Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Master of Science

TITLE Estimating cloud droplet concentration using machine learning methods.

PRESENTED BY Aditya Biyani

ACCEPTED BY THE DEPARTMENT OF

Chemical Engineering

HAMISH GORDON, PROFESSOR AND ADVISOR 12/10/2020 DATE

ANNE S. ROBINSON, PROFESSOR AND DEPARTMENT HEAD 12/11/2020 DATE

Estimating cloud droplet concentration using machine learning methods

FALL 2020

Aditya Biyani

abiyani@andrew.cmu.edu

Graduate Student

Department of Chemical Engineering

Carnegie Mellon University

5000 Forbes Avenue, Pittsburgh, PA 15213, USA

Hamish Gordon

gordon@cmu.edu

Assistant Research Professor

Engineering Research Accelerator, Chemical and Mechanical Engineering

Carnegie Mellon University

5000 Forbes Avenue, Pittsburgh, PA 15213, USA

December 16, 2020

Abstract

Cloud microphysical data obtained from aircraft measurements during the GoAmazon2014/5 campaign were analyzed to understand the relationship of cloud droplet concentration (cdp_conc) with several other variables like cloud condensation nuclei (ccn_conc), liquid water content (cdp_mass), altitude (h), relative humidity (RH) and mean updraft velocity and it's standard deviation (w_mean and w_std). The clouds studied during the campaign are mostly stratocumulus type and analyses show that these clouds have inhomogeneous mixing (IM). Thus, with the limitations of not accurately predicting the cloud parameters using radar observations it becomes important to also consider aircraft observations. The variables that affect the cdp_conc as mentioned above do not have a linear relationship with cdp_conc. This was the motivation behind this study which would help corroborate the non-linear relationship with various machine learning regression algorithms. The neural networks applied to the data yielded that with an accuracy of 63%, it was possible to estimate the cloud droplet concentration with the above features.

Table of Contents

Introduction

Clouds are accumulated mass of water in the atmosphere occurring in several microphysical states. Cloud drops are the smallest particles and the basic building blocks of clouds. These particles can exist in liquid state even in freezing conditions of the atmosphere because of the high surface tension of water. Pollutants and other aerosol particles derived due to anthropogenic activities or natural events rise up in the atmosphere (mainly troposphere). When the cloud droplets come in contact with the aerosol particles their surface tension breaks and they condense. These aerosol particles are now called the cloud condensation nuclei (CCN).

Cloud height can range from a few meters to hundreds of meters. Depending on the height of the cloud it can be either classified as shallow or deep cloud. The number concentration of CCN is responsible for the growth of droplets into other hydrometeor species. By the process of coalescence, the droplets agglomerate and combine to form bigger droplets as a result the cloud's mass increases. Now, depending on the height of the cloud, the agglomerated droplets develops into other hydrometeor species like rain, ice, snow or graupel.

Number concentrations of cloud droplets range from 20 to 1000 cm^{-3}[1]; number concentration of ice crystals range from less than 10 to more than 100,000m^{-3}[2] and so on. Because number concentration influences particle size and hence gravitational settling, collision/ coalescence and cloud radiative properties such wide variations in cloud particle number concentrations can be expected to produce large differences in the precipitation efficiency and optical properties of clouds.

As the altitude helps in determining the amount of every hydrometeor species in the cloud, the updraft wind is responsible for interaction between these species. Due to high vertical wind speeds the particles collide with each other and as a result either precipitate or get reduced in mass. The terminal falling velocity of the precipitating hydrometeor species is also governed by the mass of the droplet and the frictional forces. These forces can be responsible for change in the mass and concentration of the species.

Thus, as a part of this study the variables that are most like to affect the droplet concentration are the droplet mass, CCN concentration, updraft wind speed and relative humidity. One way to measure these features are using the radars. Using the principles of reflectivity, radars send IR impulses of varying intensities/wavelengths. When these signals hit a cloud particle in the atmosphere, they return back the reflected signal and get recorded. Some examples are the millimeter radars and Doppler radars. The radars are quite useful in weather predictions but certain drawbacks of them are that they depend on the optical properties of the cloud hydrometeor species. Uncertainty in the weather conditions like high winds leading to change in location of the species can lead to incorrect results. More accurate results in a given timeframe can be acquired using aircraft observations. Some of the benefits are that aircrafts can sample in all 3 dimensions of latitude, longitude and altitude. The quality of observations obtained from each reporting aircraft is routinely monitored by meteorological centers and the information is fed back to airlines. These benefit directly by using it to help maintain the high standard of aircraft performance [3]. Cloud base and cloud top can be effectively estimated using these observations. Thus, this research uses Aircraft observations for cloud droplet estimation.

Manaus, an industrial city and with a metropolitan population of more than 2 million people, is the largest city in the Amazon basin. The prevailing wind is from the east with the nearest major upwind city, Belem, approximately 1250 km away. The surrounding tropical forest emits 20 vast quantities of biogenic gases and aerosol. Few roads connect Manaus to the rest of Brazil and most freight and traffic from outside the city is via ship or plane. Thus, Manaus acts as a large point source of anthropogenic emissions which are transported to the surrounding and nearly pristine Amazon basin. As part of this campaign, the DOE Gulfstream-1 research aircraft conducted two, six-week-long missions in which it investigated the evolution of the Manaus plume as it was transported into the surrounding Amazon tropical rainforest. Here we report on measurements from instruments deployed on the G-1 [4].

The findings in this report are also inspired by the research study on impact of secondary droplet activation on cloud microphysical relationships during the wet and dry seasons in the Amazon [5]. The research paper compares different microphysical cloud properties in shallow cumulus clouds during the wet and dry season. The analyses was done from the data obtained during the GoAmazon2014/5 campaign aircraft measurements.

In warm clouds, usually during the wet season the cloud droplets increase in size and set the stage for precipitation. From the process of condensation they grow in size even as the altitude increases. Theory suggests otherwise that as particle sizes grow the condensation rate decreases with time. This gap between observation and theory is addressed in the paper.

Two ideal types of clouds are mainly studied: One where the cloud air is well mixed with the entrained air also called as homogeneous mixing (HM) whilst the other one is inhomogeneous mixing (IM). Laboratory experiments explain IM as the phenomenon where droplets of cloudy air adjacent to entrained air evaporate completely while rest of the cloudy air remains intact [6].

