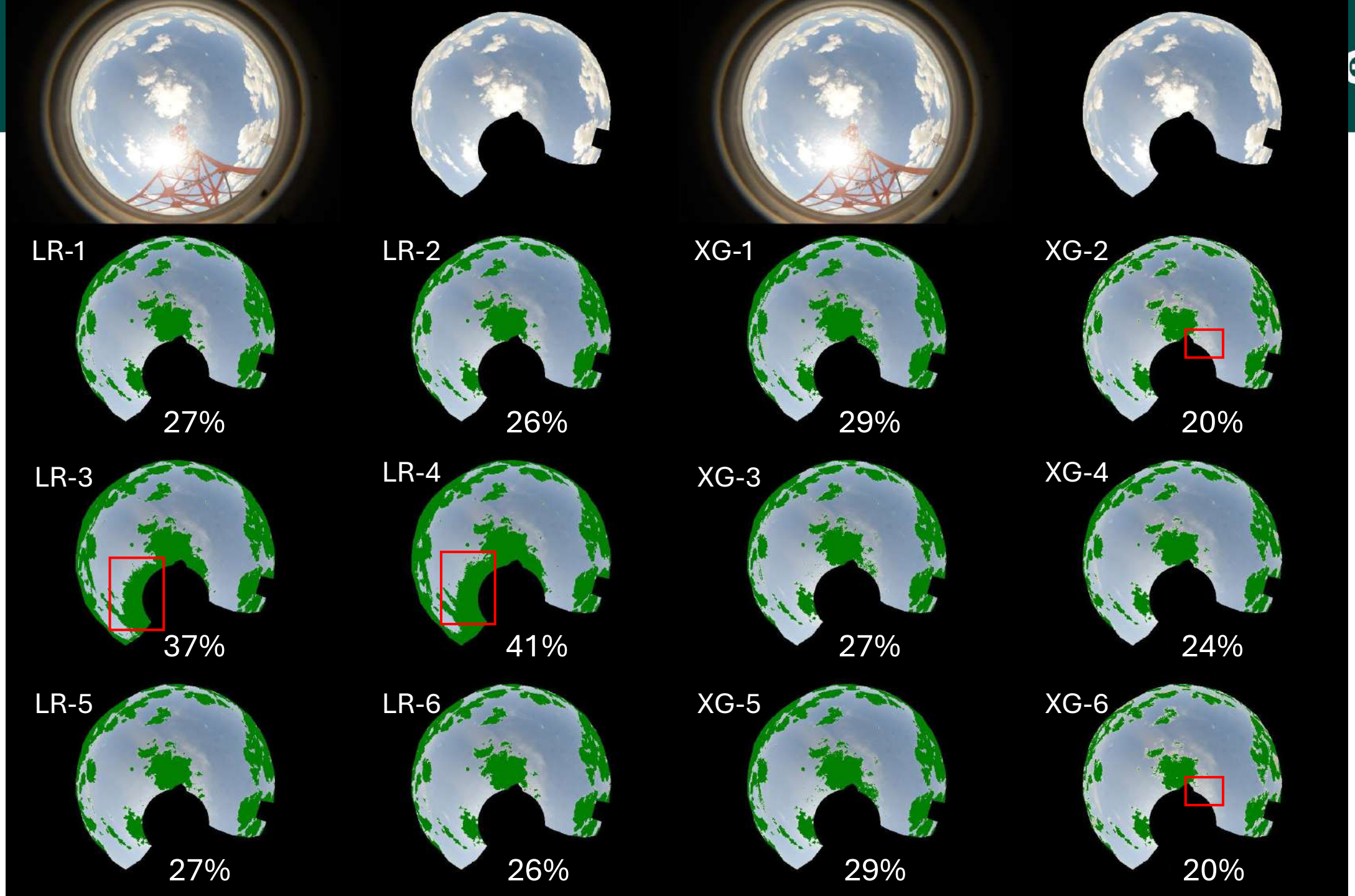


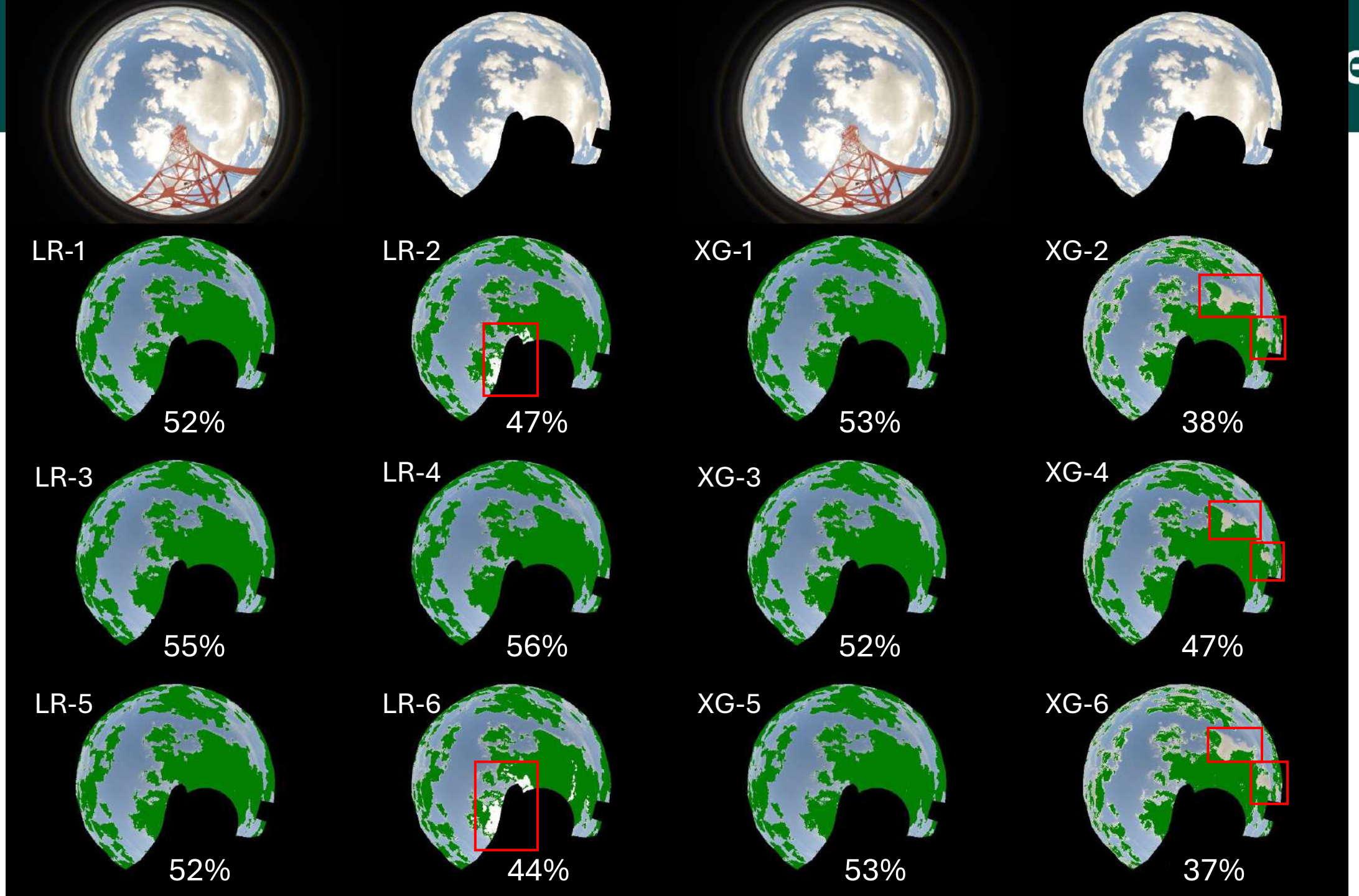
Quoted from : https://toukei-lab.com/xgboost

