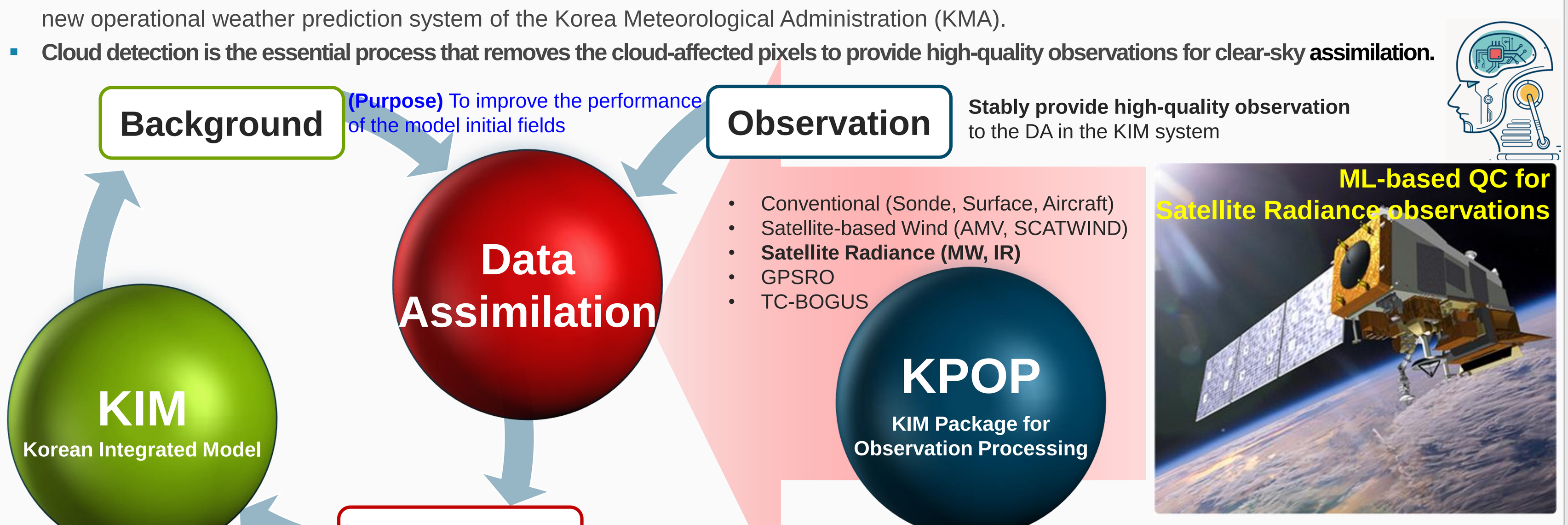
KIAPS
KOREA INSTITUTE OF ATMOSPHERIC PREDICTION SYSTEMS

A Study on Machine Learning-Based Cloud Detection for Microwave Radiance Assimilation

*Eun-jin Kim (ejkim@kiaps.org), Hyeon-Ju Jeon, Jeon-Ho Kang, In-Hyuk Kwon
Data Assimilation Group / Korea Institute of Atmospheric Prediction systems (KIAPS)

1 Introduction

- Microwave radiance observation significantly contributes to improving the accuracy of numerical weather prediction models through Data Assimilation (DA).
- The Korean Integrated Model (KIM) Package for Observation Processing (KPOP) is used to stably provide high-quality observations to DA in the new operational weather prediction system of the Korea Meteorological Administration (KMA).
- **Cloud detection is the essential process that removes the cloud-affected pixels to provide high-quality observations for clear-sky assimilation.**