This entrained air contains aerosol particles which also affect the cloud droplet effective radius and optical thickness. Slawinska et al., 2012 [7] found that 40% of clouds were activated above cloud base, thus, maintaining the mean cdp_conc consistent with altitude. Activation was related to the aerosol particle size and degree of turbulence. This activation is also possible because of dilution by entrainment, detrainment and coalescence losses.

LES Model used by Heus et al. (2008) [8], shows that entrained air originated from observation altitude because lateral mixing is dominant in the shallow cumulus clouds. It can be understood from the frequency distributions in Figure 1; that the lower slope between cdp_conc (N) and cdp_mass (L) in dry season have more data with cdp_conc > 600 cm^{-3}. This is also justified by the cdp_mass (L) and droplet volume (V) in dry season. The slope becomes less steep which means the droplets are smaller to their counterparts in wet season.

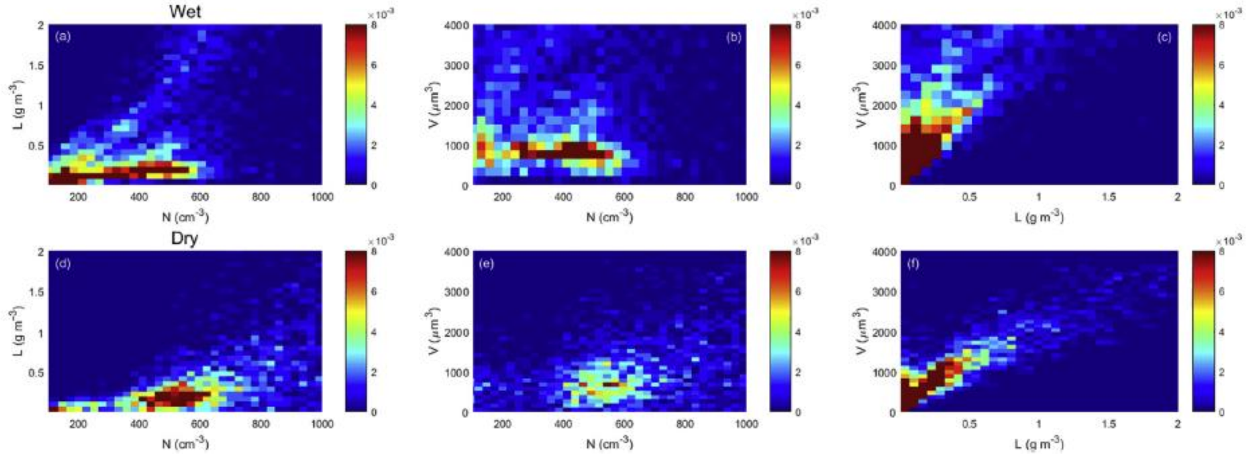


Figure 1: Frequency distributions that depict relationships between cloud microphysical variables (N, L, and V) for the wet (a, b, and c) and dry (d, e, and f) seasons. [N: cdp_conc, L: cdp_mass and V: droplet volume]. [Source: Yeom et al, 2020].

Thus, from the above comparison we can conclude that higher ccn_conc leads to reduced cdp_mass per droplet and in turn, reduced effective supersaturation. The clouds droplets competes with the entrained aerosol particles for water in the atmosphere making it difficult for them to grow in size and convert in some form of precipitation, thereby reducing droplet concentration and prevents additional aerosol activation.

Yeom et al. [5] have also explained that all droplets activation above the cloud base keeps the mean cdp_conc constant with altitude which serves as the basis for this research as well (see Figure 2). With this as the basis, this study also relies on no change of cloud droplet concentration with altitude. In theory, the droplet size increases with altitude. This change reverses when it reaches the planetary boundary layer (PBL). The increase in cloud mass due to increase in cloud diameter remains consistent with the number concentration of cloud droplets with altitude. This has also been studies using the Large Eddy Simulations. The droplet concentration plot of one flight day as shown in Figure 2, explains that with increase in altitude the concentration number is within the range of 200 – 800 #/cc with a mean of 400 #/cc. Some observations in the plot also show that as the altitude is increasing, the number is decreasing which can be because of several reasons like precipitation, effects of high wind turbulence or an effect of aircraft sampling. This also results in activation of new cloud droplets and become more pronounced as the effective turbulence rate (ε) increases.

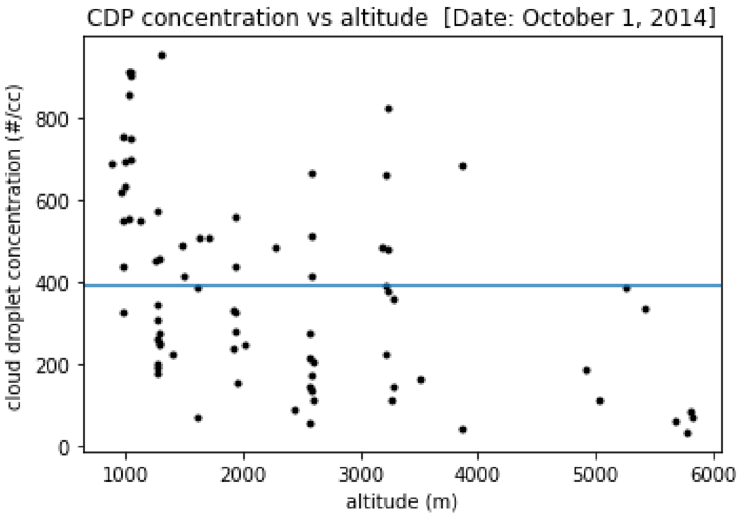


Figure 2: change effect of droplet concentration with altitude plot for one day.

There is a lot of fluctuation in the w_mean (W) as understood from Figure 3. Factors contributing to this fluctuation are IM and pressure difference. It is a cyclic process where the enhanced entrained air increases the evaporation. Increase in evaporation increases the turbulences, large eddies start to form and in turn accelerate the rise in concentration of entrained aerosol particles in the clouds. The process is very prominent in the dry season due to favorable conditions like large sized entrained aerosol particles, high fluctuation in the updraft speed and ε.

