# *“ Analysis and Visualization of Atmospheric Data ”*

## B.E. Project Report

Submitted in partial fulfillment of the requirements

For the degree of

**Bachelor of Engineering**
**in**
**Computer Engineering**

Submitted by


**Mr. Shubham Belgaonkar 16CE2003**

**Mr. Bhushan Bamble 16CE1049**


Guided by


**Dr. Amit Barve (Associate Professor, RAIT, Navi Mumbai)**

**Dr. Bipin Kumar (Scientist ’E’, IITM, Pune)**

**Mr. Manmeet Singh (Scientist ’C’, IITM, Pune)**

**Department of Computer Engineering**

**Ramrao Adik Institute Of Technology**

Dr. D. Y. Patil Vidyanagar, Sector-7, Nerul, Navi Mumbai-400706.

(Affiliated to University of Mumbai)

**Indian Institute of Tropical Meteorology,Pune**

Dr. Homi Bhabha Road, Pashan, Pune 411 008, Maharashtra, India


**April 2020**

# Analysis and Visualization of Atmospheric Data

## B.E. Project Report

Submitted in partial fulfillment of the requirements

For the degree of

**Bachelor of Engineering**
**in**
**Computer Engineering**

Submitted by


**Mr. Shubham Belgaonkar 16CE2003**

**Mr. Bhushan Bamble 16CE1049**


Guided by


**Dr. Amit Barve (Associate Professor, RAIT, Navi Mumbai)**

**Dr. Bipin Kumar (Scientist 'E', IITM, Pune)**

**Mr. Manmeet Singh (Scientist 'C', IITM, Pune)**

**Department of Computer Engineering**

**Ramrao Adik Institute Of Technology**

Dr. D. Y. Patil Vidyanagar, Sector-7, Nerul, Navi Mumbai-400706.

(Affiliated to University of Mumbai)

**Indian Institute of Tropical Meteorology,Pune**

Dr. Homi Bhabha Road, Pashan, Pune 411 008, Maharashtra, India


**April 2020**

# Ramrao Adik Institute of Technology

(Affiliated to the University of Mumbai)

Dr. D. Y. Patil Vidyanagar, Sector-7, Nerul, Navi Mumbai-400706.

## CERTIFICATE

*This is to certify that, the project 'B' titled*

**" Analysis and Visualization of Atmospheric Data "**

*is a bonafide work done by*

**Mr. Shubham Belgaonkar 16CE2003**
**Mr. Bhushan Bamble 16CE1049**

*and is submitted in the partial fulfillment of the requirement for the degree of*

**Bachelor of Engineering**
in
**Computer Engineering**
to the
**University of Mumbai**

| Supervisor | Co-Supervisor | Co-Supervisor |
|---|---|---|
| **(Dr. Amit Barve )** | **(Dr. Bipin Kumar)** | **(Mr. Manmeet Singh)** |

| Project Co-ordinator | Head of Department | Principal |
|---|---|---|
| **(Mrs. Smita Bharne )** | **(Dr. Leena Ragha)** | **(Dr. Mukesh. D. Patil)** |

# Project Report Approval for B.E

This is to certify that the project 'B' entitled ***" Analysis and Visualization of Atmospheric Data "*** is a bonafide work done by ***Mr. Shubham Belgaonkar*** and ***Mr. Bhushan Bamble*** under the supervision of ***Dr. Amit Barve***, ***Dr. Bipin Kumar*** and ***Mr. Manmeet Singh***. This dissertation has been approved for the award of ***Bachelor's Degree in Computer Engineering, University of Mumbai***.

Examiners :

1. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

2. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Supervisors :

1. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

2. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

3. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Principal :

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Date : . . . /. . . /. . . . . .

Place : . . . . . . . . . . . .

# Declaration

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

**Mr. Shubham Belgaonkar** **16CE2003** ____________________

**Mr. Bhushan Bamble** **16CE1049** ____________________

Date : .../.../......

# Abstract

Data visualization uses charts, graphs and maps to illustrate some information. Visualizing data helps us understand information faster. This places strain on software developers to create a visualization of complex datasets accurately under the guidance of meteorological scientists. The study of droplet dynamics is a critical part of cloud physics that includes studying droplet properties. The aim of this work is to visualize the droplet dynamics obtained from DNS (Direct Numerical Simulation) data due to the evaporation and condensation of the droplets. This simulation contains coupled Eulerian and Lagrangian frames. Animation is created for both Eulerian grid data and Lagrangian droplet movement. Scientific visualization provides a way to analyze these turbulent properties in a part of a cloud and learn about the nature of droplets and the mixing process in such highly turbulent areas.

# Contents

# List of Figures

# Chapter 1

# Introduction

Weather forecasting centers manage a large amount of meteorological data. Such data must be interpreted within a short period of time. Scientific visualization can be a great help in supporting meteorological researchers and scientists. Decision-making depends on data that comes in a large quantity that it cannot be understood without some level of abstraction, such as a visual one. There are various types of meteorological data namely rainfall data, temperature data, radiation data, radar data and many more. This work involves examining various droplets' properties. It is difficult to examine the properties by just looking at Direct Numerical Simulation. Direct Numerical Simulation is a method in Computational Fluid Dynamics where the range of spatial and temporal scales of the turbulence is resolved. Small scale simulation is one method in understanding the physics behind the phenomena and generate data. This work shows the data obtained from Direct Numerical Simulation of entrainment and mixing processes in the cloud affect the droplet dynamics due to condensation and evaporation. The turbulent entrainment and mixing of dry air with moist or cloudy air influences the cloud's lifespan by changing the properties. Such modifications are responsible for the droplet dynamics. This mixing process is quite complex, and it is difficult to study it during field observations. It illustrates how all of the droplets will evaporate by the same number, or the subset of the droplets evaporates entirely, while the other droplets remain intact in the cloud turbulence. However, it is not clear how the size of the droplets is influenced by the mixing process and has a major impact on the spectrum of the droplets. The types of mixing processes of the droplets are known as homogeneous mixing and in-homogeneous mixing[1].

Many visualization techniques can be used to visualize the given data[2]. This work uses ParaView and Avizo for visualization because of their ability to analyze extremely large datasets. ParaView is an open-source application while Avizo is a commercially available software and both are used for the analysis of data and visualization. They can be used on supercomputers to process large data and also on laptops for smaller data. They can quickly build visualization pipelines for the analysis of data using both qualitative and quantitative techniques.

## 1.1 Overview

The scripts are designed to accurately visualize mixing in cloud turbulence in 3D. The input given for Paraview is Lagrangian data and Eulerian data in VTK and NetCDF formats respectively. Whereas for Avizo, the same input is given in CSV and NetCDF formats respectively. Lagrangian and Eulerian are the two approaches to describe fluid flow. Eulerian data acts as a container where only the particles inside the container are being observed. In this case, Eulerian data is the grid data for cloud formation. Animation is created for both of these data in both the visualization software using the python bindings.

## 1.2 Objective

The objective of this project is

- To study various data formats like NetCDF, GRIB and VTK
- To study the visualization softwares 'Paraview' and 'Avizo'
- To write a script to generate quality visualization of the Lagrangian and Eulerian data
- To create animation of the visualizations obtained

## 1.3 Motivation

For decades, from maps and graphs to the invention of the pie chart, the idea of using images to explain data is used. Computers have allowed large quantities of data to be processed quickly. The visualization of data today has become a rapidly evolving mix of science. Visualization is an important instrument in the everyday work of weather forecasters and atmospheric researchers, from the meteorological perspective. Typical roles in modern meteorology include reviewing data to understand the meteorological situation or a particular atmospheric mechanism. Even a bad visualization can be useful if it suits the audience's objectives. Effective visualization helps viewers to spot patterns they would never see by using numerical data.

Thus, animating the visualization of droplet movements would make us understand the data more clearly than the numerical representation. This system provides a way to visualize this data.