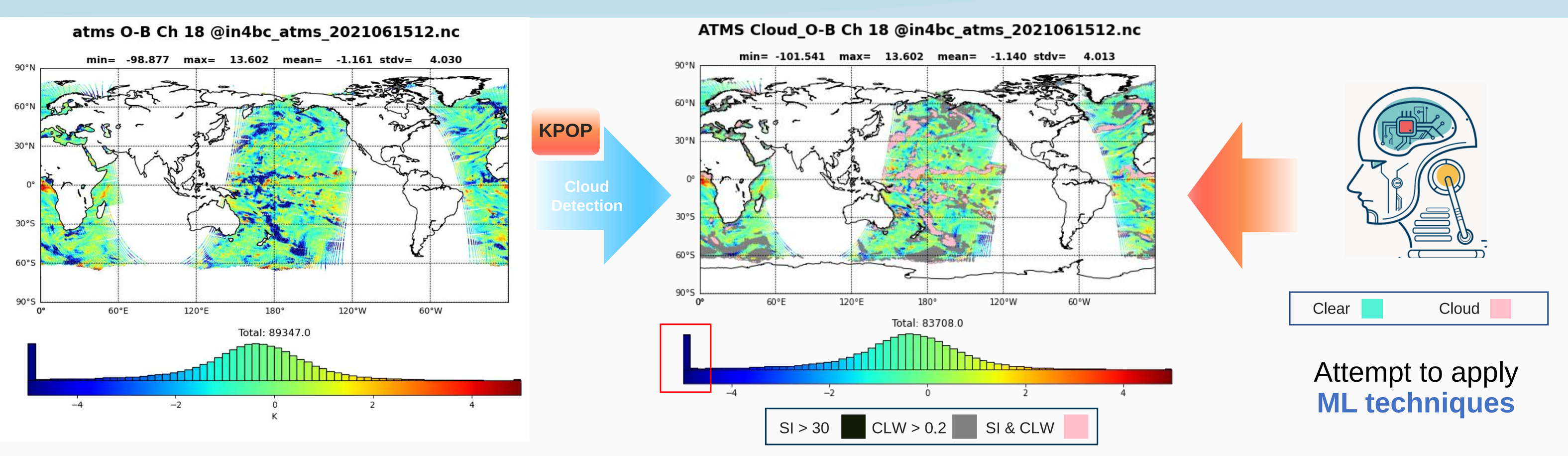
- Satellite observations contribute to improving weather forecast accuracy significantly by providing data in an area that conventional observations cannot cover.
- **KPOP** contains **gross-error QC**, **background ingests** retrieving simulated observation at the observation space from the model grid states, **bias correction, innovation QC, thinning**, and so on.

2 Cloud Detection for ATMS*

*Advanced Technology Microwave Sounder

- The cloud detection for the Advanced Technology Microwave Sounder (ATMS) temperature and humidity sounding channels in the KPOP system is performed based on **the Cloud Liquid Water (CLW), the Scatter Index (SI), and the first-guess departure (Observation-minus-Background, O−B) in the window channel.**
- We found that **the cloud-contaminated pixels remained even after the cloud screening** through the KPOP system. This indicates that it needs to adjust the threshold values of the indices for cloud detection through the challenging process requiring numerous sensitivity experiments for optimization.
- Recently, cloud detection techniques have been developed in many studies based on support vector regression, decision tree, and deep learning in the field of satellite data utilization.
- This study aims to develop and compare **unsupervised and supervised learning-based cloud detection** techniques to provide DA systems with high-quality observations of ATMS.