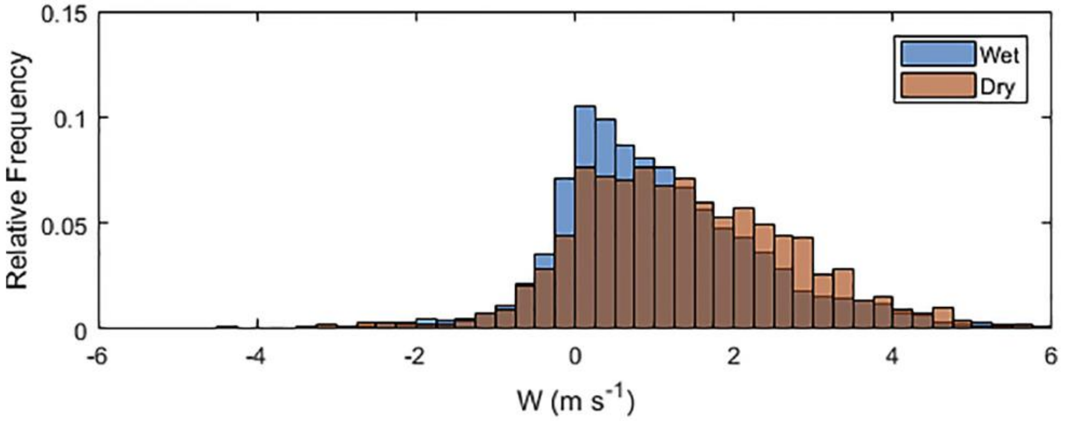


Figure 3: Histogram for W for all cloud segments in the wet and dry season. [Source: Yeom et al, 2020].

Finally, this study corroborates the learning from previous studies that there is a strong correlation between the cdp_conc with entrained aerosol particles (ccn_conc), turbulence (w_mean, w_std), altitude and the cdp_mass (L). This fact indicates that larger droplets preferentially exist in more adiabatic cloud parcels than in diluted cloud regions. This is contrary to IM scenario where super-adiabatic droplets would eventually emerge in the entrainment affected and diluted cloud region.

Data

The data used in this report were obtained from the GoAmazon2014/5 campaign carried out about 6 years ago in the central Amazon close to the city of Manaus, Brazil from January 1, 2014 to December 31, 2015 [9]. The aircraft measurements were made on-board the ARM Aerial Facility[10] Gulfstream-1 (G-1). The period of measurements were from February – March, 2014 and September- October, 2014, characterized as the wet and dry periods. According to Martin et al. [9] the flight tracks were designed to follow the easterly winds that carried the plume from Manuas City.

Cloud microphysics data were measured using the Fast Cloud Droplet Probe (FCDP). It measures cloud droplet size distribution with 20 bins covering a diameter of 1 – 50 µm with a resolution of 3 µm.

Condensational Particle Counter (CPC, TSI model 3010), Fast Integrated Mobility Spectrometer (FIMS) and Ultra High Sensitivity Aerosol Spectrometer (UHSAS) were used to measure number of aerosol concentration of different diameters < 10 nm, 10 – 400 nm and > 400 nm respectively.

Vertical velocity (w_mean) and true air speed were observed by an AIMMS – 20 instrument (Aircraft Integrated Meteorological Measurement System).

In order to fit a machine learning algorithm to calculate the cdp_conc, a data rich value table was needed. This meant to perform calculations on days with significant cloud cover. NASA's worldview Earth Data provided satellite images of the central Amazon. This was used to determine the cloud cover area. After basic analysis in python dataframe, 17 files were selected 12 from the wet season and 5 from the dry season (ARM Research Facility, 2020).

Data Analysis Methodology

The data analysis can be divided into 5 segments: Geospatial visualization, data cleaning, data collation, analytical comparison and machine learning model development.

Geospatial Aircraft path visualization

The satellite images provide the top view of Manaus city and specifically the clouds above that region. This helps in estimating the cloud top height. A snapshot of NASA's MODIS Satellite image of a region Manaus city over which aircraft flew was imported in using cartopy library. The two main platforms of MODIS: Terra and Aqua were used to monitor the cloud activity each with an image resolution of 2km and sensor resolution of 5km. Terra satellite image was used if the aircraft flew before 1700 UTC and Aqua image was used if it flew after. The change of satellites was to ensure that the cloud cover that was to be monitored is updated with time. (See Figure 4) A heat map drawn from the GIBS archive files of NASA earth data was also used to compare if the aircraft actually flew through the cloud or above or below it [11]. The standard color bar of heat map was used for comparison and can be referenced from Worldview: Explore Your Dynamic

Planet website [12]. As seen in Figure 4, the aircraft on an average flies in the range 3000 – 4000m. March 17, 2014 seems to have puffy clouds during the duration of the flight. As compared from the heat map, the purple color on the top left corner of the plot is for the regions above 5000m which suggests that the aircraft was flying under them in that region.