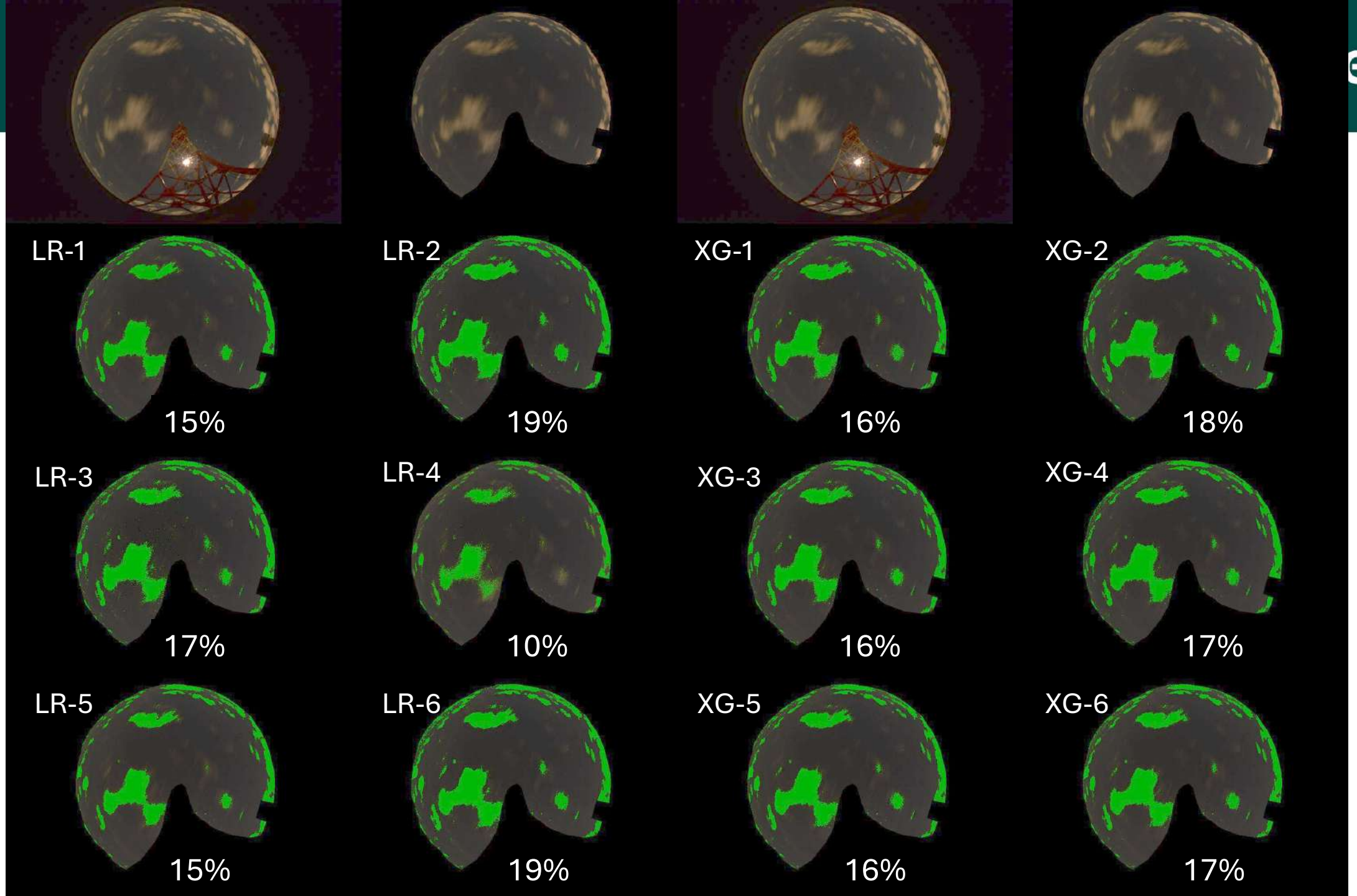
Identification of critical parameters

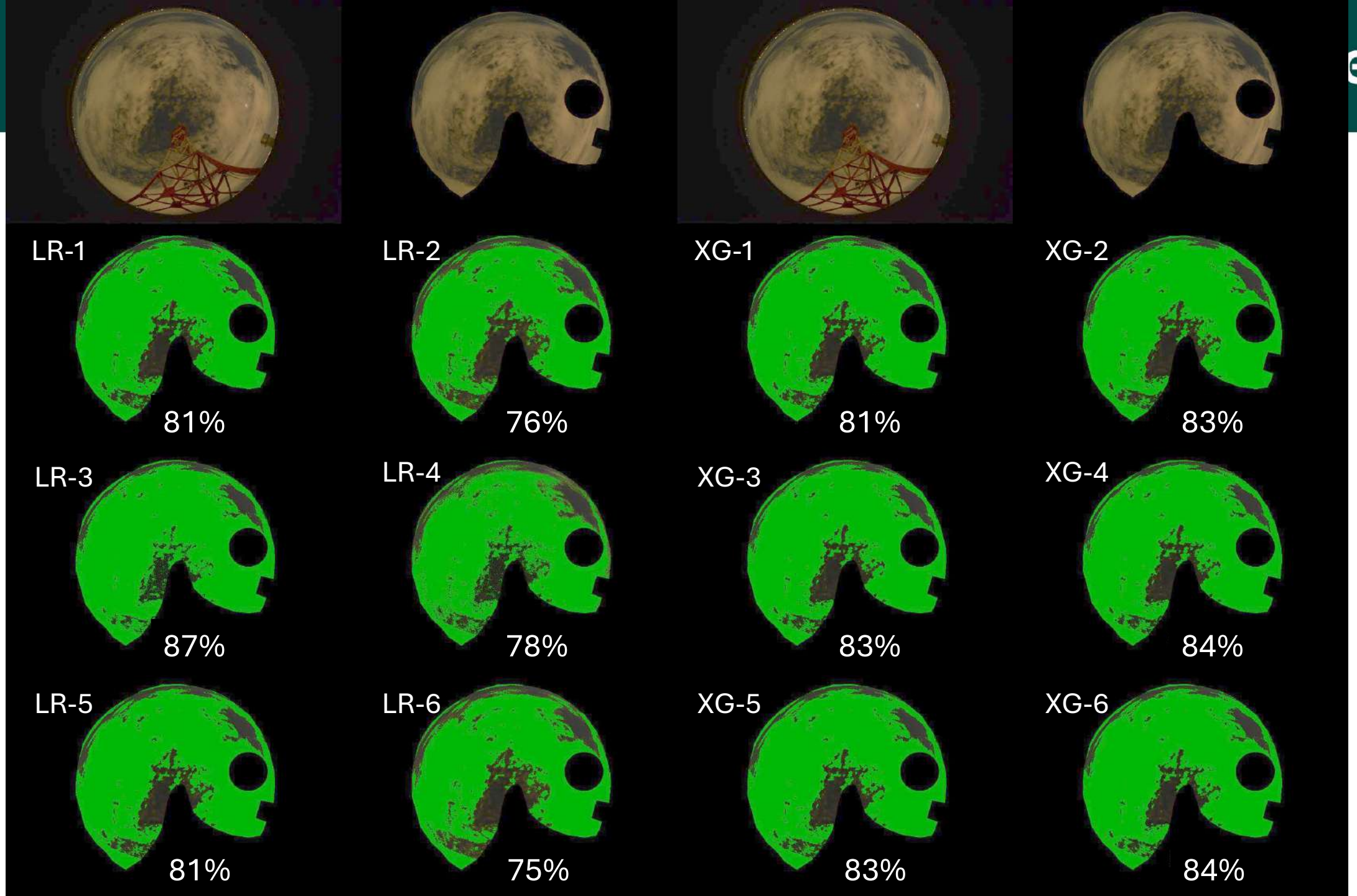
- Identifying important parameters by creating combinations of explanatory variables
  1. [R, G, B]
  2. [R, G, B, R_ave, ,G_ave ,B_ave]
  3. [SI, BI]
  4. [SI, BI, SI_ave, BI_ave]
  5. [R, G, B, SI, BI]
  6. [R. G. B, R_ave, G_ave, B_ave, SI, BI, SI_ave, BI_ave]