## 1.4 Organization of report

The report is divided into the following chapters:
Chapter 1: This chapter introduces the project topic,purpose of this project and what is being done in the project.
Chapter 2: This chapter contains all the literature review, analysis of various file formats, surveying existing methodology and problem statement.
Chapter 3: The proposal chapter contains the proposed methodology along with a diagram which illustrate the workflow of the system. This chapter also specifies hardware and software requirements of the system.
Chapter 4: This chapter gives the tentative timeline for the completion of the project.
Chapter 5: This chapter contains the detailed explanation of the system design and flow of the system along with the data flow diagram.
Chapter 6: This chapter has results that we are expecting as an outcome.
Chapter 7: Conclusion. This chapter has final concluding words of the report.
Chapter 8: Future work. Work that can be done in the future to make the project better.
Chapter 9: References. This chapter has a list of research papers, journals, articles and links used for the project.

# Chapter 2

# Literature Survey

## 2.1 Survey of various scientific file formats

The project involves the study of various data formats and visualization tools ParaView and Avizo. Given below is a brief summary of the above.

### 2.1.1 NetCDF File Format

NetCDF is the popular data format defined by Unidata program at the University Corporation for Atmospheric Research (UCAR), used widely in climatic, meteorological and oceanographic applications such as climate change and forecasting of weather, etc.

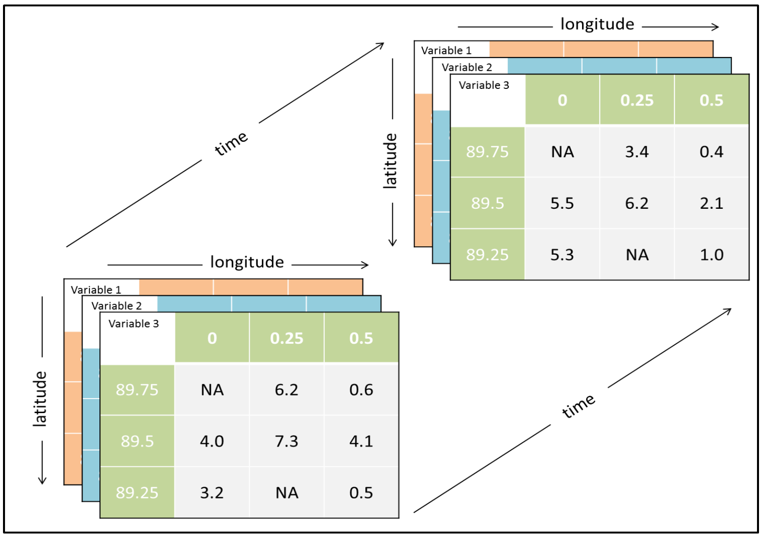


Figure 2.1: Basic structure of NetCDF file[3]

It is a data format that is created to support the array-oriented data[4]. All NetCDF formats contain a header that describes the layout of the rest of the file, particularly the data arrays.

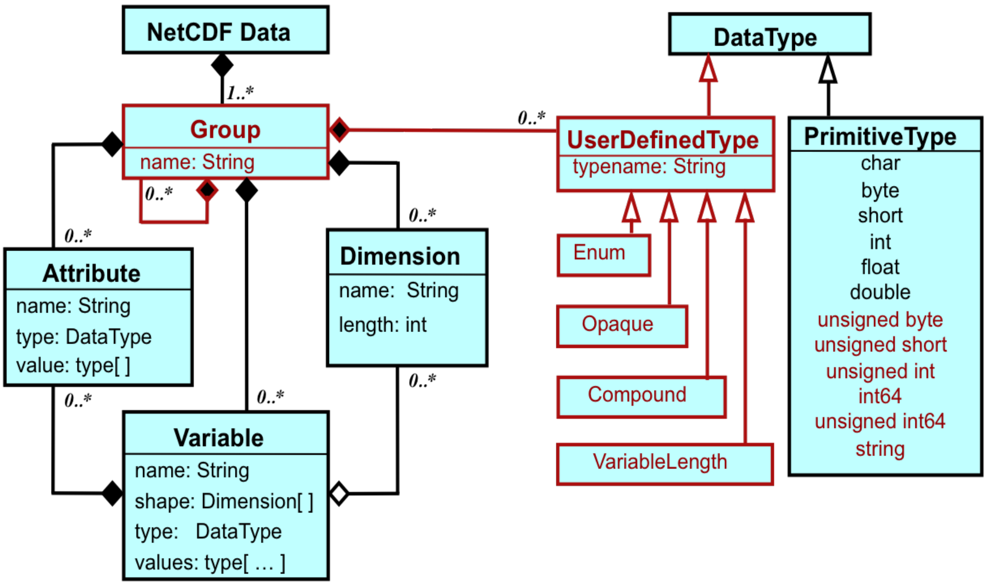


Figure 2.2: Data model of NetCDF file[5]

The NetCDF model of data abstraction creates scientific data set as a group of named multidimensional variables along with their coordinate systems. Every NetCDF file has three components: dimensions that are used to define the shape of data, variables to store the array values, and attributes to provide a title or this history of the data set. They help understand the data more clearly. These NetCDF files can be accessed in python with the help of libraries such as 'xarray' and 'netCDF4'[6][7]. Xarray introduces the dimensions, coordinates and attributes on raw Numpy-like arrays. It is heavily inspired by pandas.

### 2.1.2 GRIB File Format

GRIB (Gridded Binary or General Regularly-distributed Information in binary form) is a data format that is widely used in meteorology to store historical and weather predictions. It is a uniform format prepared by the Commission for Basic Systems of the World Meteorological Organization. GRIB files are a set of self-contained 2D data records, and individual records are standing alone as valid data, without references to other records or an overall scheme. So GRIB record collections can be appended to each other or isolated records[8].

There are two components in each GRIB record the section that defines the record (the header) and the real binary data itself. Each record contains information such as grid resolution, time, variable, level who generated the field. GRIB decoders are available in a set. For the analysis of GRIB data format files, the 'pygrib' module is used. This module has a dependency of Linux operating system. It returns the list containing pygrib.messages object with which one can access the properties such as unit, dimensions, shape, values of the variable.GRIB is slightly more compact than netCDF files but is difficult to process. NetCDF is preferred because it can handle more than two dimensions and is comparatively faster than GRIB as it has fewer data stored there. Also, there are standard programming interfaces for NetCDF.

### 2.1.3 Vol File Format (.vol extension)

Vol is a simple file consisting of a header and a set of values storing volume data. The header is of 10,000 bytes. First five characters identify the file as a volume with this file type and other three to indicate the initial display. 'wradlib'[9] is a weather radar data processing library in python which can access these type of file formats but after thorough studying of this library it was found there is no reader of this extension. Vol files are encoded so they cannot be accessed by normal file I/O in python and need to be decoded first.

### 2.1.4 VTK (Visualization ToolKit) files

Visualization ToolKit is a powerful high-level visualization library for 3D graphics, image processing and visualization. VTK was created in 1993 and is maintained by Kitware Inc. VTK can be used by anyone wanting to visualize data but is commonly used by data visualization engineers and data scientists. It offers numerous writer object with the ability to read and write standard file formats along with their own file formats. The purpose of creating other data file format is to provide a consistent data representation scheme for a number of data set forms and to provide an easy method for transmitting data[10].

There are two file formats available in VTK, Legacy and Serial. These formats are less flexible as compared to XML file formats. While these formats are much more complicated than the Legacy format. By just providing a simple XML description any VTK source or filter can be added. Python has a library called 'pyevtk' which is used to write the x, y, z and radius points to VTK file by using the function 'pointsToVTK'.