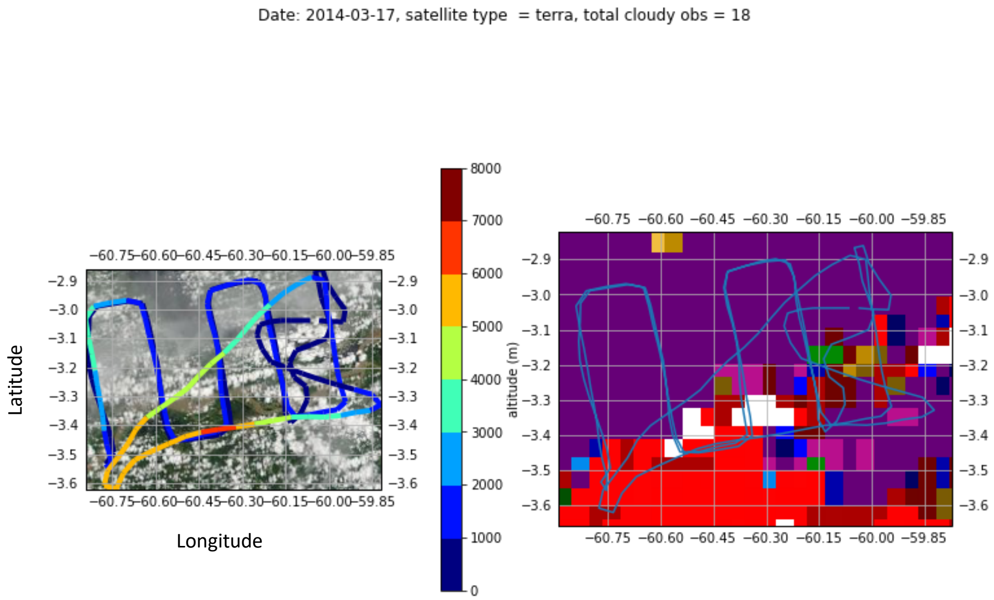
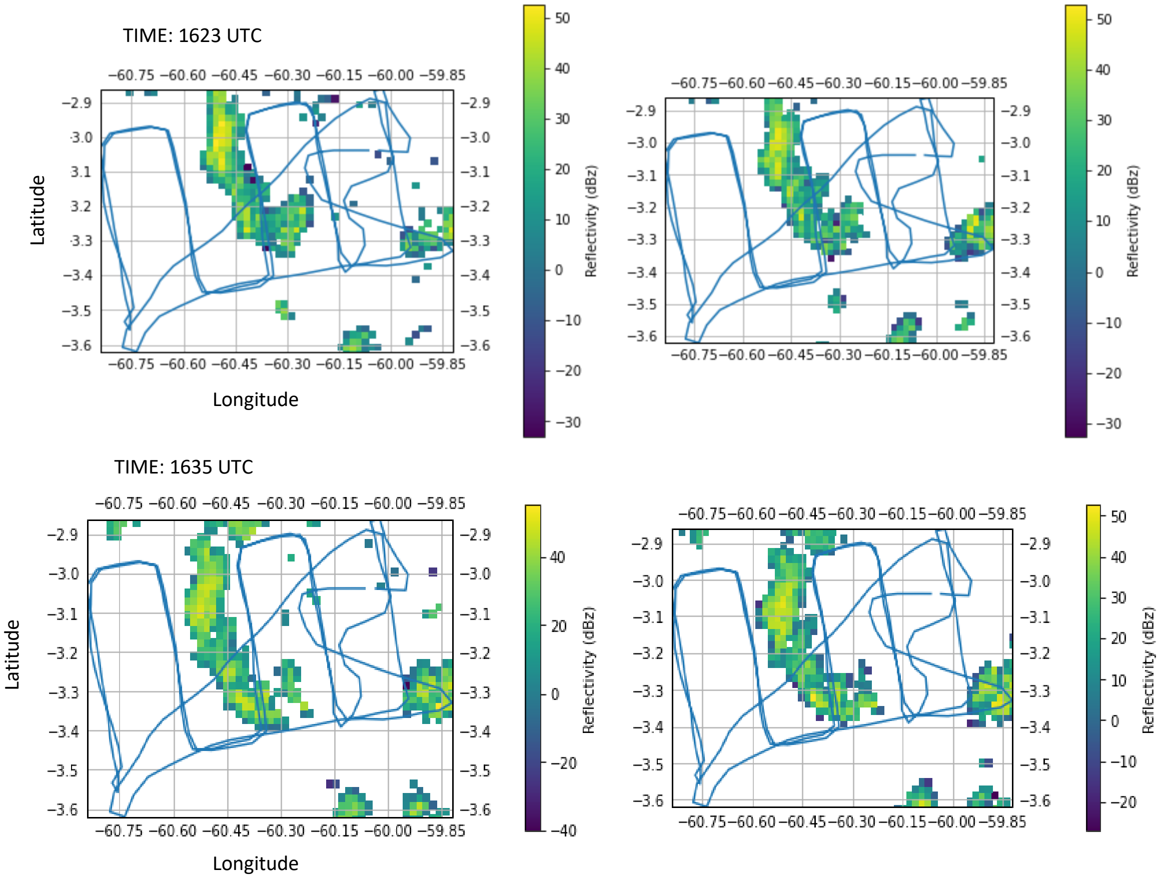

Date: 2014-03-17, satellite type = terra, total cloudy obs = 18

Figure 4: Left plot - snapshot of Manaus region with the aircraft path over it as a function of altitude. Right Plot – Heat map of the same region as given by the MODIS heat map.

Although, the satellite images presented a 2-D representation of the cloud top height, the estimation of the base was still required to understand if the clouds were shallow or deep which would then require the aircraft to alter its path accordingly thus, giving one of the explanations for any noise or erratic observations while recording. Thus, Radar reflectivity data was used to combine the image analysis from the atmosphere with the cloud images as seen from the earth's surface. Figure 5 explains that a range of radar reflectivity plots of the clouds and confirms if the flight was in the cloud cover or not. Three different time instances as an hour interval each from the nearest time instance of flight take off were analyzed. This analysis would confirm if there was any change in the location and volume of cloud cover from the time we start recording observations. Radar images for two different height instances for every interval was plotted. The minimum and maximum height that the aircraft reached during its course. Providing a sense of the cloud base and cloud top as seen from the radar.

Left plot altitude = 1.5 km, Date: 2014-03-17 Right plot altitude = 6.5km



TIME: 1647 UTC

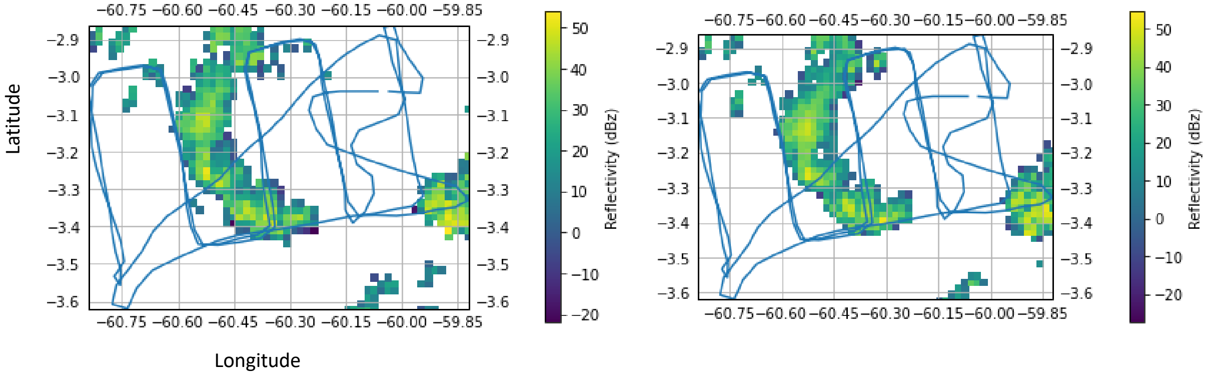


Figure 5: Radar Reflectivity plots for March 17, 2020

Data Cleansing

The files used for aerosol particle and cloud droplet concentration extraction were of '.ict' format. Updraft speed and other parameters like dew point, ambient temperature and altitude files were of the format '.ppc' and '.txt'. All these files could be easily read using the read_csv () function of pandas dataframe. 4 different data frames were created for cdp_conc, cdp_mass, and w_mean and RH calculations. Different functions were created to input each file and return the variables cdp_mass, cdp_conc, w_mean, w_std, RH and rain_conc.