Quality Control Settings for ATMS in KIAPS, ECMWF, UM



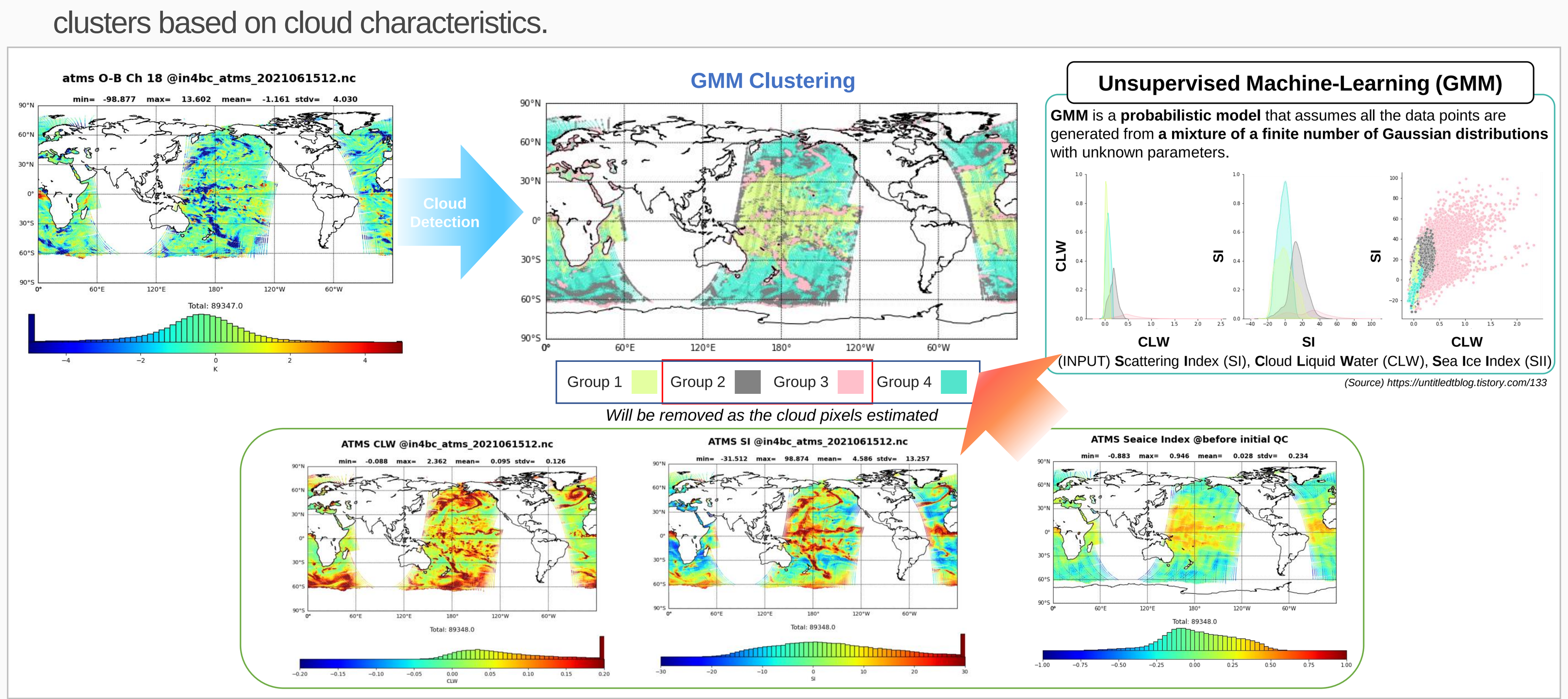
✓ There are slight differences in the way the index is calculated for each institution, and also in the channels and thresholds for applying cloud detection.

Criterion		KIAPS	ECMWF *Bennartz et al. (2002)*	UM
Scatter Index (SI)	Equation	TB(89 GHz) −TB(166 GHz) −(−46.94+0.248xSZA)		TB(89 GHz) −TB(166 GHz) −(−35.60+0.14xSZA)
	Threshold	SI > 30	SI > 10	SI > 10
	Remarks	Ch 6, Ch 18~20	Ch 18~22	Ch 6~9, 18~22
Cloud Liquid Water (CLW)	Equation	cosza x [aa + 0.754 x log(285.0 − TB(23.8 GHz, Ch1)) − 2.265 x log(285.0 − TB(31.4 GHz, Ch2))] aa = 8.240 − (2.622−1.846 x cosza) x cosza, cosza = cos(Satellite Zenith Angle * d2r), d2r = π / 180.0		
	Threshold	CLW > 0.2	CLW > 0.12 (Ch 6, 7, 18) CLW > 0.15 (Ch 8)	
	Remarks	Ch 6~7, 18~22	Ch 6, 7, 8, 18	
Window channel departure	threshold	$\lvert (O-B)_{50.3\ GHz} \rvert > 5$ K	$\lvert (O-B)_{50.3\ GHz} \rvert > 5$ K	
	Remarks	Ch 6~7, 18~22	Ch 6~8, 18~22	

▲ Hyeyoung Kim (2019. 10. 1.) "ATMS static bias correction and QC adjustment"

