The Influence of Aerosols on Cirrus Clouds during an Aircraft Campaign

**Introduction & Motivation**

Climate change is disrupting societies and harming the delicate ecosystems on Earth (IPCC, 2023). Unfortunately, global-scale action does not seem to be taking place to curb climate change, thus it is necessary that we appropriately evaluate all the potential approaches to mitigate it (Lawrence et al. 2018). One such approach requires the use of an efficient ice nucleating particle (INP) to thin ice clouds, also referred to as cirrus clouds, which are ubiquitous in the atmosphere and uniquely have a net warming effect on Earth due to trapping outgoing longwave radiation (OLR) (Sassen et al., 2008).

Cirrus clouds are composed entirely of ice crystals and form at high altitudes where temperatures can get colder than 235K. Thus, by 'thinning' them with an INP, such as mineral dust, their optical thickness can decrease allowing OLR to exit the atmosphere and causing a cooling effect on the climate (Mitchell & Finnegan, 2009; Storelvmo et al., 2014). But the success of this approach is reliant on the way that the ice crystals nucleated in the cloud as this controls whether it has a warming or cooling effect on the climate. Ice crystals are formed by either homogeneous or heterogeneous nucleation, the former produces ice crystals that are both numerous and small while the latter creates fewer ice crystals that are larger in size (Storelvmo, 2017). Adding an INP to a heterogeneously formed cirrus cloud would increase its optical thickness resulting in a warming effect as ice crystals will preferentially form upon the extra INPs rather than growing on the already available ice crystals. Therefore, it is imperative that a method is developed in order to understand the spatial patterns for the two cirrus cloud nucleation formation mechanisms.

**Data & Methods**

For this project I am using meteorological, cloud, and aerosol measurements from the Atmospheric Tomography Mission (ATom), an airborne field campaign with wide spatial coverage (Wofsy et al., 2021). The flight tracks for this campaign can be seen in Figure 1 where they are separated by each season. The Cloud Aerosol and Precipitation Spectrometer (CAPS) provides cloud indictor data which is a categorical variable that can be a value from 0 to 4 (0 = cloud free, 1 = aerosol cloud transition regime (ACTR), 2 = liquid cloud, 3 = cloud in the mixed-phase temperature regime, and 4 = cirrus cloud). I plan to use a model to predict this variable with aerosol measurements from the Aerosol Microphysical Properties (AMP) instrument suite which includes the aerosol sizing instruments: 5-chan CPC (N-MASS), ultra-high sensitivity aerosol size spectrometer (UHSAS), and the laser aerosol spectrometer (LAS). These size dry aerosols encountered during the campaign into the fine (0.0027 um - 0.5 um) and coarse (0.5 um - 4.8 um) size ranges based on the ammonium sulfate optical equivalent diameter.

To begin, I created a feature matrix composed of the time, season, latitude, longitude, altitude, temperature (T), vertical velocity (w), saturation vapor pressure with respect to water, saturation vapor pressure with respect to ice, fine dry aerosol number concentration, coarse dry aerosol number concentration, fine dry aerosol surface area concentration, coarse dry aerosol number concentration, and total aerosol number concentration. I cleaned the data by replacing the fill in value of -99999 with nan and then removing all the rows that contain a nan in any one of the features. After cleaning the data, I was left with 1211447 samples as seen in Figure 1. Next, I also removed any samples that were either cloud free or in the ACTR because I was only interested in the in-cloud data (2 to 4), which left me with 18980 samples. Figure 2 shows histograms for each step that cleans or filters the dataset from its original number of samples.

Next, I created a feature matrix that only had aerosol measurements in order to predict the in-cloud indicator data. I decided to use a supervised model because I had labeled data as seen in the probability density function (PDF) for each cloud type binned against temperature (Figure 3). I attempted to use linear SVC, because I was predicting a category and had less than 100K samples, but it did not converge despite increasing the number of maximum iterations to 10K every $10^i$ starting from i = 1. This demonstrated to me that the variables were too colinear for this type of model, so I opted to use a Support Vector Classification (SVC) with the Gaussian Radial basis function (RBF) instead, which did converge.