Some of the VTK data set types are: structured grid, structured points, unstructured grid, rectilinear grid, polydata and field. This work involves the use of unstructured grid. In unstructured grid there is no regularity in the position of the points. One has to specify the number of points in the mesh and the data type used to specify each point. After that the positions of each point are to be specified. Lastly, how are these points connected together to form cells has to be specified. This is done by specifying the number of cells in the data set and the number of points in each cell. This is the domain structure of unstructured gird. Given below are the representations of the VTK data sets.

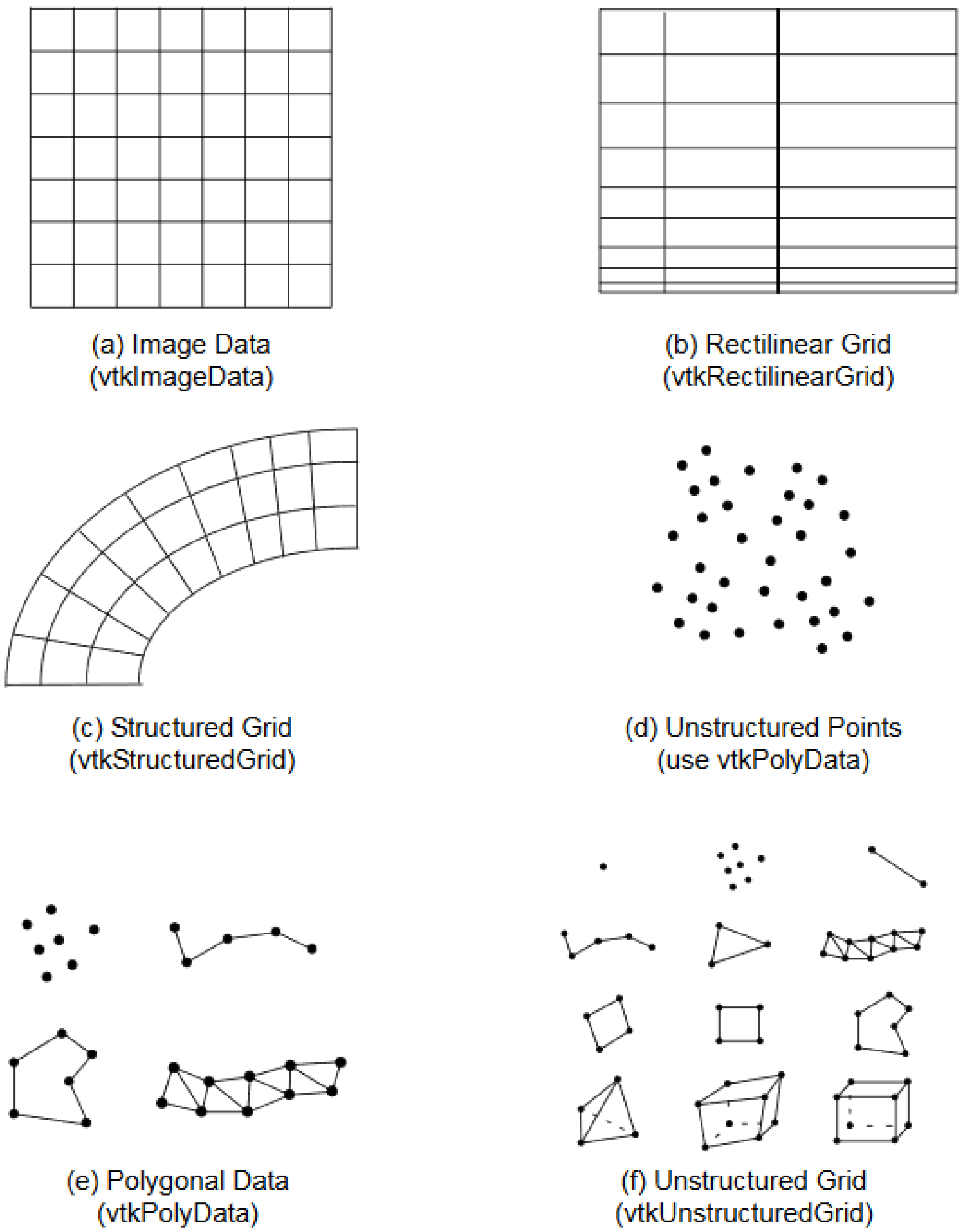


Figure 2.3: Types of data sets in VTK[10]

## 2.2 Survey of Visualization softwares

### 2.2.1 ParaView

ParaView is one of the most preferred open-source scientific visualization tools for visualization. This software is used in many communities because of its easy to use nature to analyze and visualize large data sets. ParaView uses three steps to visualizing data: reading or loading the data files, filtering the necessary details, and then rendering the images for the given data. Reading data involves loading the data files which creates pipelines for the loaded data. Filtering the data involves creating arrays or using arrays which are necessary for the visualization. Rendering involves generating and saving the images for the data set. ParaView has Python, C, C++ and Fortran bindings. In this work python shell is used for the visualization process. From 'Tools' select 'Python shell', then select 'Run file' and run the python script to obtain the visualization. The screenshots of the visualizations can be saved for producing animation.

### 2.2.2 Avizo

Avizo is general-purpose commercial software for 3D analysis of scientific data. Avizo is more focused on applications of physical and earth sciences. By using Avizo, one can do interactive visualizations on 3D datasets. Avizo allows data and modules to be connected interactively and also be controlled with specific parameters output for which is displayed in the 3D viewer window. For interactive visualizations, custom scripts can be developed in either TCL or Python bindings. In this research work, Python console scripting has been used for visualization. Python console can be found under the console tab. Python scripts are developed and stored externally and later imported into the Python console by copying the script from the external file and are executed to obtain the visualization.

## 2.3 Problem Statement

This work aims to propose a procedure for the visualization of the droplet dynamics in the cloud obtained from the DNS (Direct Numerical Simulation) data in a visualization software application to help the scientists and researchers to examine the properties of the droplets in cloud turbulence.

# Chapter 3

# Project Proposal

## 3.1 Proposed Work

The aim is to write scripts that provide 3-D visualization for DNS (Direct Numerical Simulation) data provided in SION and netCDF formats. Reading of the SION data is done using FORTRAN library SIONlib[11]. However, SION format is not accepted in both ParaView and Avizo. For ParaView, SION data is converted to VTK using python library 'pyevtk'. For Avizo, SION data is converted to CSV data. After conversion, these VTK/CSV and NetCDF files are used to obtain visualizations in both the software. These visualizations are saved in the form of images which are further used to create an animation of the movement of droplet particles.

## 3.2 Proposed Methodology

### 3.2.1 Paraview

The input data consists of two datasets - Eulerian NetCDF files and Lagrangian SION files. As far as NetCDF files are concerned, there is an inbuilt NetCDF reader in ParaView software. For SION files, there are no readers as such in ParaView. So, to load SION data, SION files have to be converted into VTK format which is the default format for ParaView. Then it is to be further read in vtkUnstructuredGrid format. The data set forms a topologically irregular set of points and cells. After the conversion to VTK, a variable known as mixing ratio is extracted from the NetCDF files and new series of NetCDF files are created which are smaller in size.

This allows ParaView to process the files faster. Now, both the VTK and extracted NetCDF files and loaded in ParaView and visualization of the data is obtained at various time steps. The purpose of ParaView is to analyze extremely large data sets. The visualizations at every time step are saved as images which are further used for the animation purpose.