3 Gaussian Mixture Model (GMM)

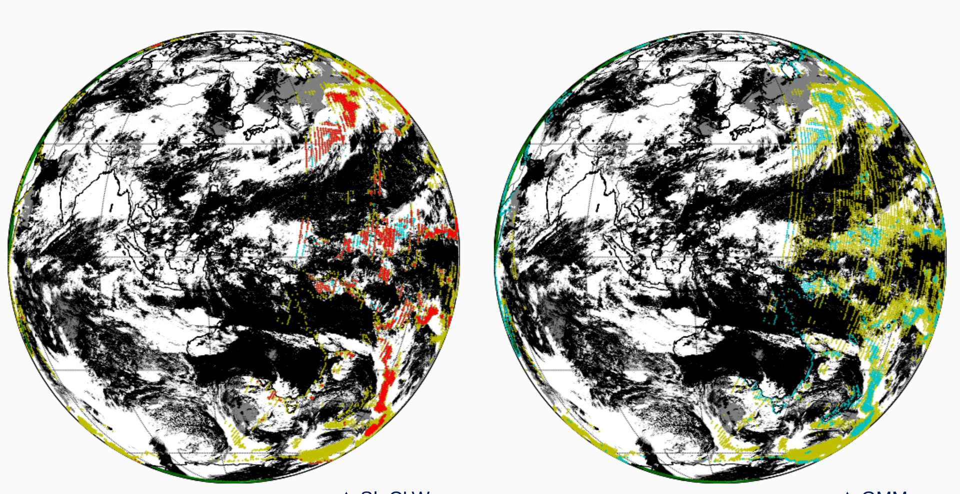
- We applied **unsupervised learning of the Gaussian Mixture Model (GMM)** which divides the entire dataset into **a fixed number of clusters using the multiple Gaussian distributions.**
- CLW, SI, and Sea Ice Index (SII) were used as the input parameters, and ATMS observations were classified into four clusters based on cloud characteristics.

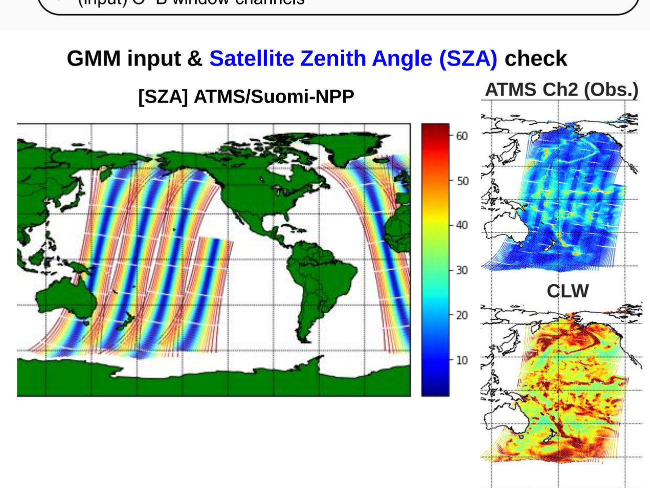


- Cloud detection using GMM & Limitations

✓ We performed various tests to improve the clustering model (GMM) ➔

- n_components = 4~7 TEST
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
- Bayesian Gaussian Mixture (BGM)
- Remove edge pixels of scan position
- (input) O−B window channels





Comparison of KPOP and GMM cloud detection results (2021. 06. 15. 12UTC)