As per the 30 different bins the cloud droplet number concentration was measured as number per unit volume per unit droplet-size (#/L/ µm). Thus, to calculate the aggregate cloud-droplet mass each value was multiplied by its corresponding diameter, volume and water vapor density (assumed to be 1 g/cc). Assuming spherical shape of the cloud droplets, the volume is calculated as $V = \frac{\Pi {D_p}^3}{6}$ (where, D_p = droplet particle diameter).

The cdp_mass calculated in the above function was used in the next function where the cdp_conc was determined from the given observations. Since, it is difficult to estimate the aerosol concentration inside the cloud we estimate the concentration outside and around that cloud. Most of the clouds as that were formed due to IM in the Manaus region were assumed to be strato-cumulus type clouds. The average liquid water content (LWC) of these clouds is 0.2 – 0.25 g/m^3 [13]. The observations of LWC can include all cloud types from stratus, fog, cirrus and cumulus and thus, average threshold mass for the smallest cloud droplet particle is taken as 0.1 g/m^3 and all the cdp_mass values > 0.01 g/m^3 are masked to give ccn_conc and < 0.1 g/m^3 gives cdp_conc inside the cloud to include most cloud types.

The updraft speed calculations were pretty straightforward with the same assumption that wind velocity inside the cloud with cdp_mass < 1 g/m^3 is required. According to earlier studies there is a lot of eddies in the cloud which account for huge wind velocity fluctuations and thus, it was necessary to find the standard deviation of the same. The formula used is:

$$\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N} {x_i}^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2}$$

An important point to note is that all the observations were recorded per second. The data had to be grouped by minutes and taken mean of all the seconds' observations in that minute. The updraft velocity equipment has a frequency of 20 Hz meaning observations in every second also have 20 values. We have to first take the mean every second and then every minute.

The rain droplet mass and rain droplet concentration variables were also calculated in a similar fashion. Except, the instruments used to calculate the rain concentration values had a higher recording frequency than the cloud droplet recording instruments. Another challenge was to match the shapes of both the files because the recording start times and end times were different. So, an empty data frame was created with the columns equal to that of rain observations but rows equal to that of cloud observations. The start-time and end-time of rain observations was found and matched with the clouds. If the rain observations started recording before the cloud instruments then the observations before that moment were dropped from the calculations, similarly all the observations taken after the cloud recording were dropped too. If the observations were recorded later than cloud observations then the missing time instances were filled with rows of zeros. This was one assumption made for iterating over rain observations. Similar to calculating the cloud mass, rain mass was calculated and a threshold value of 0.1 g/m^3 was used to filter out rain droplets inside and outside the cloud. One more important assumption made was droplet sizes ranging from 3 – 50 µm were that of cloud and not of rain drops thus, all the columns within this range were dropped too from the calculations.

Data Collation

The next step is to collate all these variables from all the dates the observations are sourced from, into one dataframe. The functions created to calculate these variables are called repeatedly in 'for' loop for all the observation files, the appended values are flattened and each variables is added to a new data frame. Since, there are about 17 files, this increases the collation process time and makes it computationally complex. Thus, the collated file is stored as a csv file in the local backup folder.

The file now has the following variables cdp_mass, cdp_conc, ccn_conc, w_mean, w_std, temp and ambient temperature. From the last two variables we calculate the RH as follows:

$$100\,.\frac{e^{\frac{17.625\,*TD}{243+TD}}}{e^{\frac{17.625\,*T}{243+T}}}$$

TD = dew point temperature and T = ambient temperature.

It is important to compare our work with the literature and the correlation used to compare our work was given by Pinsky et al.[14]).

$$N = \,C_3{}^{\frac{2k}{2+k}}\,N_0{}^{\frac{2}{2+k}}\,w^{\frac{3k}{4+2k}}$$

C_3 is a complex function of temperature, N_0 = ccn_conc and w = updraft velocity.

We have also included a non-nan value counter in the velocity standard deviation function calculations to see how many real valued observations are contributing to the updraft speed. Hence, for comparison with the analytical studies we use w_std for updraft velocity calculations. It gives us a better estimate than the mean value as the distribution is widespread (see Figure 2).

Machine Learning Model

The final step was to try fitting a machine learning model. The aim of the project is to estimate the droplet concentration using the above mentioned variables. In feature engineering it is critical to understand the correlation between the variables. df.corr () function is used to determine the interdependence of each variable on other. It uses the Pearson method and creates a covariance matrix. After understanding the covariance, variables with least dependence can be dropped out. Now, with the remaining variables, the first obvious choice is to use linear regression and understand the linear dependence of variables on the y-values (cdp_conc). The next machine learning model used was Neural Networks. Scikit learn packages were used for linear regression and multi-layer perceptron regression (MLP or neural-nets). GridCV package from sklearn is used for hyper parameter optimization of MLP model, namely hidden layer sizes and learning rate.

Results

The covariance table is as follows (Figure 6):

