Conference:
12th Workshop of International Precipitation Working Group; July 7-10, 2026, Kraków Poland

Title:
Decomposing Uncertainty in ML-Based Precipitation Type Retrievals

Authors:
Veljko Petković[1], Marko Orescanin[2], Barret Manfre[2], Malarvizhi Arulraj[1], Lei Ji[1]

Affiliations:
1: University of Maryland, College Park, MD, USA
2: Naval Postgraduate School, Monterey, CA, USA

Abstract:
As AI-based precipitation retrievals move into operational use, transparent uncertainty characterization is essential for assessing algorithm maturity, ensuring cross-agency comparability, and building user confidence. Yet most ML models provide point estimates without meaningful confidence measures. Here we demonstrate a Bayesian deep learning approach, a ResNet with Monte Carlo dropout, applied to GPM Microwave Imager (GMI) convective/stratiform precipitation classification, showing that this framework yields uncertainty estimates that are physically interpretable, spatially coherent, and informative for retrieval improvement. The method decomposes total predictive uncertainty into its aleatoric (inherent data ambiguity) and epistemic (model knowledge gap) components, offering distinct and complementary diagnostic information. Analysis of global GMI observations demonstrates the value of this decomposition. Aleatoric uncertainty dominates total uncertainty by approximately two orders of magnitude, revealing that the fundamental ambiguity of passive microwave signatures, where two different precipitation types produce similar brightness temperature signatures, is the primary limitation, rather than insufficient model training. Both uncertainty components exhibit strong regime dependence, directly identifying training data gaps, such as those in high-latitude precipitation regimes, which are known to pose a challenge for passive microwave retrievals (e.g., mixed-phase processes, frozen surfaces, light precipitation). The posterior predictive distributions are found to be consistently unimodal with moderate logit-space spread, suggesting stable performance suitable for operational use. These results illustrate the benefits of decomposed uncertainty quantification for ML-based retrievals, with aleatoric dominance pointing to the need for richer information content rather than simply more training data, and epistemic contribution indicating where training data augmentation would be most impactful.