Now that the relationship between aerosols and clouds was established, I sought to go one step further and use a model that would use data identified as cirrus cloud by the cloud indicator to establish whether it was formed homogenously or heterogeneously. For this I did not have labeled data, so I needed to turn to a clustering algorithm, an unsupervised method, to find patterns across my feature matrix. From my prior knowledge, I know that cirrus cloud formation ice nucleation mechanisms are highly dependent on the temperature, relative humidity, vertical velocity, aerosol number concentration and aerosol surface area concentration. For homogeneous ice nucleation to take place for cirrus clouds, the temperature needs to be colder than about 235K and the relative humidity with respect to ice greater than approximately 140%. This is a high energy threshold to be crossed, but as previously mentioned, in the presence of an INP this threshold can be lowered as heterogenous ice nucleation can take place at warmer temperatures and a lower relative humidity with respect to ice. This can be observed in the PDF for cirrus clouds in Figure 3 which is not limited to temperatures colder than 235K. The vertical velocity is also an important feature as it can uplift water vapor or even ice crystals to colder regions in the upper atmosphere. Specifically, I know that I only have 2 categories (homogeneous and heterogeneous) and that I have less than 10K samples. Thus, I decided to use the K-means model to predict these categories based on these nine features. Before I did the K-means I scaled, or standardized, the feature matrix in order to diminish the sensitivity for different scales across the various features during the clustering process. Then I rescaled the data along with the cluster centroids to plot each feature used in the K-means against temperature for reference (Figure 4).

Finally, I wanted to predict the locations of these ice nucleation mechanisms, so I attempted to use the labels identified by the K-means clustering algorithm to train a multiple output regression model that predicts latitude and longitude. To identify which label represents which ice nucleation mechanism, I plotted Figure 5 which has two PDFs based on the labels from the k-means clustering against temperature. I decided that the PDF at warmer temperatures was most likely heterogeneous nucleation while the one at colder temperatures was homogeneous nucleation. Using this information, I set up two feature matrices and target variables for samples identified as homogeneous nucleation and heterogeneous nucleation, respectively. For the feature matrices, I kept the time, month, year, temperature vertical velocity (w), saturation vapor pressure with respect to water, saturation vapor pressure with respect to ice, fine dry aerosol number concentration, coarse dry aerosol number concentration, fine dry aerosol surface area concentration, coarse dry aerosol number concentration, total aerosol number concentration and the season. The season was encoded from a string into a number that corresponds to each season. Then I used a multiple output regression model with the Random Forest Regressor to predict the latitude and longitude of these two ice nucleation mechanisms.

**Results & Limitations**

For my first model I used the SVC with a RBF kernel in order to predict cloud types based on a feature matrix with only aerosol measurements. I obtained an accuracy of 74% as seen in Table 1 which shows the classification report for this model. This report provides valuable insights into the model's performance across different classes. Precision represents the accuracy of positive predictions among the instances classified as positive, emphasizing the reliability of the model's positive classifications. Recall, on the other hand, measures the ability of the model to identify all relevant instances of a particular class, emphasizing sensitivity. F1-score combines precision and recall, providing a balanced metric for model evaluation. Looking at the specific class results, the model exhibits high precision (85%) for class 2.0, indicating a low rate of false positives. However, class 3.0 shows lower precision (71%), suggesting a higher rate of misclassifying instances as positive. Class 4.0 demonstrates a trade-off between precision and recall, resulting in a relatively balanced F1-score (83%). The overall accuracy of 77% indicates the proportion of correctly classified instances across all classes. The macro-average and weighted-average metrics provide summarized insights across all classes, considering either equal weight for each class or accounting for class imbalances, respectively. Overall, the values for the classification report in Table 1 suggest a reasonably good performance, especially considering the balanced F1-scores and the weighted average metrics, thus, demonstrating that aerosols can adequately predict the different types of clouds.

The Root Mean Squared Error (RMSE) values for the multiple output regression model were 84.79 and 45.28 for homogeneous and heterogeneous nucleation, respectively. The RMSE is a measure of the average magnitude of the residuals, representing the differences between the predicted and actual values. Lower RMSE values are desirable, as they signify smaller prediction errors and, consequently, higher precision in the model's output. Unfortunately, the RMSE values for this model are high indicating that there are areas for improvement. It is interesting to note that the RMSE is lower for heterogeneous nucleation, which actually has more samples (5344) compared to homogeneous nucleation (3434). Thus, this could be an issue of simply not having enough samples.

