Improving Infrared Rainfall Retrieval Using Machine Learning–Based Probabilistic Cloud Classification

Dong-Cheol Kim and Dong-Bin Shin
Department of Atmospheric Sciences, Yonsei University, Seoul, South Korea

INTRODUCTION

- Precipitation retrieval from geostationary satellite infrared (IR) observations remains a fundamental challenge due to the indirect relationship between cloud-top radiative properties and surface rainfall.
- The Geostationary-Korean Multi-Purpose Satellite-2A (GK2A) rainfall rate (RR) algorithm classifies five precipitating cloud types (PCTs) by combining brightness temperature differences (BTDs) from multiple IR channels and applying optimized thresholds (Shin *et al.*, 2025)
- This study presents a machine learning–based probabilistic framework for improving RR retrieval from the GK2A satellite using a Gaussian Mixture Model (GMM) (Kim and Shin, 2026).
- The proposed approach replaces threshold-based PCT classification with hierarchical multi-stage clustering to better represent precipitation variability.

METHODS

- GMMs are capable of distributional feature extraction from unlabeled datasets through the application of multiple Gaussian distributions.
- A Gaussian Mixture Model is a weighted sum of ***K***th component Gaussian densities.

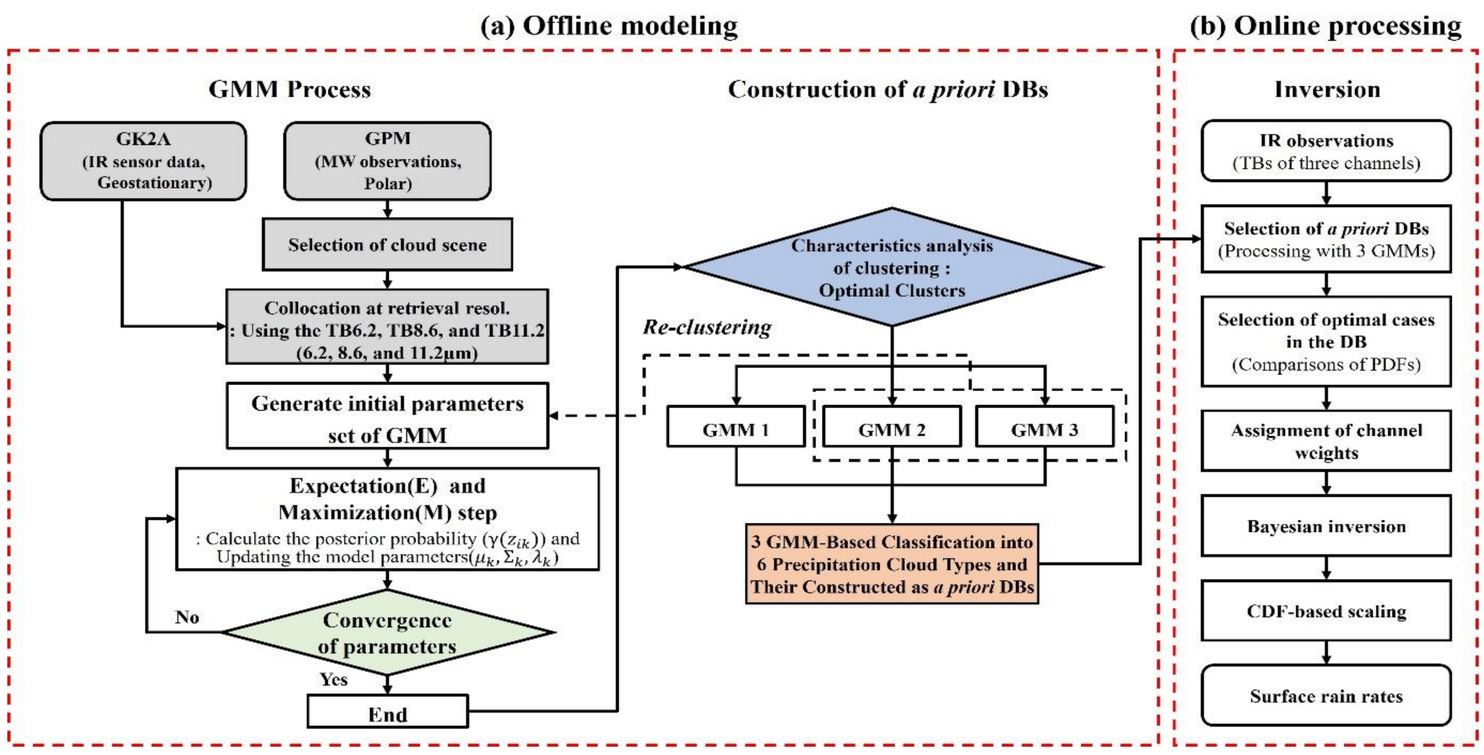
$$p(X) = \sum_{k=1}^{K} \pi_k N(X|\mu_k, \Sigma_k)$$