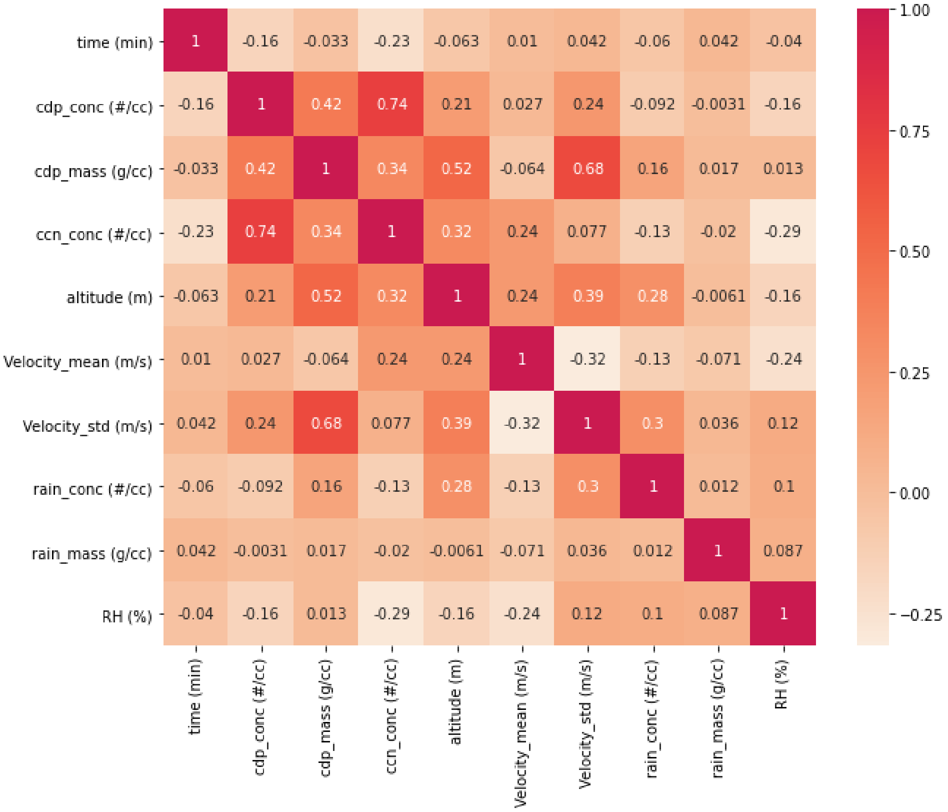


Figure 6: Covariance Table of all the variables with the droplet concentration.

As inferred from the cdp_conc row in covariance table above, the variables which adversely affect the increase in droplet concentration are RH and w_mean, but since their contribution is not very significant they were still kept in the feature matrix. Also, from the matrix it is clear that rain mass and concentration are poorly related to the any of the variables. When these variables were added to the covariance table, they had an adverse effect on the efficacy of the model.

Contrary to the concept that increase in rain drop concentration, the cloud drop concentration should decrease; the correlation was still weak and adverse.

This hypothesis was tested by plotting a 2D histogram of cloud drop concentration and cloud mass by dividing the dataset into rain and no rain data. The histograms (in Figure 7) show that during the rainy days there is steep increase in cdp_mass but a reduction in number of droplet concentration. This proves that as the droplets coalesce to form rain drops the mass increases but the droplet number should decrease. Whereas during a cloudy/sunny day, the number droplet concentration values from the histogram are linearly increasing with droplet mass values.

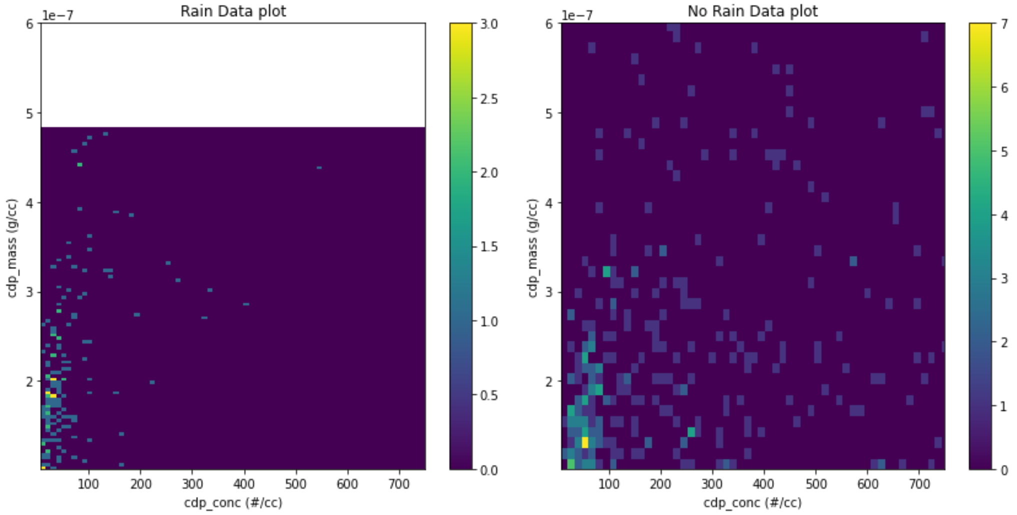


Figure 7: 2-D histogram of cloud droplet mass vs number concentration during a rainy and non - rainy day

It is however interesting to note that despite the fact that observations prove the above hypothesis they still have a poor dependency as seen in the covariance table. Thus, although there is a relation between the rain and cloud drop attributes but the dependency is weak. Hence, it was better to exclude them from the model (see Figure 8).

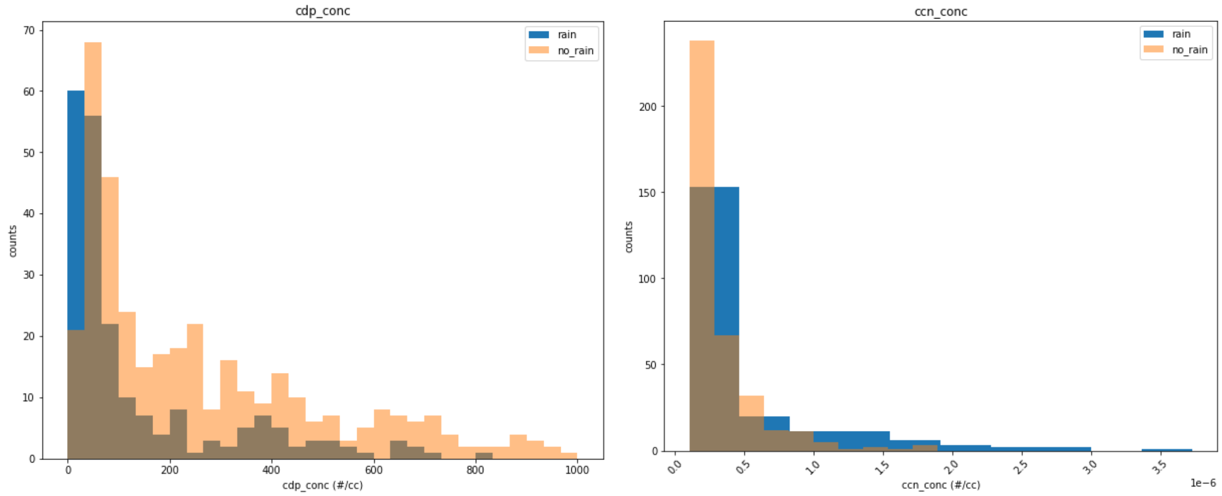


Figure 8: Histograms of droplet concentration and mass when dataset is divided as rain and no-rain.