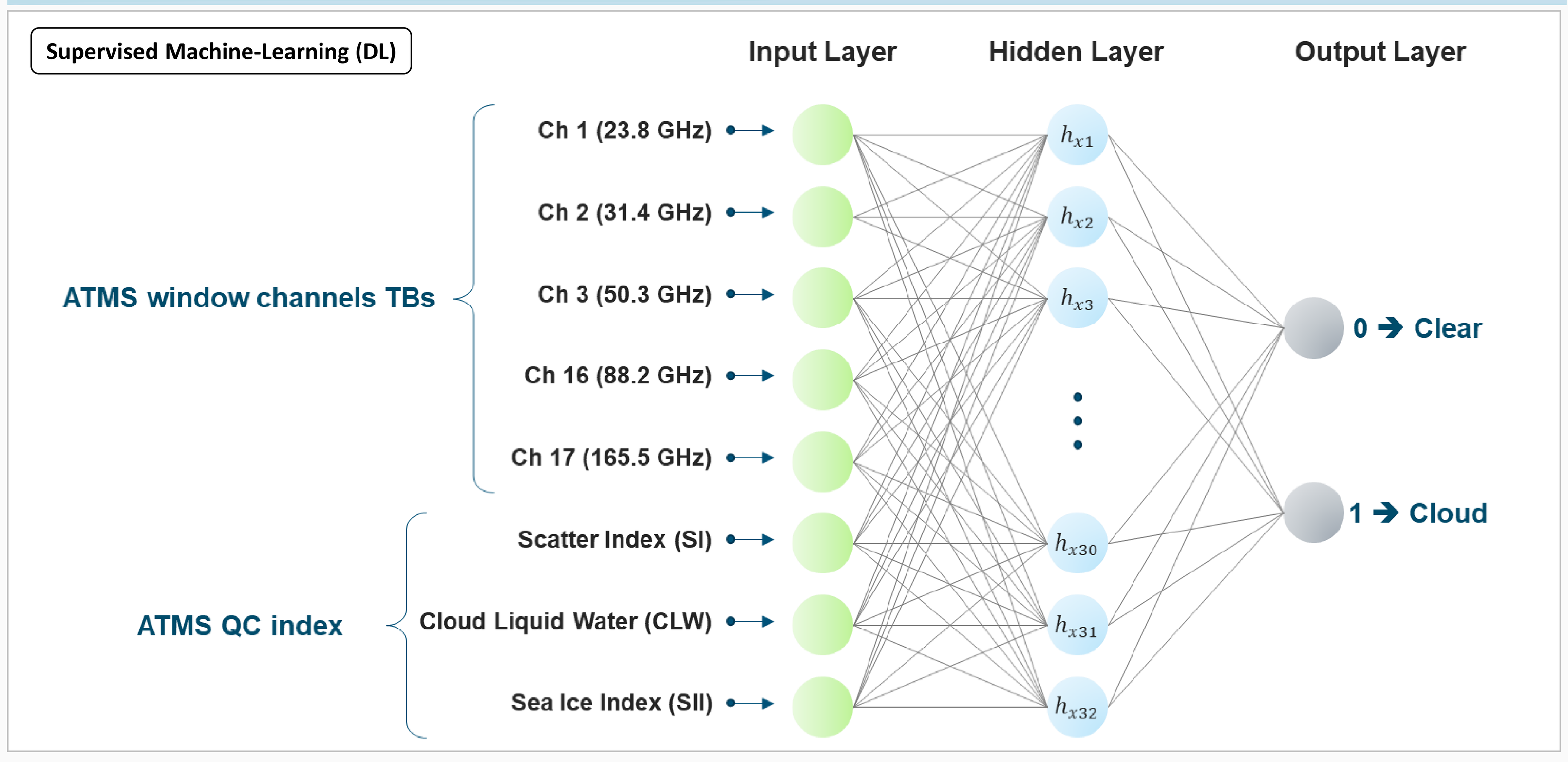
- Due to the observation characteristics of the ATMS satellite, there are many pixels classified as clouds in the area corresponding to the scan edge moving away from the nadir.
- When calculating CLW and SI, a term that considers the Satellite Zenith Angle (SZA) is included to correct it.
- This problem appears to remain because data before scan bias correction (BC) of actual ATMS satellite data is used.

Attempt to apply Deep Learning (DL) techniques

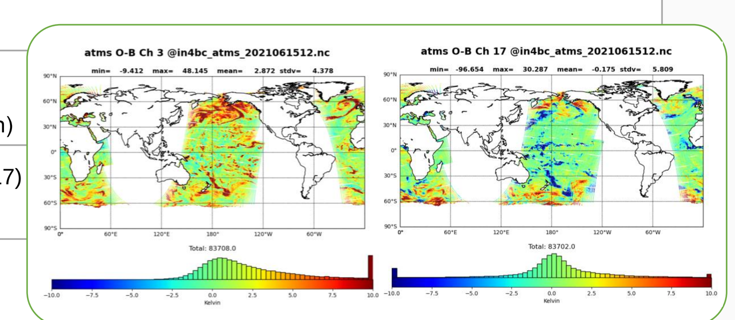
4 Deep Learning Model (binary classification)

- Also, we developed the cloud detection technique using **the supervised learning-based binary classification model.**
- The binary classification model, a **Neural Network (NN) consisting of input, hidden, and output layers, classifies the output data into two classes (0 or 1).**
- In this study, using CLW, SI, and O−B in the window channel of ATMS, the supervised learning model was developed that **conservatively distinguishes clear and cloud regions based on the KPOP results.**