- **X** : The data point vector (Set of brightness temperatures from AMI).
- ***K*** : The number of component (hyperparameter).
- π_k : Mixing coefficient (weights).

$$0 \le \pi_k \le 1 \text{ and } \sum_{K=1}^{K} \pi_k = 1$$

- μ_k, Σ_k : The mean vector and covariance matrix of the ***K***th component.

GMM-RR Framework



A. OFFLINE MODELING—GMM CONSTRUCTION

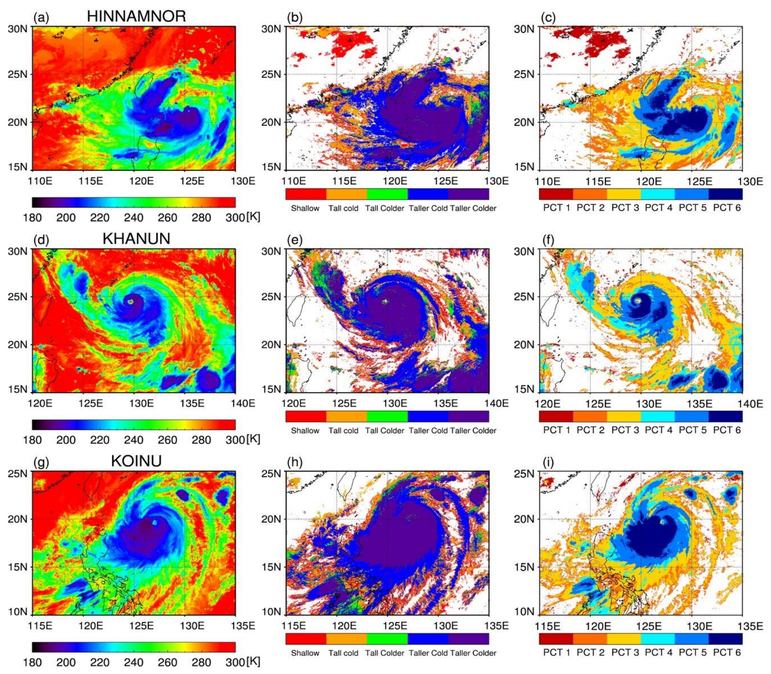
- Train three GMMs using the DPR–GK2A collocation database.
- Classify precipitating clouds into six probabilistic PCTs through multi-stage re-clustering.
- Construct PCT-specific *a priori* databases for rainfall retrieval.

B. ONLINE PROCESSING—RR RETRIEVAL

- Classify precipitating clouds probabilistically from real-time GK2A IR observations.
- Select the corresponding PCT-specific *a priori* database.
- Retrieve rainfall rates using the Bayesian retrieval framework.