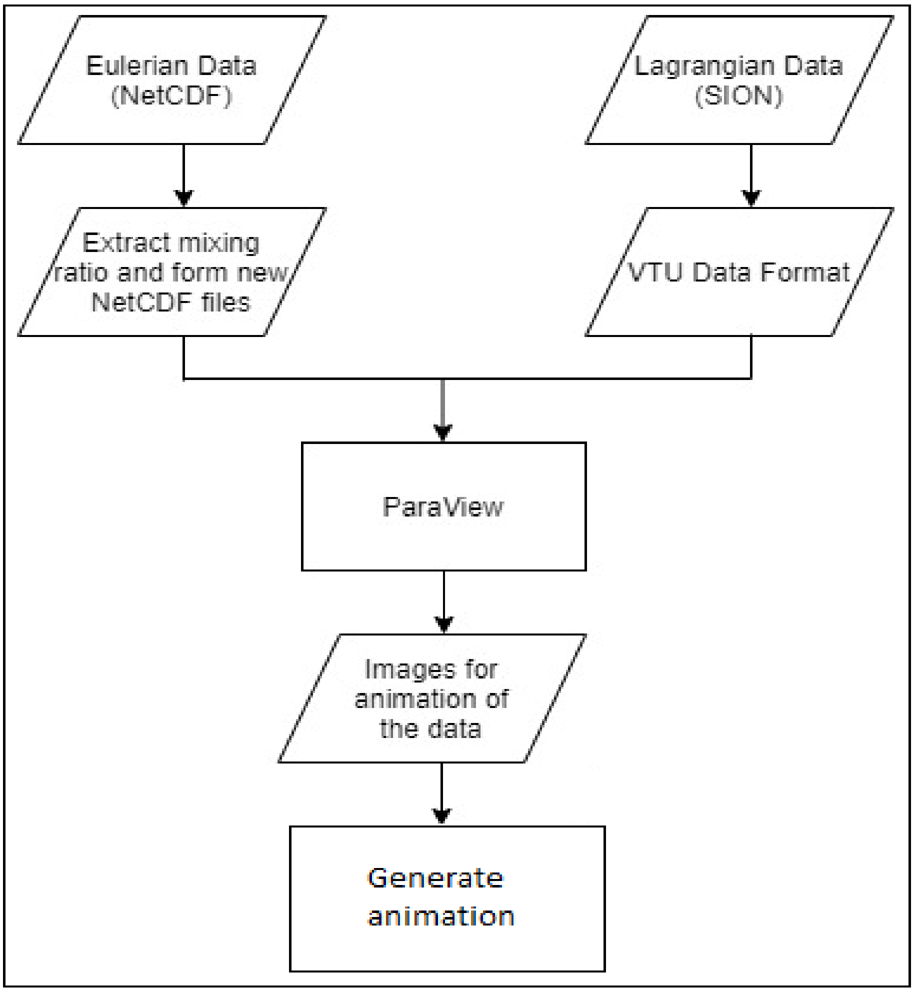


Figure 3.1: Basic flowchart of proposed methodology in ParaView

### 3.2.2 Avizo

The original input data is similar to that in ParaView, i.e. Eulerian NetCDF files and Lagrangian SION files. NetCDF files can be read in Avizo directly. Avizo does not support SION data format and thus it has to be converted into a format acceptable by Avizo like CSV. A variable known as mixing-ratio is extracted from each NetCDF file and a new series of NetCDF files is created. These NetCDF files are ready to be processed and are also much memory-friendly as compared to original files. The modified Eulerian NetCDF and Lagrangian CSV files are loaded in Avizo and visualizations are obtained for each time-step. Avizo can analyze and generate outputs for extremely large data sets. The images captured are then used for creating an animation.

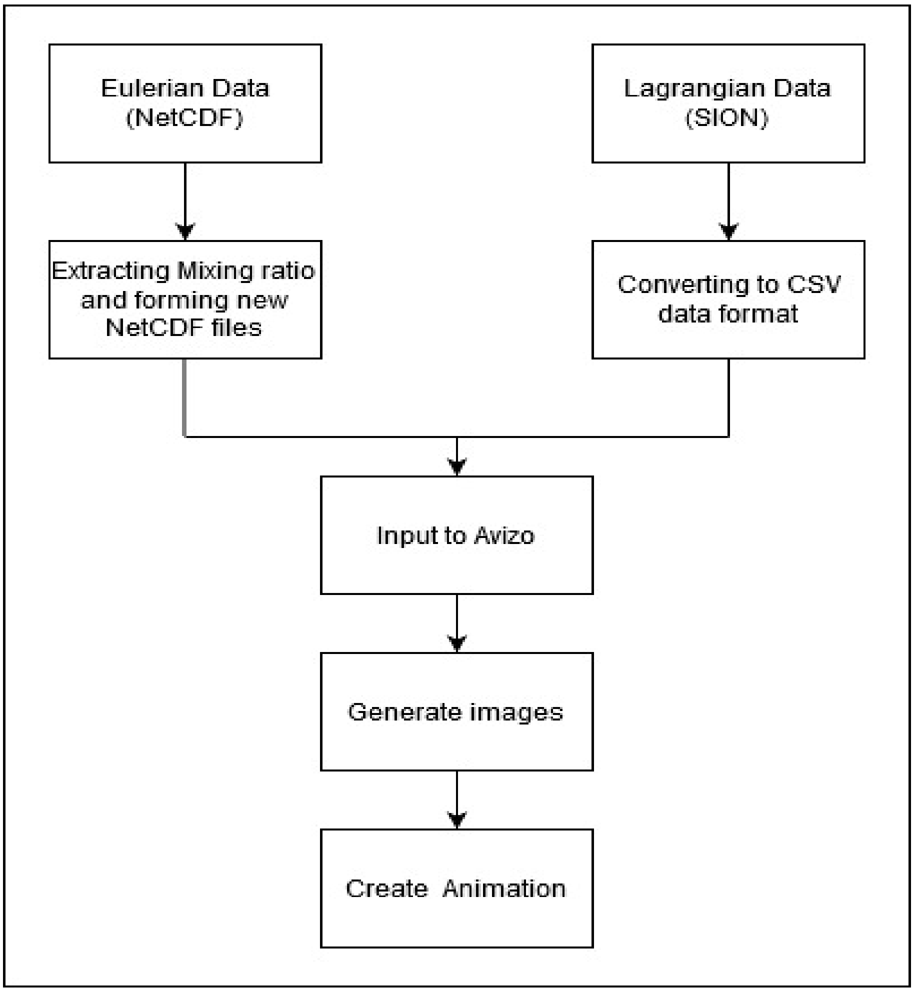


Figure 3.2: Flowchart of proposed algorithm in Avizo

## 3.3 Details of Hardware/Software Requirement

**Hardware Requirements:**

- 50 GB or greater RAM.
- Graphics card for better rendering.
- Disk space depending on the size of input data.

**Software requirements:**

- Linux OS
- Python 3.6 and above
- Paraview 5.2.0 and above
- Avizo version 2019.2 and above
- xarray (for accessing netCDF files)
- opencv (for the animation of images)