There are some limitations to this work. Firstly, cleaning and filtering the original dataset to just in-cloud and cirrus cloud data decreased the number of samples from 1490913 to 18980 and 9154, respectively. I believe it would have been better to have some more samples, especially for the cirrus clouds, in order to train the model. Additionally, I encountered some issues with the relative humidity with respect to ice measurements and was not able to use it in the feature matrix. This variable is preferable to saturation vapor pressure with respect to ice in ease of interpretation, but unfortunately it decreased the number of samples for in-cloud data for summer to none. Finally, the aerosol measurements I used in this project are not from a full-size distribution which paints an inaccurate picture of aerosols in the atmosphere as mineral dust is a coarse aerosol. This is because the coarse size range upper limit is only 4.8 um and doesn't capture larger sizes of dust which could be better at predicting heterogeneous nucleation. This is because surface area directly controls ice nucleation as larger INPs can have a higher occurrence of ice activation sites as there is simply more space on its surface.

**Future Work & Conclusion**

I am planning on refining and extending this work in the future. I am interested in adding other sources and types of data to get more spatial and temporal coverage. I also would like to further down this work and specifically look at mineral dust as it is an efficient INP that can have implications for climate geoengineering (Cziczo et al., 2013; Lohmann and Gasparini, 2017). As seen in Figure 1, the ATom campaign does not include many samples over Africa, Europe, and Asia, which are important source regions of aerosols and certain deserts are better at supplying mineral dust into the upper troposphere than others (Froyd et al., 2022). Unfortunately, there is a lack of mineral dust concentration measurements at cirrus cloud forming altitudes, but it is possible to use ice crystal number concentrations (ICNC) to estimate this value (Burrows et al. 2022). Although, field campaigns that have captured hydrometeor imaging from the 2D-S Stereo Probe (2DS), which is an optical imaging instrument that obtains stereo cloud particle images and concentrations using linear array shadowing, could provide a promising avenue to obtain this information. For example, a group used hydrometeor images captured by the 2DS instrument during the Southern Ocean Clouds, Radiation, Aerosol Transport Experimental Study (SOCRATES) campaign to discriminate between liquid and ice in the mixed-phase cloud regime by using a random forest algorithm (Atlas et al., 2021). I would be interested in using a similar method to discriminate between ice crystals that were formed heterogeneously with the assistance of an INP and those that were formed by homogeneous freezing. However, it is important to note though that this may cause some overestimation issues in the derived mineral dust concentration due to secondary ice production processes being difficult to account for. If possible, I would like to figure out how I can account for this type of ice production when attempting to estimate the mineral dust concentration using ICNC, but I would have to look into this issue further within the scientific literature.

Ultimately, this work could eventually inform cirrus cloud geoengineering by highlighting the regions on Earth where introducing mineral dust would produce a cooling effect. Cirrus cloud seeding specifically targets the OLR which makes it potentially much more effective at mitigating climate change (Lawrence et al., 2018). It is also quite appealing to conduct further research into because dust is relatively chemically inert. Geoengineering as a whole has been largely ignored in practice due to its risky nature, but it is necessary to conduct this research now least we require it in the future and have not yet attempted to solve this longstanding question.

