

Identifying Ice Crystal Chain Aggregates in Cold-Season Storms: Leveraging XGBoost to Map Occurrence and Distribution

Christian M. Nairy¹, David J. Delene¹, Shawn W. Wagner¹, Joseph A. Finlon^{2,3}, John E. Yorks²

¹University of North Dakota, Department of Atmospheric Science, Grand Forks, North Dakota, USA

²National Aeronautics and Space Administration (NASA), Goddard Space and Flight Center, Greenbelt, Maryland, USA

³Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, Maryland, USA

In-situ observations of electrically induced aggregation of cloud ice and frozen droplets have primarily been observed in mid- to upper-level clouds of summertime storms. These aggregates, distinguished by their elongated, quasi-linear structure, are specifically termed as chain aggregates. Cloud chamber experiments reveal that chain aggregation is temperature-dependent, and their formation is enhanced in an electric field exceeding approximately 60 kV m^{-1} . However, various difficulties arise when connecting the laboratory experiments to in-situ observations. While there is evidence that significant electric fields are required for chain aggregate formation, the precise locations and the mechanisms for chain aggregation within storms remain poorly understood. This knowledge gap hinders the accurate parameterization of chain aggregate formation processes in cloud models, impacting precipitation formation, radiative transfer, remote sensing retrievals, and precipitation forecasting.

During NASA's Investigation of Microphysics and Precipitation for Atlantic Coast-Threatening Snowstorms (IMPACTS) field campaign, chain aggregates were observed in 30 of 34 research flights, across temperatures from -38.2 to 2.5 °C and altitudes from 1.5 to 9.7 km—including in weakly electrified winter storms. These frequent observations challenge prevailing assumptions and underscore the need for comprehensive analysis. Given that the Cloud Particle Imager (CPI) captured millions of particle images during IMPACTS, manual classification is infeasible. To address this, we developed a supervised XGBoost classifier that leverages particle morphological features (e.g., maximum dimension, curl, compactness, and complexity) along with environmental parameters, as chain aggregates possess a unique structural appearance that differentiates them from other ice particle habits. The model was trained on a labeled dataset of tens of thousands of CPI particle images, manually classified across several IMPACTS flights that included both chain aggregate and non-chain observations within varying environments and storm types. The XGBoost classifier achieves higher accuracy ($\sim 96\%$), reduced overfitting, and more balanced predictions compared to earlier efforts. Applying this classifier across the campaign enables robust mapping of chain aggregate distributions and supports comparisons with collocated ER-2 radar and lidar data. This automated framework enhances our ability to identify, quantify, and contextualize chain aggregates, ultimately improving our understanding of their formation and impacts in winter storm environments.