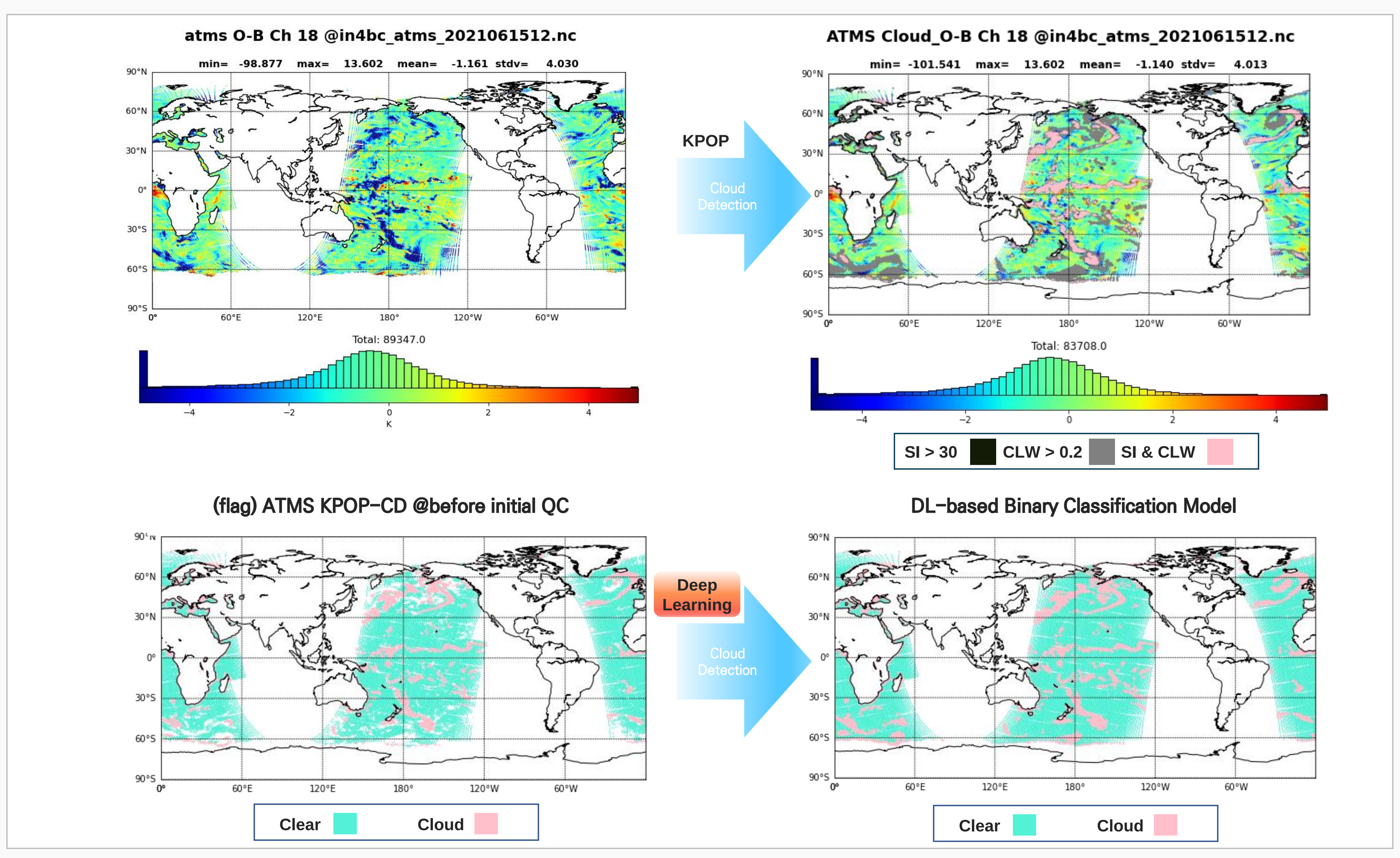
Binary Classification Model



Data type	• Data containing QC index information prior to cloud detection (in4bc_atms_$date.nc)
Dataset (7,229,102)	• (Training & Test) 2022. 07. 01. 00UTC~2022. 07. 25. 18UTC (25 days)
Pre-processing	• Remove missing values, classify land/ocean (sfctype = 1) • Z-score Standardization : z = (x - μ) ÷ σ (μ: mean, σ: standard deviation)
Flag Criteria (Cloud / Clear)	• Clear (0) / Cloud (1) ➔ KPOP Results & O-B of window Channels (3, 17) • Convert array to binary : Y_encoded = to_categorical(y_train)
Model Composition	• **Deep Learning Framework : Tensorflow 2.7.0** • 3 layers (node = 8 → 32 → 2) • Activation function (hidden layer: relu, ouptut layer: softmax), loss function (binary_crossentropy)



DL-based Cloud Detection

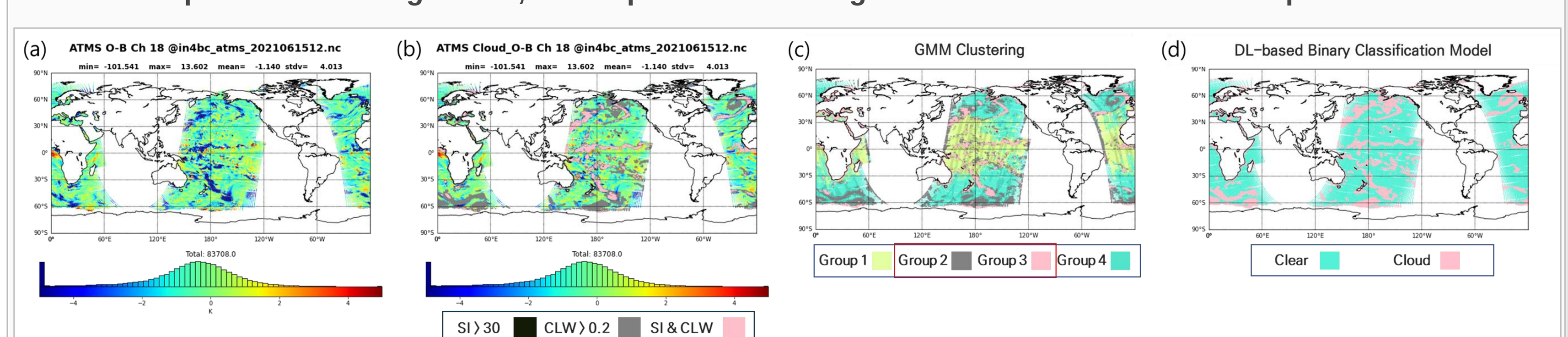


Comparison of KPOP and DL-based Binary Classification Model (2021. 06. 15. 12UTC)

5 Summary and Future Plans

Development of the ML-based Cloud Detection Techniques for the Microwave Satellite Radiance DA