# Chapter 4

# Planning And Formulation

## 4.1 Schedule for Project

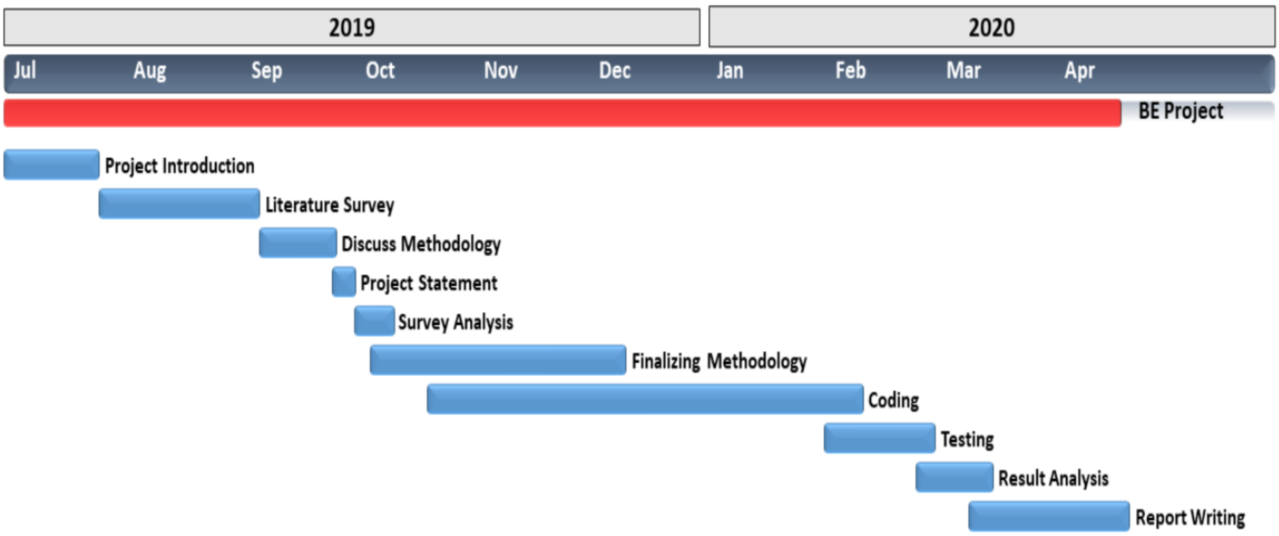


Figure 4.1: Gantt chart of project schedule

## 4.2 Detail Plan of Execution

The proposed project work plan and time schedule is as follows:

1. Requirement Analysis Phase(July 2019 to August 2019)
2. Analysis Phase (August 2019 to October 2019)
3. Designing Phase (October 2019 to November 2019)
4. Coding Phase (October 2019 to January 2020)
5. Testing Phase (January 2019 to February 2020)
6. Report Writing Phase (February 2019 to March 2020)

# Chapter 5

# Design of System

## 5.1 Process flow Diagram with Explanation

The DNS data stored in SION format could be converted to ASCII files using SIONLib in Fortran routine. ASCII files are also referred to as text files that can be read easily. They often contain letters, numbers and some line separators such as comma or spaces. ASCII files may be larger in file size but the speed of writing and reading is much faster. They are efficient, but one has to specify the file's content before reading or writing the files.

So, to open and visualize the data a program must be used. There are many bindings available in all the popular languages to access these text files. The procedure for the visualization of droplets dynamics includes extraction of the positions of the marker and radius of the droplets from the ASCII files. After extraction of the above properties, preparation of this unstructured data using Python script to create a series of VTK and CSV files is done. These can be coupled with the NetCDF Eulerian grid data files and imported into ParaView and Avizo respectively.

### 5.1.1 Process flow diagram for ParaView

ParaView uses three steps to visualizing data: reading or loading the data files, filtering the necessary details, and then rendering the images for the given data. Data must be read in an acceptable file format into ParaView. Next, the data must be analyzed in different ways available and is to be filtered with the number of filters available. Finally, the visible image from the data is rendered and the images at different time steps can be stored. Above steps are performed for every file present in the directory in a loop. All these images are stored in the image directory. Now, these images are used to form an animation video using python script. The process flow diagram below shows the flow of the whole process.

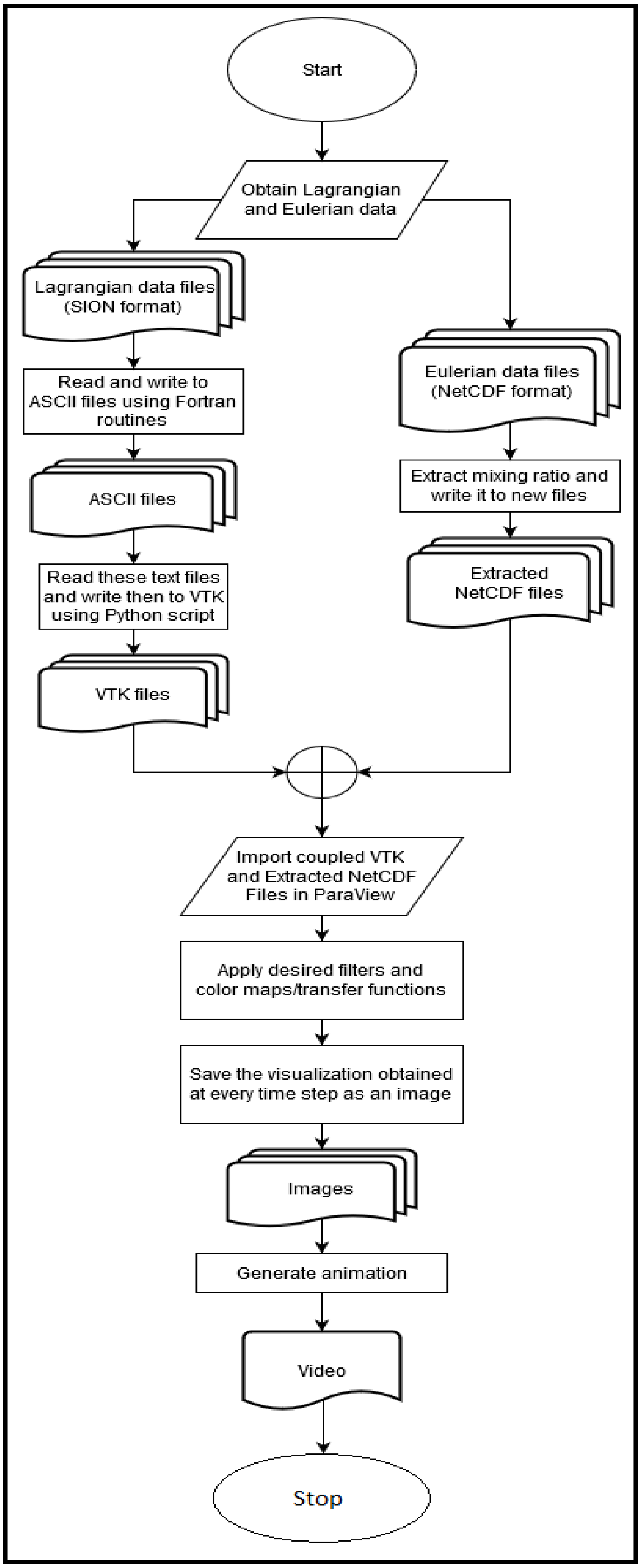


Figure 5.1: Process flow Diagram for ParaView

### 5.1.2 Process flow diagram for Avizo

Avizo uses the following steps in visualizing data: loading the data files, filtering the necessary details and then rendering the images for the given data[12]. The data files are first loaded into Avizo. Next, the data must be analyzed and processed in different ways available to obtain the optimal representation of the data. Finally, this optimal representation is rendered and the image is stored. The above steps are performed for every file present in the directory in a loop. All these images are stored in the image directory. Now, these images are used to form an animation video using python script. The process flow diagram below shows the flow of the whole process.