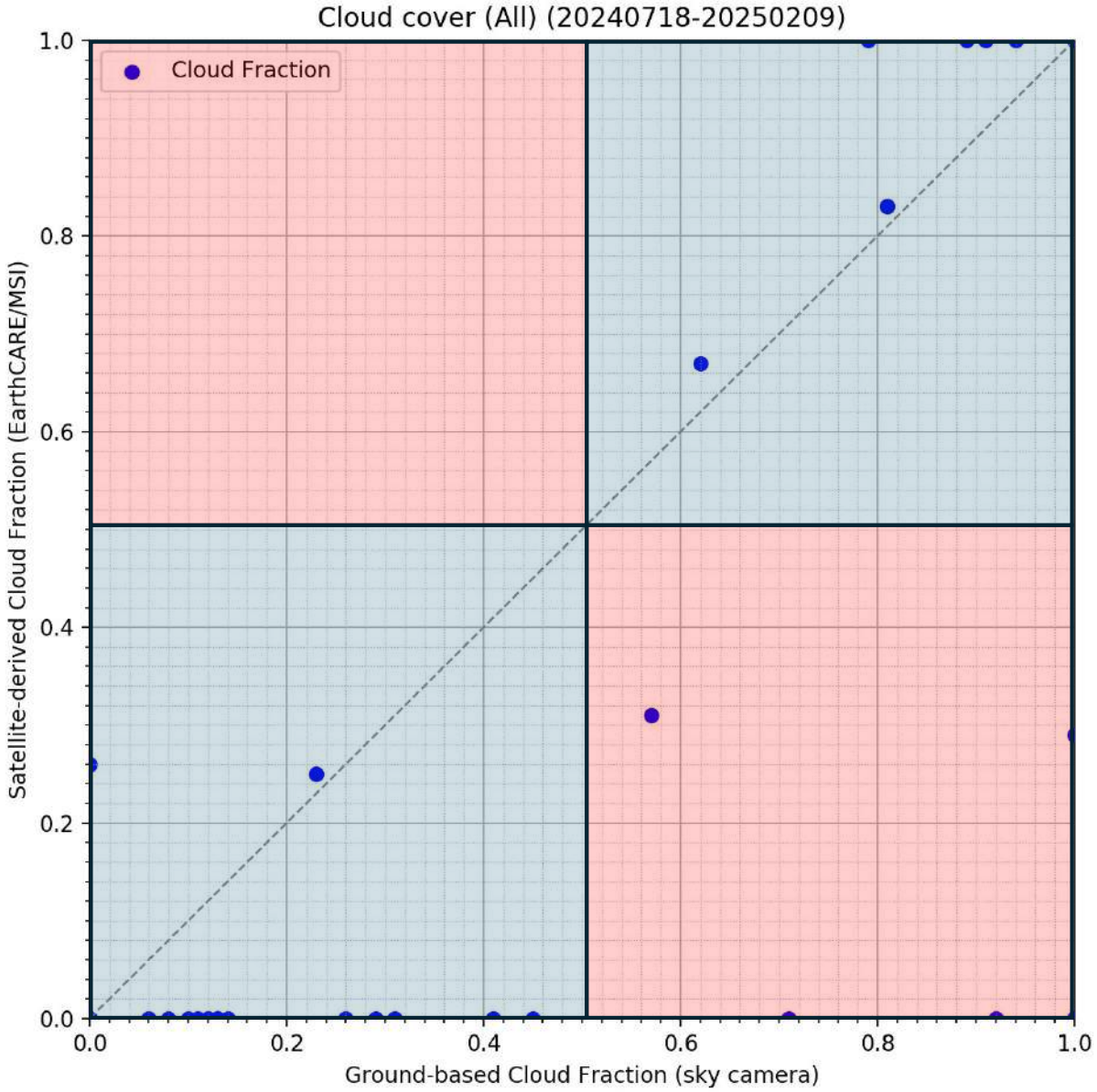
- 6 patterns for each method (LR / XG-Boost)
  - LR-1… LR × [R, G, B]
  - XG-6… XG × [R, G, B, RGB_ave, SI, BI, SIBI_ave]