- We evaluated the classified the cloud area during the training and validating processes and confirmed the **feasibility of ML-based ATMS cloud detection** by conducting a comparative study with conventional, **unsupervised learning-based, and supervised learning-based cloud detection techniques.**



Index	KPOP-ATMS				Gaussian Mixture Model (GMM)								DL-based Binary Classification Model			
	Clear		Cloud		Group1		Group2		Group3		Group4		Clear		Cloud	
	mean	stdv	mean	stdv	mean	stdv	mean	stdv	mean	stdv	mean	stdv	mean	stdv	mean	stdv
CLW	0.066	0.056	0.379	0.230	0.032	0.031	0.178	0.090	0.556	0.291	0.069	0.049	0.067	0.057	0.385	0.232
SI	1.217	10.130	20.902	14.481	-0.075	10.288	13.931	10.158	24.882	20.719	-1.882	6.733	1.207	9.976	21.738	14.588
Count	70457		13251		24845		21505		5056		32302		70961		12747	

Future Plans

- Establishment the new ML-based Cloud Detection system within the KPOP system [Fortran ↔ Python]
- Verification of the DA performance through the analysis-forecast cycle with these techniques.
- Study on field application : Statistical analysis of the SI, CLW ➔ Reset the threshold & Test

4th ESA-ECMWF Workshop on Machine Learning for Earth Observation and Prediction
ESA-ESRIN, Frascati, 7 - 10 May 2024