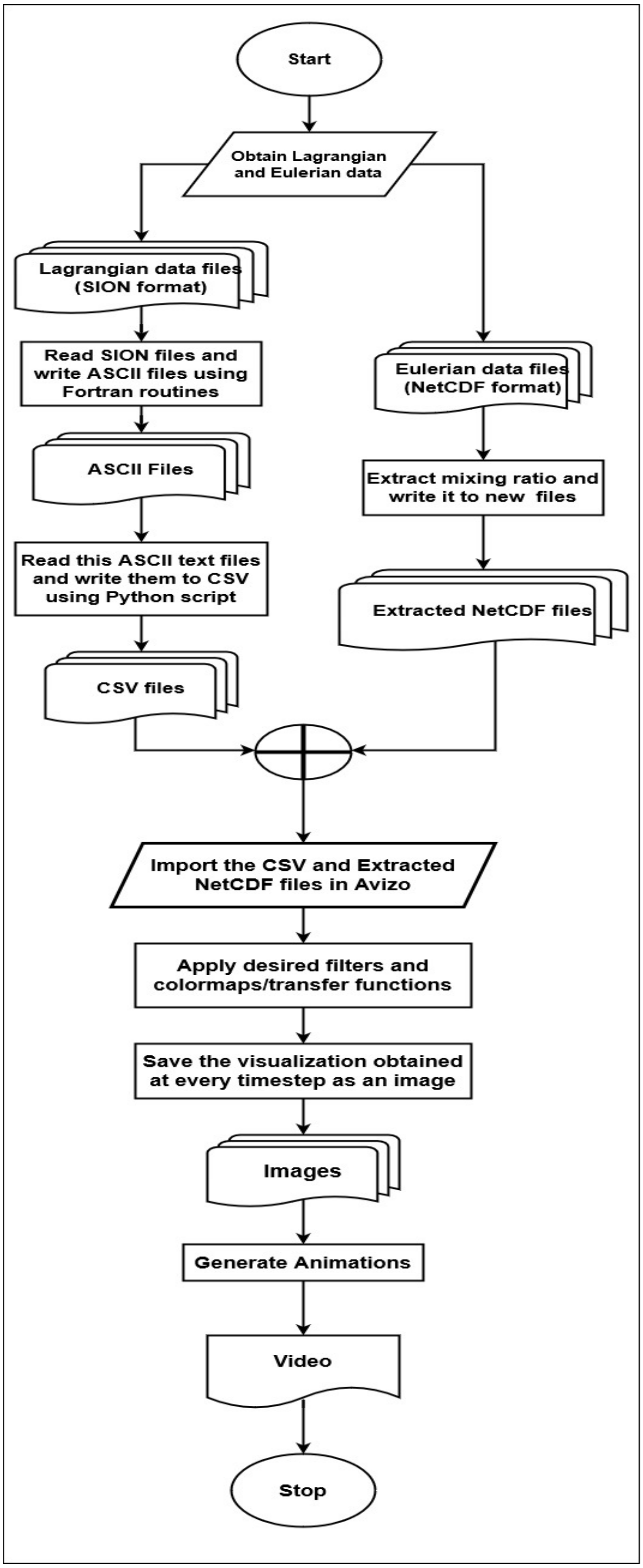


Figure 5.2: Process flow Diagram for Avizo

## 5.2 Implementation Details

### 5.2.1 For Visualization in Paraview

- Extract the mixing ratio variable from the Eulerian data files and create series of NetCDF files containing only mixing ratio data of same dimensions x, y, z. These Eulerian data points were reduced by spacing every two grid points. We extract this variable and form new NetCDF files to reduce the memory consumption by ParaView and also to speed up the computational time.

- Get the Lagrangian Data files converted to ASCII text files by extracting the marker positions and radius of the droplets using Fortran routine. SION files can only be read in Fortran using SIONlib.

- Write this unstructured droplet data present in these ASCII files, consisting of coordinate values of the droplet and its radius to VTK files using 'pyevtk' library in python.

- Load these coupled files in ParaView by creating NetCDF Reader for extracted Eulerian data and XMLUnstructuredGrid Reader for Lagrangian data.

- Set scalar coloring for both and create render view with best possible resolution. Results for this work have 2555 x 1376 pixels.

- Change representation type bye applying color maps/ transfer functions until desired visualization is obtained. Edit color legend/bar for the datasets respectively and choose the scale based on minimum and maximum values in data.

- Save the Screenshots of the desired visualization and run above steps in loop for each extracted Eulerian file and Lagrangian file.

- Use python program to animate the saved images and create an animation video.

### 5.2.2 For Visualization in Avizo

- Convert the Lagrangian data originally in SION format to ASCII and later to CSV format readable by Avizo.
- Extract the mixing ratio variable from Eulerian data files and create a new series of files containing only mixing ratio data with same dimensions x,y,z.
- Load these files in Avizo separately.
- Create Volume rendering and Point Cloud View pipelines for Eulerian and Lagrangian data respectively and connect them to the data files loaded.
- Select the appropriate colormap for the respective data files. Connect the colormaps to the pipelines created.
- Choose the scale of values for colormap legend based on the minimum and maximum values extracted from the data files of all time steps.
- Save the screenshot of the obtained result of visualization using the inbuilt Snapshot module.
- Run the above steps in loop for every timestep.
- Create a animation video of the saved images using a python program.

# Chapter 6

# Results And Discussion

## 6.1 Result analysis

### 6.1.1 Cloud Droplet Visualization in ParaView

In visualization the most important part is giving appropriate color. In ParaView, applying color map/transfer function is the part where the color is applied to the visualization. ”GetColorTransferFunction(’variable_name’)” and ”ApplyPreset(’colormap_name’)” functions help to apply the color map/transfer function to the variables. A gray scale color map is imported (eul_gray.8) for the visualization of cloud slab. Usually the cloud slab is represented by white coloration. The range of the mixing-ratio variable is decided to be 0 to 0.0012. For radius variable ’warm to cool’ color map/transfer function is used, which is already present in ParaView. This color map is useful to highlight the contrast between smaller radii and larger radii. It would otherwise be difficult to distinguish the difference in sizes due to the high density of the data. For the desired visualization of the droplet particles, the range of the radius data was changed many times. By trying several different ranges, the desired visualization was obtained when the range was 0 to 18. These ranges were set based on the values present in the data. ”RescaleTranferFunction(start_point, end_point)” function was used to set the range.

One can always start the trace window if they don’t know how to do something in python. Trace window helps to get the syntax of the function for any activity done in the Graphical User Interface. Use tools =>Start.

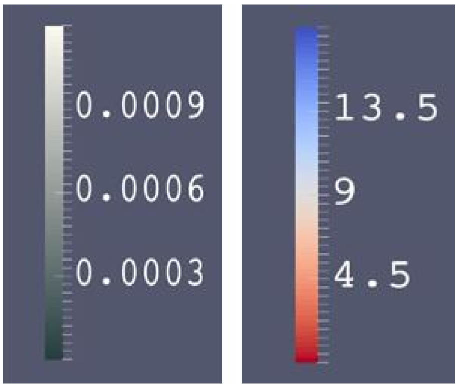


Figure 6.1: Color maps/Transfer functions

Shown in figure 6.1, is the color maps/transfer functions for mixing-ratio and droplet radius respectively. Scale of the color bar depends on the maximum and minimum values in the data. For mixing-ratio variable we use gray scale color bar and for radius variable we use 'warm to cool' color bar.

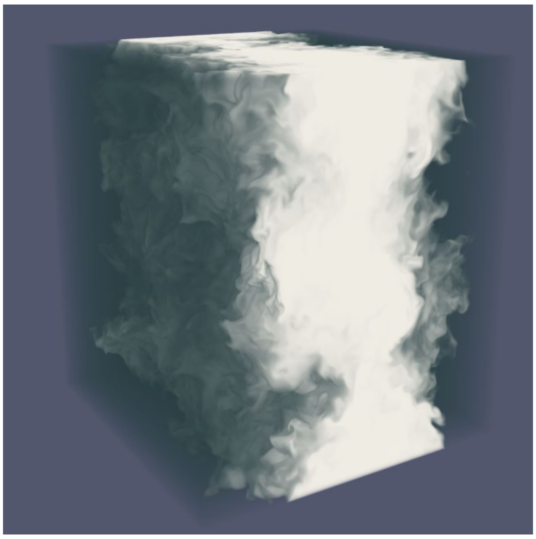

Figure 6.2: Visualization of Mixing-Ratio variable at time step 20

Shown in figure 6.2, is the visualization of Eulerian grid data colored by Mixing-Ratio variable. Central cloud slab is visible in white.