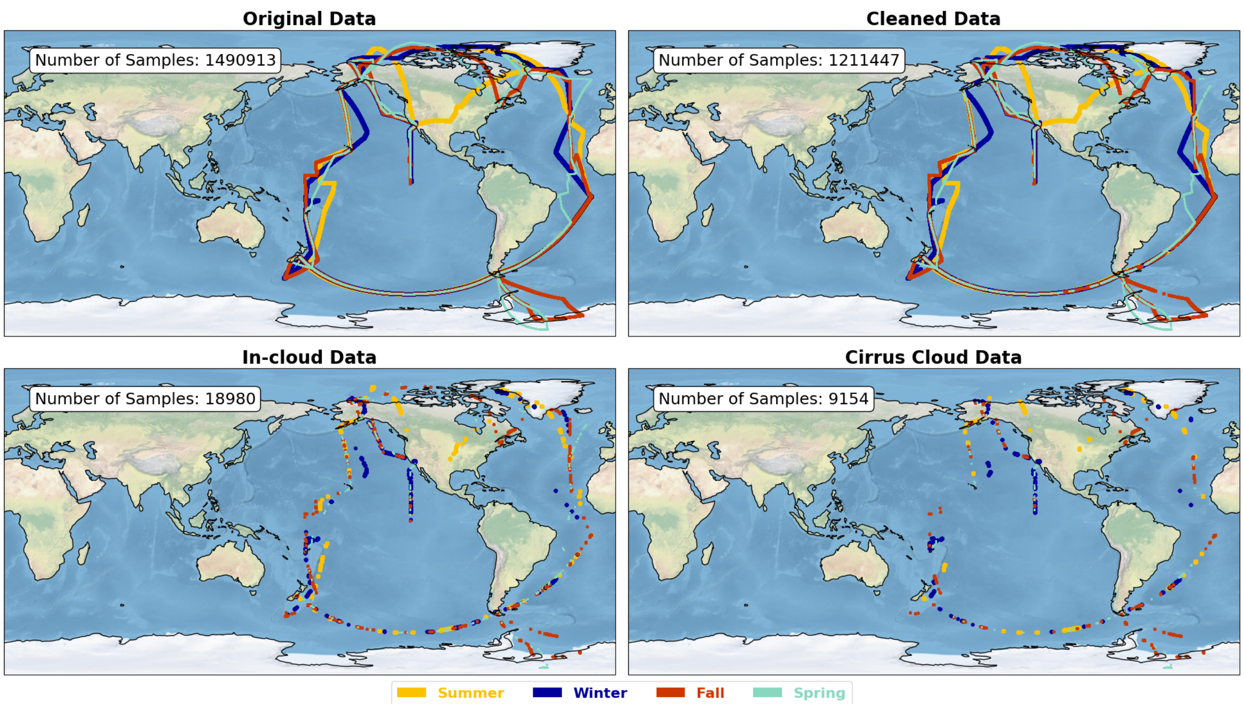


**Figure 1 |** The flight tracks for the ATom campaign for each season separated by original data, cleaned data, in-cloud data, and cirrus cloud data.