The variables used as predictors of droplet concentration were added to a linear regression model (aka Feature Matrix). The linear regressor model gave a R^2 = 0.59 with all the variables (excluding time) from Figure 6. It is not a very bad predictor as compared to the Pinsky correlation (analytical). The comparison table and Figure 9 better explain the analyses.

Table 1: Regression and Pinksy Comparison plots

Comparison plots	**Observed drop conc. vs F.M**	**Pinsky vs Observed drop conc.**	**Regressor Predicted vs Observed drop conc.**
R^2 values (%)	59	45	49

Note: F.M = Feature Matrix

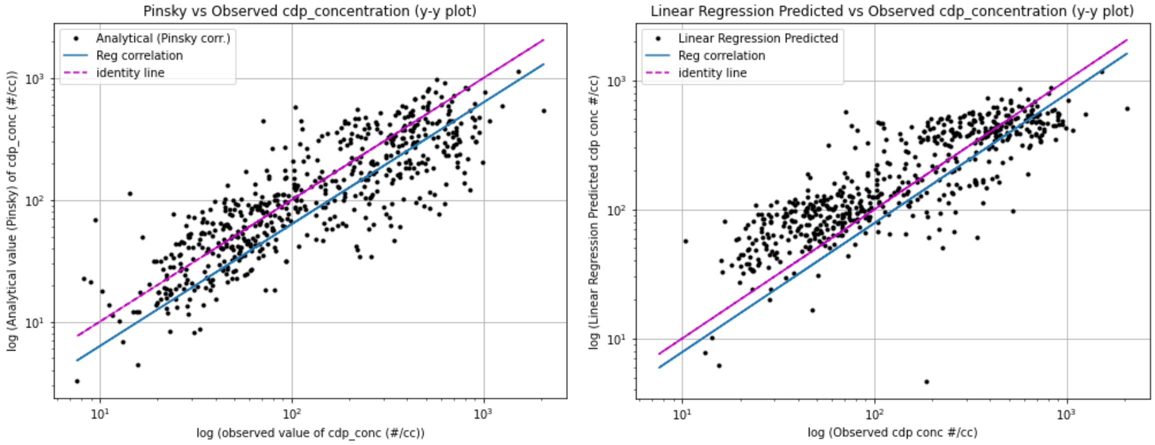


Figure 9: log-log plots of droplet concentration between observed, predicted by analytical (Pinsky) correlations and Linear Regression.

Figure 9 shows the log-log comparison plots of the observed target values (droplet concentration) with Pinsky correlated droplet concentration and Linear Regression predicted droplet concentration. The observations seemed to be clustered around the range 5 – 200 #/cc and sparse later. Thus, a log normalization is helpful for better visualization. The regression predicted target values have 4% improvement from the analytically correlated drop concentration values. Thus, the features used to determine the concentration significantly reduce the number of variables required for estimation. An inference about the outliers can also be made from the plot of Linear regression vs observed cdp_conc. Extremely low concentration like 2 #/cc (bottom left) and higher values of 200 #/cc (bottom right) are both skewing the regression fit which make them possible outliers.

The linear regression results were put to test with a more complex non-linear multi-layer perceptron or Neural Networks. After optimizing the hyper parameters using GridSearch CV package in sklearn the results were to use two hidden layer neural network of 6 neurons each and constant learning rate as 0.05. The results from the trained neural net was 5% improvement from the linear regression with a mean R^2 = 64% ($\pm 12\%$). A 5-fold Shufflesplit cross validation was applied and the activation function ‘relU’ along with ‘lbfgs’ gradient method were used to determine the complex parameters of the network. The results are tabulated below:

Table 2: Neural Net CV model score

CV fold	R^2 (Train set) (%)	R^2 (Test set) (%)
1	66	59
2	63	53
3	67	44
4	66	74
5	68	69

Note: The train and test scores are bound to change as the splitting is random but a random state of ‘0’ was assigned to the model to ensure reproducibility.

ShuffleSplit is random permutation cross-validator. It yields indices to split data into training and test sets. ReLU stands for rectified linear unit, and is a type of activation function. Mathematically, it is defined as y = max (0, x). Limited-memory BFGS (L-BFGS or LM-BFGS) is an optimization algorithm in the family of quasi-Newton methods that approximates the Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) using a limited amount of computer memory.

As also seen from Figure 10, the neural network target predictions vs the observed predictions seem to be better correlated with a score of 53%. It is definitely an improvement to the linear regression.

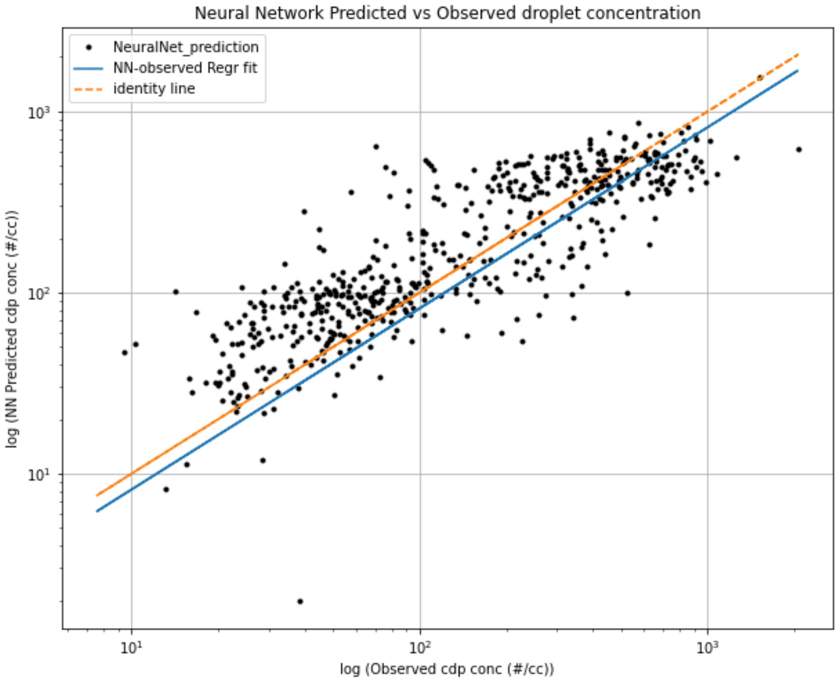


Figure 10: Target values comparison plot between neural network predictions and observed values.

