

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

Dong-Cheol Kim and Dong-Bin Shin

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. This study presents a machine learning–based probabilistic framework for improving rainfall rate (RR) retrieval from the GEO-KOMPSAT-2A (GK2A) satellite using a Gaussian Mixture Model (GMM). The proposed approach replaces threshold-based classification of precipitating cloud types (PCTs) with a hierarchical multi-stage clustering strategy designed to better represent the complexity and variability of precipitation systems.

Three sequential GMMs are applied using physically relevant IR features, including brightness temperature (TB) and brightness temperature differences (BTDs), to progressively refine cloud classification. This multi-stage framework enables the identification of six distinct PCTs, each associated with dedicated *a priori* databases representing specific precipitation characteristics. By integrating these probabilistic classifications into the operational GK2A RR algorithm, the proposed GMM-based rainfall retrieval (GMM-RR) enhances the physical consistency of cloud representation and reduces ambiguity in intermediate cloud regimes.

The performance of GMM-RR is evaluated using comprehensive multi-scale validation against satellite-based observations (DPR and GMI), ground-based radar measurements, and the IMERG precipitation dataset. Results from multiple typhoon case studies and seasonal analyses indicate that GMM-RR consistently reduces systematic overestimation in high-intensity rainfall and improves agreement with reference datasets. Additional evaluation at monthly scales demonstrates improved consistency, with reduced bias and error metrics and increased correlation relative to the operational GK2A-RR. Quantitative accuracy is improved while maintaining comparable precipitation detection capability.

These results indicate that machine learning–based probabilistic cloud classification provides a robust and scalable approach for improving IR-based rainfall retrieval. In particular, the proposed framework achieves enhanced accuracy and consistency using only three TB channels, highlighting its efficiency and potential for practical application under diverse observational conditions.