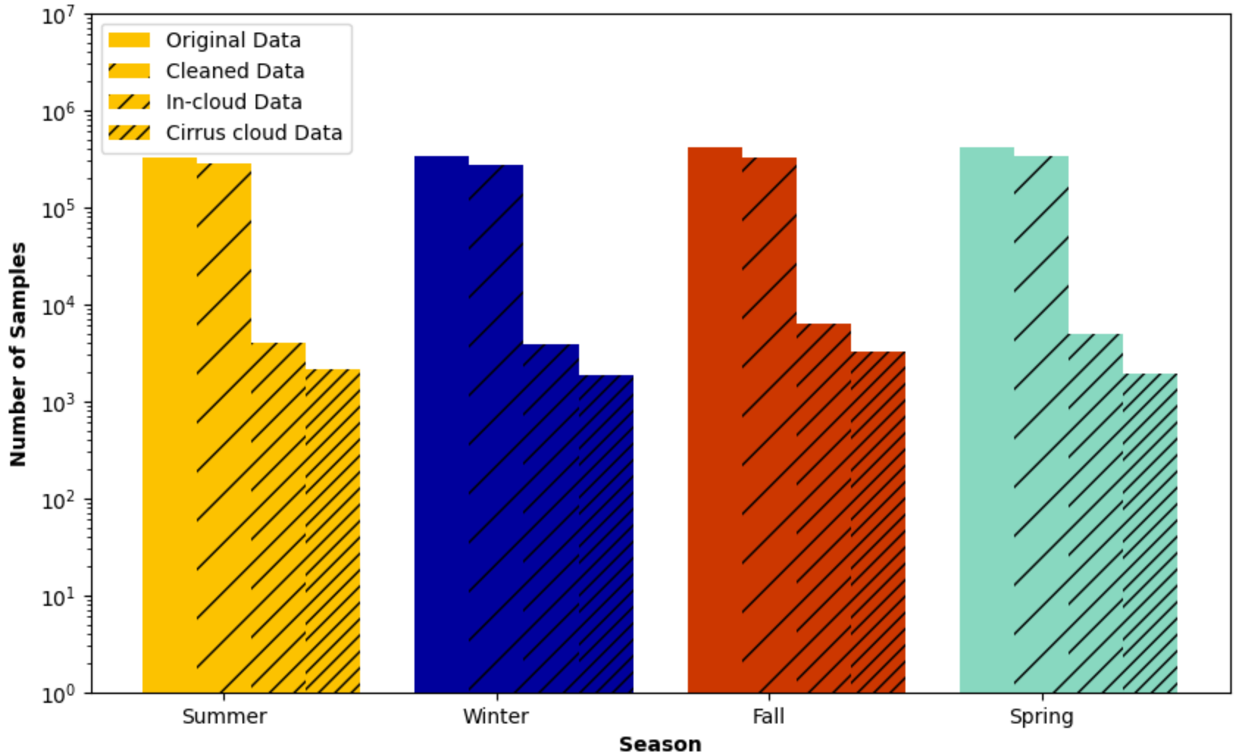


**Figure 2 |** Histograms that show the number of samples for each season in the ATom campaign further separated into the original data, the cleaned data, in-cloud data, and cirrus cloud data.

***Table 1 |*** *A classification report for the SVC with RBF analysis.*

| | Precision | Recall | F1-score | Support |
|---|---|---|---|---|
| 2.0 | 85% | 78% | 81% | 1046 |
| 3.0 | 71% | 39% | 50% | 912 |
| 4.0 | 74% | 95% | 83% | 1838 |
| Accuracy | | | 77% | 3796 |
| Macro Average | 77% | 70% | 72% | 3796 |
| Weighted Average | 77% | 77% | 75% | 3796 |