RESULTS

PRECIPITATING CLOUD-TYPE CLASSIFICATION

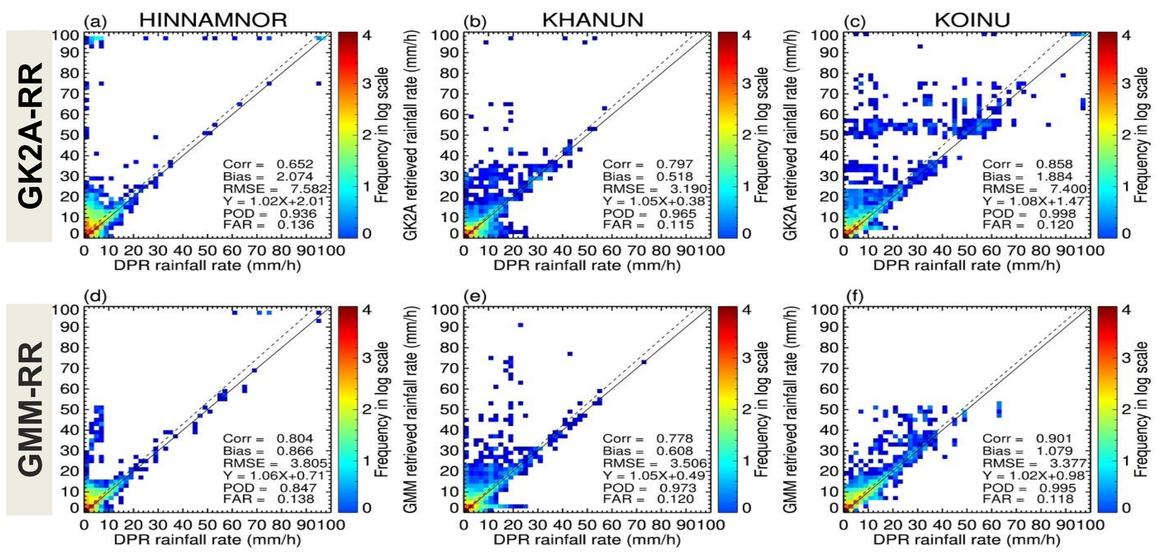


- Comparison of PCT classifications between the operational GK2A algorithm and the GMM-based approach for three typhoon cases (HINNAMNOR, KHANUN, and KOINU).
- The GMM-based framework provides a more detailed and physically consistent PCT classification than the operational threshold-based algorithm.
- It progressively separates outer and inner convective regions of typhoons into six PCTs and clearly delineates typhoon-eye structures and intense precipitation regimes.

PCTs	Scalar accuracy						Categorical accuracy			
	Corr		Bias		RMSE		POD		FAR	
	DPR	GMI	DPR	GMI	DPR	GMI	DPR	GMI	DPR	GMI
PCT 1	0.41	0.35	0.32	0.33	1.29	0.93	0.93	0.98	0.27	0.26
PCT 2	0.50	0.40	0.15	0.13	0.68	0.46	0.98	0.92	0.28	0.32
PCT 3	0.43	0.30	0.26	0.25	1.41	1.16	0.96	0.91	0.26	0.22
PCT 4	0.44	0.33	0.51	0.66	2.12	1.78	0.98	0.96	0.25	0.17
PCT 5	0.53	0.47	2.14	2.04	6.30	5.81	0.99	0.99	0.15	0.08
PCT 6	0.64	0.64	6.27	4.97	13.77	12.60	0.99	0.99	0.10	0.05

- Rainfall retrieval accuracy differs considerably across the six PCTs, reflecting differences in the TB–RR relationships and cloud microphysical characteristics.
- PCTs associated with colder cloud tops and stronger precipitation, particularly PCTs 5 and 6, show the highest agreement with DPR-RR and GMI-RR.

TYPHOON CASE VALIDATION USING DPR-RR



- GK2A-RR tends to overestimate light and high-intensity rainfall, whereas GMM-RR exhibits reduced overestimation and closer agreement with the 1:1 reference line.
- For all three typhoon cases, GMM-RR shows higher correlation and lower error statistics than GK2A-RR when compared with DPR-RR.

CONCLUSIONS

- A GMM-based probabilistic classification framework was developed to replace the threshold-based PCT classification in the operational GK2A-RR algorithm.
- Three-stage GMMs classify precipitating clouds into six PCTs using only three IR TB inputs and construct a refined a priori database for rainfall retrieval.
- Validation against DPR, GMI, IMERG, and ground-based radar showed that probabilistic GMM-based PCT classification better represents precipitation-system variability and leads to more accurate and consistent rainfall estimation.

- ***D.-B. Shin, D.-C. Kim, and D. So, "Rainfall Rate Measurement for Advanced Meteorological Imager of the GEO-KOMPSAT-2A Satellite," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 18, pp. 15725–15739, Jun. 2025.***
- ***D.-C. Kim and D.-B. Shin, "Enhancing GK2A Rainfall Retrieval with GMM-Based Classification of Precipitating Cloud Types," IEEE Trans. Geosci. Remote Sens., early access, doi: 10.1109/TGRS.2026.3704742.***

- ***This work was funded by the Korea Meteorological Administration Research and Development Program under Grant (RS-2025-02221093).***

(Corresponding author: Dong-Bin Shin. E-mail: dbshin@yonsei.ac.kr or mr_cheol@yonsei.ac.kr)