LR-1
27%
LR-2
26%
XG-1
29%
XG-2
20%
LR-3
37%
LR-4
41%
XG-3
27%
XG-4
24%
LR-5
27%
LR-6
26%
XG-5
29%
XG-6
20%


LR-1
52%
LR-2
47%
XG-1
53%
XG-2
38%
LR-3
55%
LR-4
56%
XG-3
52%
XG-4
47%
LR-5
52%
LR-6
44%
XG-5
53%
XG-6
37%


LR-1
15%
LR-2
19%
XG-1
16%
XG-2
18%
LR-3
17%
LR-4
10%
XG-3
16%
XG-4
17%
LR-5
15%
LR-6
19%
XG-5
16%
XG-6
17%


LR-1
81%
LR-2
76%
XG-1
81%
XG-2
83%
LR-3
87%
LR-4
78%
XG-3
83%
XG-4
84%
LR-5
81%
LR-6
75%
XG-5
83%
XG-6
84%



Date: 2024/07/18-2025/02/09
Both daytime and nighttime
Sample from Shibuya: 23
Sample from Kumamoto: 25

Accurate
Inaccurate

		Cloud Fraction from sky images		
		Clear	Cloudy	Total
Cloud Fraction from MSI data	Clear	26	11	37
	Cloudy	0	11	11
	Total	26	22	48

Case All_daynight

- Daytime + nighttime：
- Total Accuracy: 77.1%, clear accuracy: 100%, cloudy accuracy: 50%

AUC score from ROC curve graph

表 4-10 昼間のパターン別 AUC スコア

	昼	夜
LR-1(RGB)	0.996	0.996
LR-2(RGB, RGB_Ave)	0.998	1.000
LR-3(SIBI)	0.988	0.993
LR-4(SIBI, SIBI_Ave)	0.988	0.994
LR-5(RGB, SIBI)	0.996	0.996
LR-6(RGB, RGB_Ave, SIBI, SIBI_Ave)	0.997	1.000
XG-1(RGB)	0.995	0.994
XG-2(RGB, RGB_Ave)	0.999	0.999
XG-3(SIBI)	0.994	0.999
XG-4(SIBI, SIBI_Ave)	0.999	1.000
XG-5(RGB, SIBI)	0.995	0.999
XG-6(RGB, RGB_Ave, SIBI, SIBI_Ave)	0.999	1.000

K-Fold Cross-Validation

表 4-11 昼間のパターン別 k 分割交差検証スコア

	昼	夜
LR-1(RGB)	0.975	0.980
LR-2(RGB, RGB_Ave)	0.981	0.976
LR-3(SIBI)	0.939	0.976
LR-4(SIBI, SIBI_Ave)	0.932	0.972
LR-5(RGB, SIBI)	0.975	0.979
LR-6(RGB, RGB_Ave, SIBI, SIBI_Ave)	0.983	0.976
XG-1(RGB)	0.962	0.985
XG-2(RGB, RGB_Ave)	0.971	0.991
XG-3(SIBI)	0.973	0.984
XG-4(SIBI, SIBI_Ave)	0.972	0.989
XG-5(RGB, SIBI)	0.972	0.983
XG-6(RGB, RGB_Ave, SIBI, SIBI_Ave)	0.973	0.989