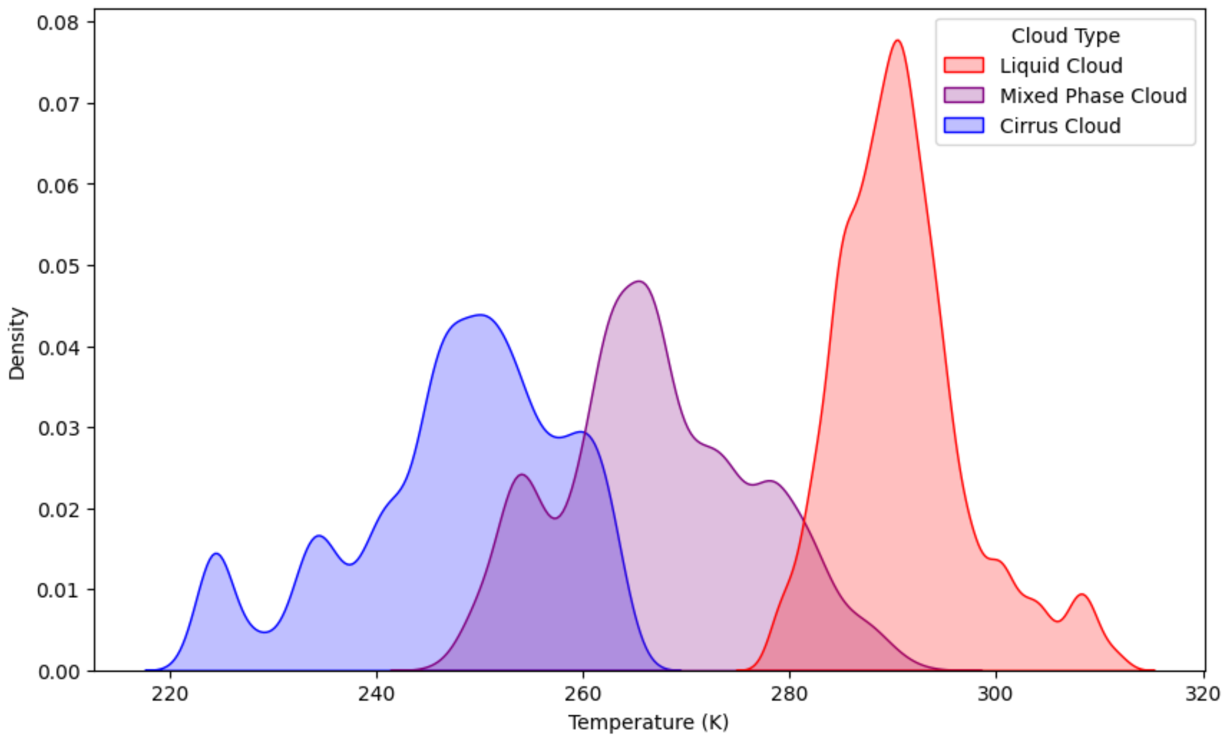


**Figure 3 |** Probability distribution functions for the three types of cloud in the ATom dataset

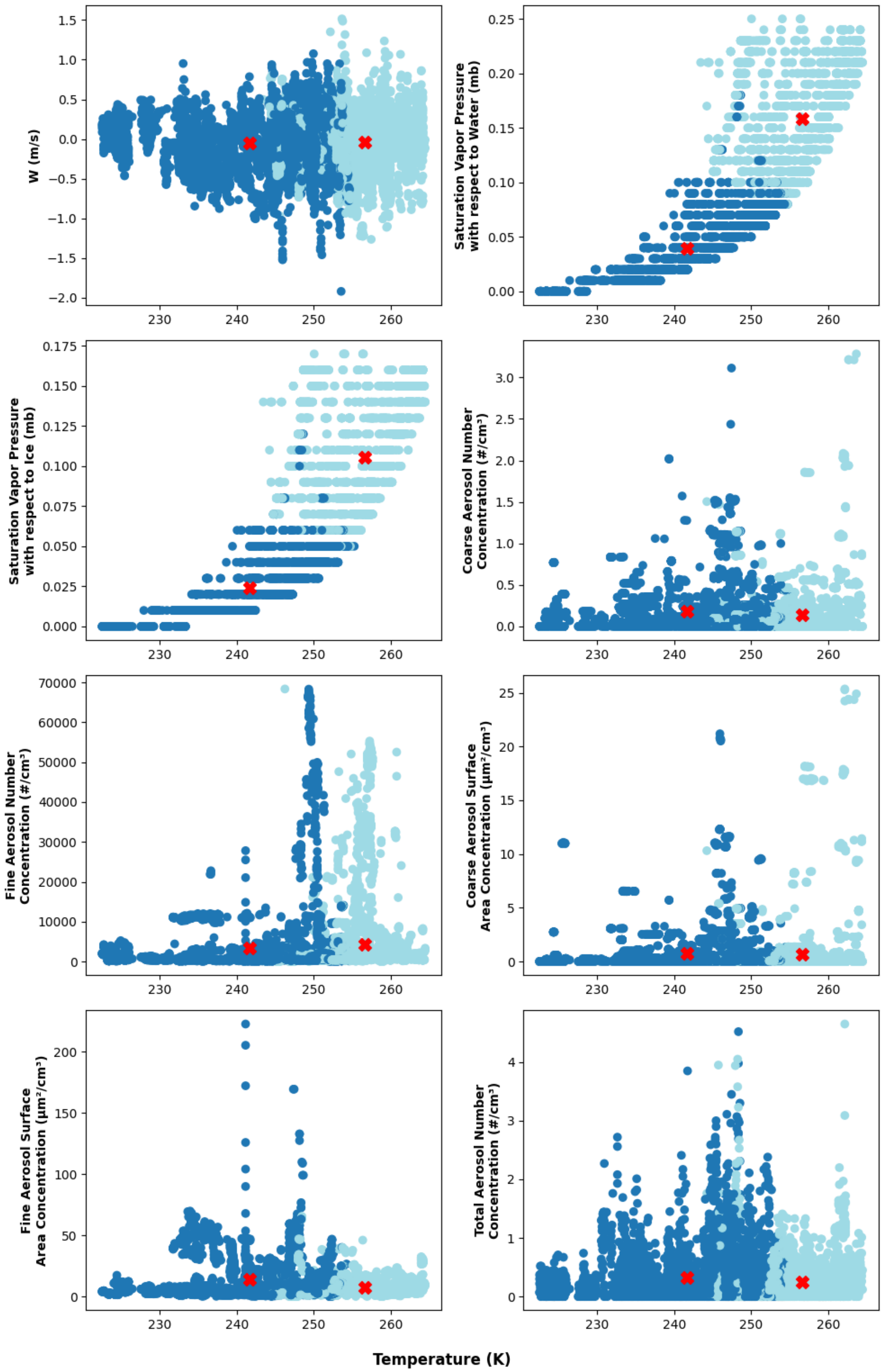


**Figure 4 |** Scatter plots for the features used for the K-means clustering plotted against temperature. Red Xs signify the cluster's centroid.