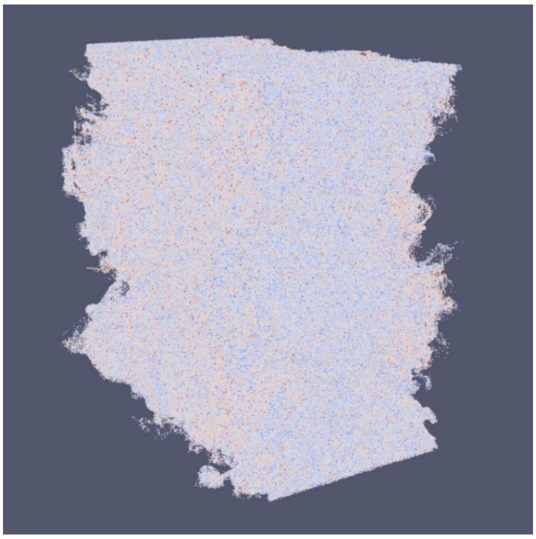

Figure 6.3: Visualization of radius variable at time step 20

Shown in figure 6.3, is the visualization of Lagrangian data colored by radius variable. It shows the visualization of droplets based on their radius size. Blue droplets have larger radii and the droplets in red are with the smallest radii.

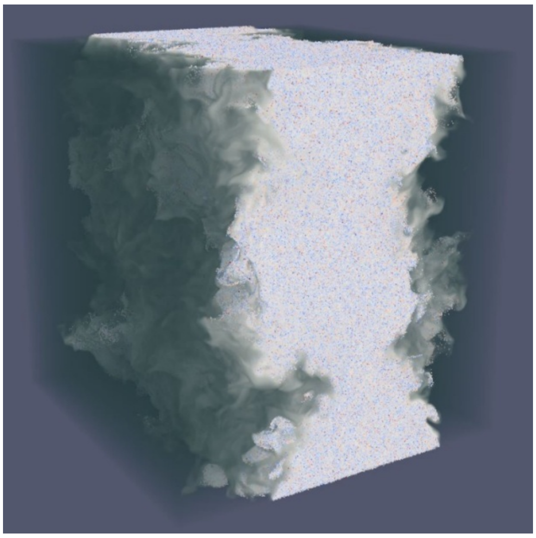

Figure 6.4: Visualization of Coupled Eulerian and Lagrangian files at time step 20

Shown in figure 6.4, is the visualization of coupled Eulerian grid data and Lagrangian data at time step 20. Both of these data files are imported simultaneously and visualized together. The droplets of different radii are visible in white cloud slab.

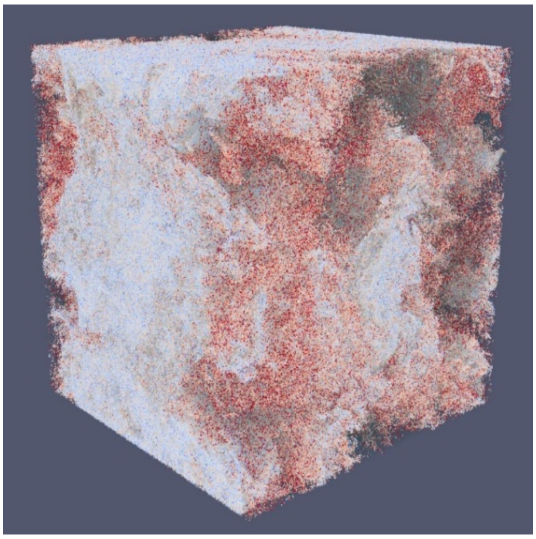

Figure 6.5: Visualization of Coupled Eulerian and Lagrangian files at time step 80

Shown in figure 6.5, is the visualization of Eulerian grid data and Lagrangian data together at time step 80. The droplets have started to evaporate and get smaller in size (red particles). Thus, the mixing process of the droplets in the cloud turbulence can be observed.

These results show the mixing process of the droplets in cloud due to evaporation and condensation. An animation is created of the visualizations obtained at every time step to help us understand the movement of droplets in the cloud turbulence.

### 6.1.2 Cloud Droplet Visualization in Avizo

In Avizo, applying color map/transfer function is the part where the color is applied to the visualization. There are color maps loaded in Avizo which can be chosen appropriately for the data in use. These color maps can also be customised in the Color map tab as per the requirements and then saved for future use. If a color map is not specified by the user, Avizo assigns a suitable color map by default. In Cloud Droplet visualization, a gray scale color map available in Avizo by default was used for visualization of mixing ratio variable which represented the cloud slab. On the other hand, the 'Physics.icol' color map is a 'cool to warm' color map.However,for droplet radius a 'warm to cool' color map is desired. The 'Physics.icol' color map was thus customized to give the effect of 'warm to cool' to droplet radius variable. This customized color map was loaded into Avizo separately and then mapped with the droplet radius variable. The minimum-maximum range for both variables was calculated over all the time steps and then the calculated minimum-maximum ranges were used for both the color maps accordingly.


Figure 6.6: Color maps/Transfer functions

Shown in figure 6.6, is the transfer functions/color maps for mixing ratio and droplet radii respectively. For the scale of the color bar, the minimum and maximum values for both the variables were calculated from the data and those values were kept constant throughout the visualization process for all time steps. For mixing-ratio variable, Gray scale color bar was used while for droplet radius variable, Physics color bar was used which is based on 'warm to cool' scaling and is suitable for tracking droplet movements.

Figure 6.7: Visualization of Mixing-ratio variable at time step 1200

Figure 6.7 represents the Eulerian grid data colored by mixing ratio variable at time step 1200. The volume rendering filter was applied to the grid data as it is suitable for visualization of liquid and gaseous phenomena. Volume Rendering can only be applied to data with uniform or rectilinear grids. In Volume Rendering method, the physical values for all grid cells are mapped to opacity and color values by using a transfer function which for the above visualization is defined by Gray scale color map. The alpha scale is set at 0.25 to obtain an optimal visualization.

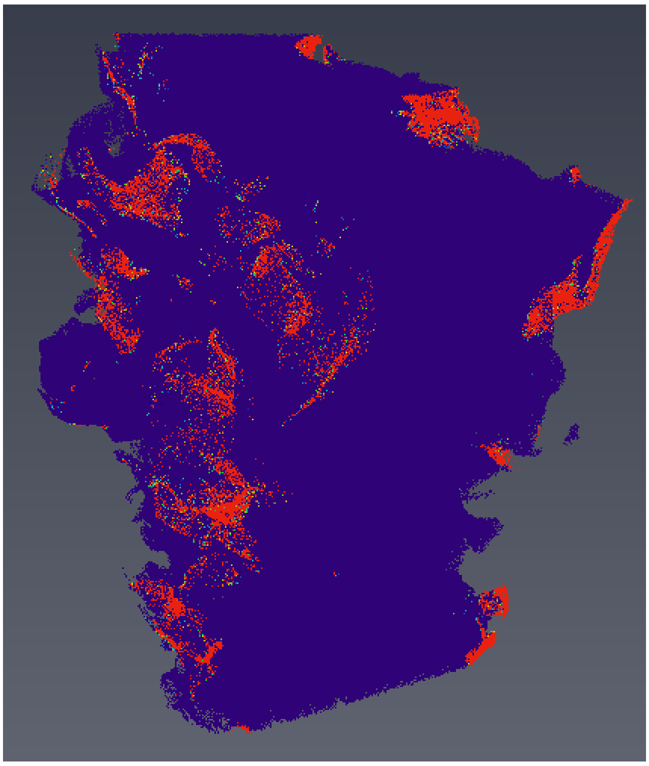

Figure 6.8: Visualization of Droplet radius variable at time step 1200

Figure 6.8 represents the Lagrangian droplet movement data colored by radius variable. Point Cloud filter was applied to this data. This filter filters a connected Point Cloud by an expression built from the coordinates and data values of each Point Cloud. It is mapped to Physics color map suitable to show droplet movement. Red color depicts droplets having smaller radii while blue color depicts droplets having greater radii.