NOTE: Python scripts for the research study can be found on this link.

Conclusion

In conclusion, the aircraft observations made during the GoAmazon campaign of 2014/15 were used for analysis. When comparing aircraft to ground site data, it is important to acknowledge differences in the datasets. First, the aircraft samples a different spatial domain in all three dimensions (latitude, longitude, and altitude) than a ground site. In addition, the G-1 typically sampled the boundary layer in the late morning (approximately 10:00 – 11:30 local time), and thus may miss peak concentrations of secondary species. Finally, the goal of the flights was often to sample the Manaus plume for the first half of the flight and clouds for the second half. The most important variables in the estimation of droplet concentration are the droplet mass, cloud condensation nuclei concentration, relative humidity, updraft wind velocity and rain drop concentration. A non-linear correlation provided by Pinsky et al. [14] was compared with linear regression and a much more complex 2 hidden layer neural network; which proved to be a better predictor than the other predictors. The study can be further extended by studying the effects of Decision Trees, Random Forests and other such regressors on the given data. It is important to note that the data sets used for analysis is not sufficient to accurately train the model. Addition of more flight days from later campaigns can significantly improve the model's accuracy. However, this a great initiative and novel approach to estimate droplet concentration in the atmosphere using aircraft observation instead of radar measurements.

Bibliography

(1) Hansen, J.; Sato, M.; Ruedy, R. Aerosol Forcing of Climate. *J. Bjerknes, Mon. Weather Rev* **1995**, *376* (March), 345–369.

(2) Hobbs, P. V.; Rangno, A. L. Ice Particle Concentrations in Clouds. *J. Atmos. Sci.* **1985**. https://doi.org/10.1175/1520-0469(1985)042<2523:IPCIC>2.0.CO;2.

(3) Grooters, F. Aircraft Observations https://public.wmo.int/en/bulletin/aircraft-observations (accessed Dec 16, 2020).

(4) Chen, Q.; Farmer, D. K.; Rizzo, L. V.; Pauliquevis, T.; Kuwata, M.; Karl, T. G.; Guenther, A.; Allan, J. D.; Coe, H.; Andreae, M. O.; Pöschl, U.; Jimenez, J. L.; Artaxo, P.; Martin, S. T. Submicron Particle Mass Concentrations and Sources in the Amazonian Wet Season (AMAZE-08). *Atmos. Chem. Phys.* **2015**. https://doi.org/10.5194/acp-15-3687-2015.

(5) Yeom, J. M.; Yum, S. S.; Mei, F.; Schmid, B.; Comstock, J.; Machado, L. A. T.; Cecchini, M. A. Impact of Secondary Droplet Activation on the Contrasting Cloud Microphysical Relationships during the Wet and Dry Seasons in the Amazon. *Atmos. Res.* **2019**. https://doi.org/10.1016/j.atmosres.2019.104648.

(6) Baker, M. B.; Corbin, R. G.; Latham, J. The Influence of Entrainment on the Evolution of Cloud Droplet Spectra: I. A Model of Inhomogeneous Mixing. *Q. J. R. Meteorol. Soc.* **1980**. https://doi.org/10.1002/qj.49710644914.

(7) Slawinska, J.; Grabowski, W. W.; Pawlowska, H.; Morrison, H. Droplet Activation and Mixing in Large-Eddy Simulation of a Shallow Cumulus Field. *J. Atmos. Sci.* **2012**, *69* (2), 444–462. https://doi.org/10.1175/JAS-D-11-054.1.

(8) Heus, T.; van Duk, G.; Jonker, H. J. J.; Van den Akker, H. E. A. Mixing in Shallow Cumulus Clouds Studied by Lagrangian Particle Tracking. *J. Atmos. Sci.* **2008**. https://doi.org/10.1175/2008JAS2572.1.

(9) Martin, S. T.; Artaxo, P.; MacHado, L. A. T.; Manzi, A. O.; Souza, R. A. F.; Schumacher, C.; Wang, J.; Andreae, M. O.; Barbosa, H. M. J.; Fan, J.; Fisch, G.; Goldstein, A. H.; Guenther, A.; Jimenez, J. L.; Pöschl, U.; Silva Dias, M. A.; Smith, J. N.; Wendisch, M. Introduction: Observations and Modeling of the Green Ocean Amazon (GoAmazon2014/5). *Atmos. Chem. Phys.* **2016**. https://doi.org/10.5194/acp-16-4785-2016.

(10) Schmid, B.; Tomlinson, J. M.; Hubbe, J. M.; Comstock, J. M.; Mei, F.; Chand, D.; Pekour, M. S.; Kluzek, C. D.; Andrews, E.; Biraud, S. C.; McFarquhar, G. M. The DOE Arm Aerial Facility. *Bull. Am. Meteorol. Soc.* **2014**, *95* (5), 723–742. https://doi.org/10.1175/BAMS-D-13-00040.1.

(11) Wiki.earthdata.nasa.gov. GIBS Available Imagery Products - Global Imagery Browse Services (GIBS) - Earthdata Wiki https://wiki.earthdata.nasa.gov/display/GIBS/GIBS+Available+Imagery+Products#expand-CloudHeight10Products (accessed Jul 12, 2020).

(12) Worldview. Worldview: Explore Your Dynamic Planet https://worldview.earthdata.nasa.gov/?v=158541.0307560143,724236.6511193844,580586.0959365916,1358323.6814148412&r=-14.3679&p=arctic&l=Reference_Labels(hidden),Reference_Features(hidden),Coastlines,VIIRS_NO

AA20_CorrectedReflectance_TrueColor(hidden),VIIRS_SNPP_CorrectedReflectance_TrueColor(hidden),MODIS_Aqua_CorrectedReflectance_TrueColor,MODIS_Terra_CorrectedReflectance_TrueColor (accessed Jul 12, 2020).

(13) Thompson, A. Simulating the Adiabatic Ascent of Atmospheric Air Parcels Using the Cloud Chamber. **2007**.

(14) Pinsky, M.; Khain, A.; Mazin, I.; Korolev, A. Analytical Estimation of Droplet Concentration at Cloud Base. *J. Geophys. Res. Atmos.* **2012**. https://doi.org/10.1029/2012JD017753.