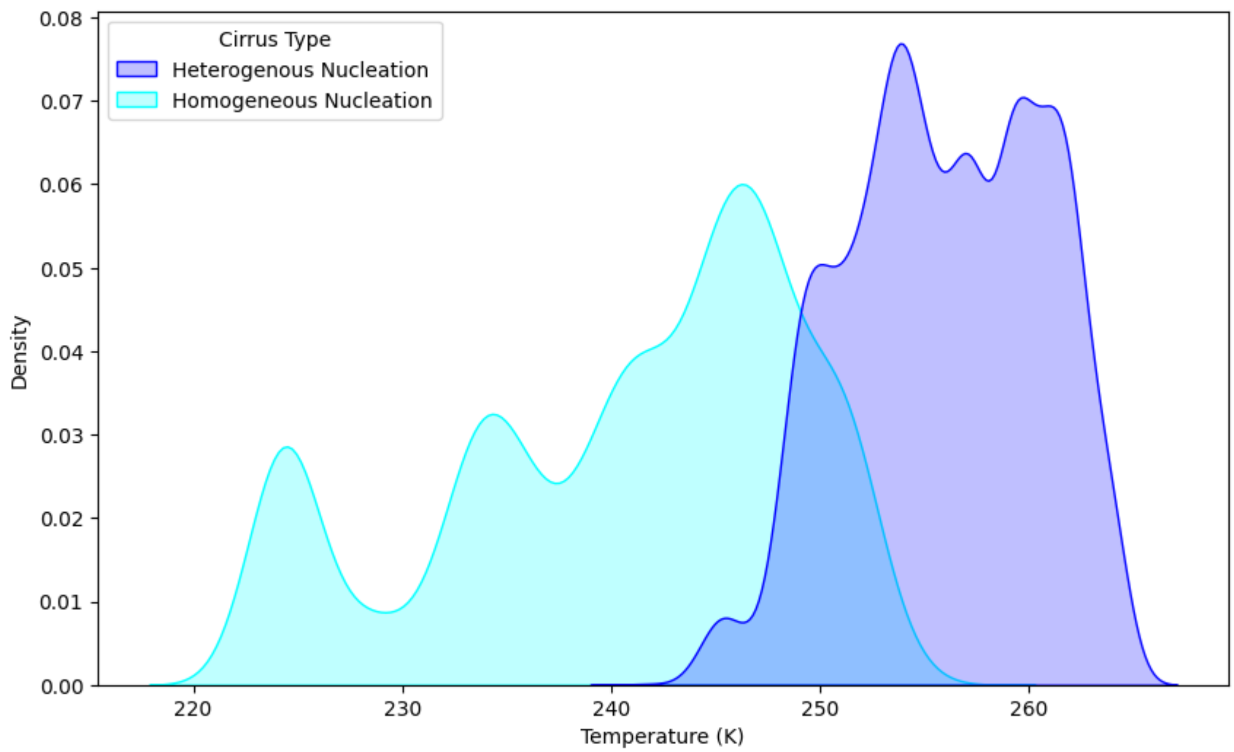


**Figure 5 |** Probability distribution functions for the two ice nucleation formation mechanisms for cirrus clouds as identified using k-means clustering.

## References


Atlas, R., Mohrmann, J., Finlon, J., Lu, J., Hsiao, I., Wood, R., & Diao, M. (2021). The University of Washington Ice–Liquid Discriminator (UWILD) improves single-particle phase classifications of hydrometeors within Southern Ocean clouds using machine learning. Atmospheric Measurement Techniques, 14(11), 7079–7101. https://doi.org/10.5194/amt-14-7079-2021

Burrows, S. M., McCluskey, C. S., Cornwell, G., Steinke, I., Zhang, K., Zhao, B., et al. (2022). Ice-Nucleating Particles That Impact Clouds and Climate: Observational and Modeling Research Needs. Reviews of Geophysics, 60(2), e2021RG000745. https://doi.org/10.1029/2021RG000745

Cziczo, D. J., Froyd, K. D., Hoose, C., Jensen, E. J., Diao, M., Zondlo, M. A., et al. (2013). Clarifying the Dominant Sources and Mechanisms of Cirrus Cloud Formation. Science, 340(6138), 1320–1324. https://doi.org/10.1126/science.1234145

Froyd, K. D., Yu, P., Schill, G. P., Brock, C. A., Kupc, A., Williamson, C. J., et al. (2022). Dominant role of mineral dust in cirrus cloud formation revealed by global-scale measurements. Nature Geoscience, 15(3), 177–183. https://doi.org/10.1038/s41561-022-00901-wIntergovernmental Panel on Climate Change (Ed.). (2014). Clouds and Aerosols. In Climate Change 2013 – The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (pp. 571–658). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.016

Huang, Y., Adebiyi, A. A., Formenti, P., & Kok, J. F. (2021). Linking the different diameter types of aspherical desert dust indicates that models underestimate coarse dust emission. Geophysical Research Letters, 48, e2020GL092054. https://doi.org/10.1029/2020GL092054

Kok, J. F., Storelvmo, T., Karydis, V. A., Adebiyi, A. A., Mahowald, N. M., Evan, A. T., et al. (2023). Mineral dust aerosol impacts on global climate and climate change. Nature Reviews Earth & Environment, 4(2), 71–86. https://doi.org/10.1038/s43017-022-00379-5

Lawrence, M. G., Schäfer, S., Muri, H., Scott, V., Oschlies, A., Vaughan, N. E., et al. (2018). Evaluating climate geoengineering proposals in the context of the Paris Agreement temperature goals. Nature Communications, 9(1), 3734. https://doi.org/10.1038/s41467-018-05938-3

Lohmann, U., & Gasparini, B. (2017). A cirrus cloud climate dial? Science, 357(6348), 248–249. https://doi.org/10.1126/science.aan3325 Mitchell, D. L., & Finnegan, W. (2009). Modification of cirrus clouds to reduce global warming. Environmental Research Letters, 4(4), 045102. https://doi.org/10.1088/1748-9326/4/4/045102

Sassen, K., Wang, Z., & Liu, D. (2008). Global distribution of cirrus clouds from CloudSat/Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) measurements. Journal of Geophysical Research: Atmospheres, 113(D8). https://doi.org/10.1029/2008JD009972

Storelvmo, T., Boos, W. R., & Herger, N. (2014). Cirrus cloud seeding: a climate engineering mechanism with reduced side effects? Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 372(2031), 20140116. https://doi.org/10.1098/rsta.2014.0116

Storelvmo, T. (2017). Aerosol Effects on Climate via Mixed-Phase and Ice Clouds. Annual Review of Earth and Planetary Sciences, 45(1), 199–222. https://doi.org/10.1146/annurev-earth-060115-012240

Wofsy, S.C., S. Afshar, H.M. Allen, E.C. Apel, E.C. Asher, B. Barletta, J. Bent, H. Bian, B.C. Biggs, D.R. Blake, N. Blake, I. Bourgeois, C.A. Brock, W.H. Brune, J.W. Budney, T.P. Bui, A. Butler, P. Campuzano-Jost, C.S. Chang, M. Chin, R. Commane, G. Correa, J.D. Crounse, P. D. Cullis, B.C. Daube, D.A. Day, J.M. Dean-Day, J.E. Dibb, J.P. DiGangi, G.S. Diskin, M. Dollner, J.W. Elkins, F. Erdesz, A.M. Fiore, C.M. Flynn, K.D. Froyd, D.W. Gesler, S.R. Hall, T.F. Hanisco, R.A. Hannun, A.J. Hills, E.J. Hintsa, A. Hoffman, R.S. Hornbrook, L.G. Huey, S. Hughes, J.L. Jimenez, B.J. Johnson, J.M. Katich, R.F. Keeling, M.J. Kim, A. Kupc, L.R. Lait, K. McKain, R.J. Mclaughlin, S. Meinardi, D.O. Miller, S.A. Montzka, F.L. Moore, E.J. Morgan, D.M. Murphy, L.T. Murray, B.A. Nault, J.A. Neuman, P.A. Newman, J.M. Nicely, X. Pan, W. Paplawsky, J. Peischl, M.J. Prather, D.J. Price, E.A. Ray, J.M. Reeves, M. Richardson, A.W. Rollins, K.H. Rosenlof, T.B. Ryerson, E. Scheuer, G.P. Schill, J.C. Schroder, J.P. Schwarz, J.M. St.Clair, S.D. Steenrod, B.B. Stephens, S.A. Strode, C. Sweeney, D. Tanner, A.P. Teng, A.B. Thames, C.R. Thompson, K. Ullmann, P.R. Veres, N.L. Wagner, A. Watt, R. Weber, B.B. Weinzierl, P.O. Wennberg, C.J. Williamson, J.C. Wilson, G.M. Wolfe, C.T. Woods, L.H. Zeng, and N. Vieznor. (2021). ATom: Merged Atmospheric Chemistry, Trace Gases, and Aerosols, Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1925