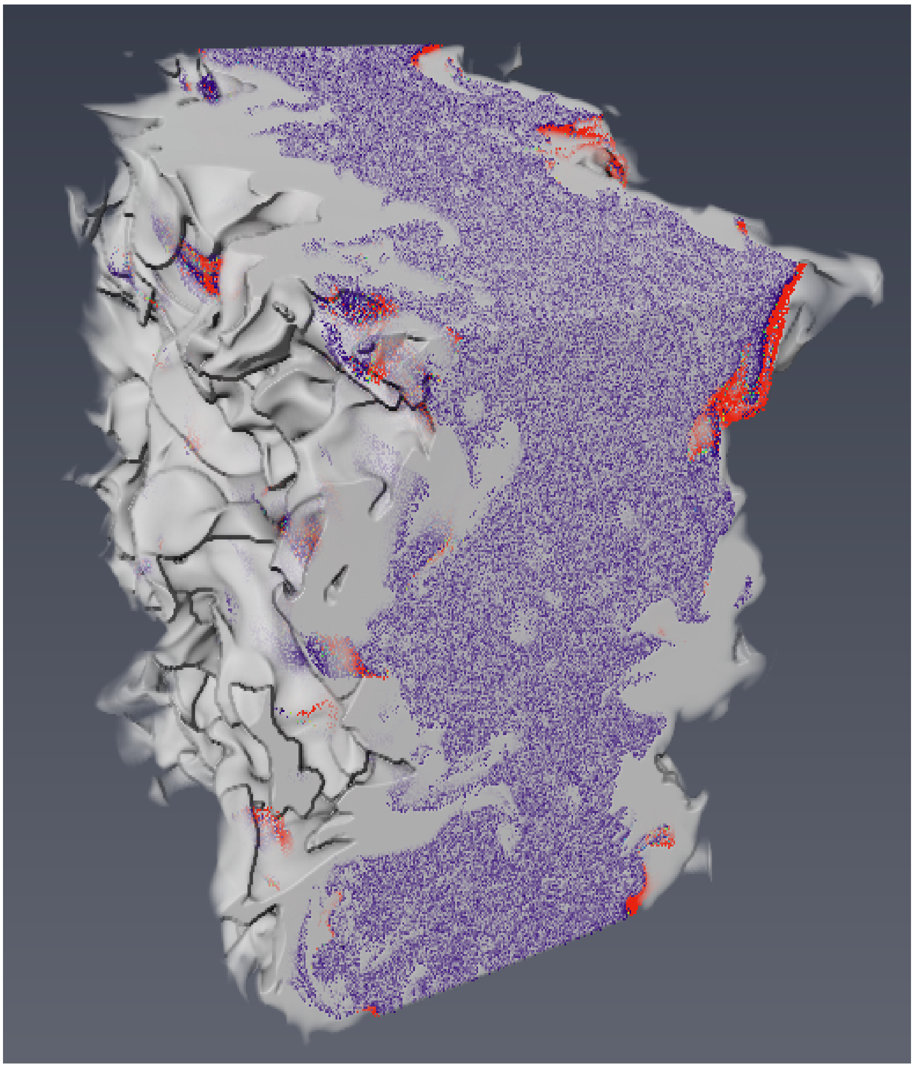

Figure 6.9: Visualization of Coupled Eulerian and Lagrangian files at time step 1200

Figure 6.9 represents the cloud droplet dynamics at time step 1200. The Eulerian gridded data and the Lagrangian data of time step 1200 are coupled together. The droplets are visible in the cloud slab. From the visualization it is clearly seen that at time step 1200, some droplets have started to evaporate(droplets in red) but the number of such droplets are still less as compared to total number of droplets.

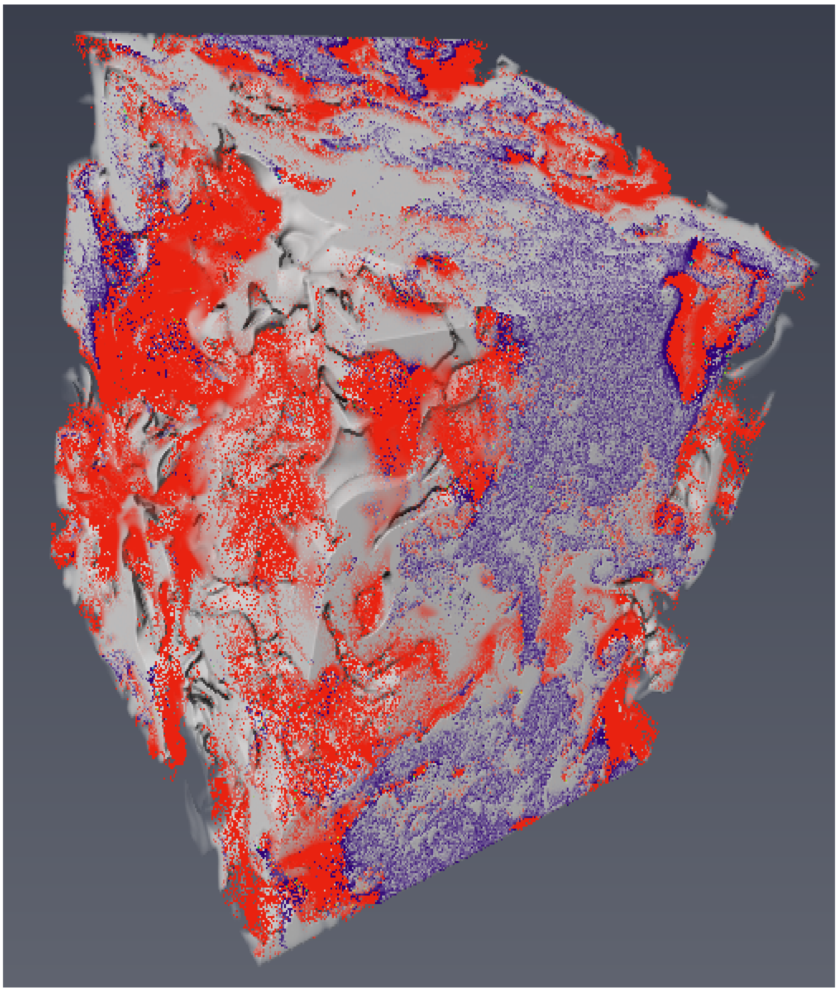

Figure 6.10: Visualization of Coupled Eulerian and Lagrangian files at time step 3500

Figure 6.10 represents the cloud droplet dynamics at time step 3500. The Eulerian and Lagrangian data at time step 3500 are coupled together. Here, the number of droplets which have undergone evaporation have increased as seen after comparing with the visualization obtained at time step 1200. The mixing process of droplets in cloud can be seen clearly by comparing the patterns of cloud and droplets in both the above visualizations. These changes occur due to evaporation and condensation of droplets in cloud stab.

Visualizations of such coupled Eulerian mixing ratio data and Lagrangian droplet data at every time step are created. The movement of droplets in cloud turbulence can be better understood by creating an animation of all the above visualized outputs.

# Chapter 7

# Conclusion And Future Work

In meteorological data analysis, visualization has emerged as a great tool to understand the data more quickly and efficiently. The results are communicated between climate researchers and scientific community. A picture is worth more than a thousand lines of numbers or words. This visualization of droplet evolution provides the opportunity to peek inside the cloud and observe properties of individual droplets. This investigation is not possible with field observations or with traditional data analysis.

This particular version of the visualization was run on an individual PC using CPU processing only. Parallel computational methods such as, GPU and multi-core processing, and those described in ParaView Visualization handbook[13] can be utilized in future studies to reduce time consumption. Due to time constraint the ASCII files were converted to VTK in this work as 'pyevtk' makes it easy to do so. In future, research can be done on how to read ASCII files directly in ParaView to avoid the step of converting them to VTK files as shown in Figure 5.1 on page no. 17.

# References


[1] Bipin Kumar, Jörg Schumacher, and Raymond A Shaw. Lagrangian mixing dynamics at the cloudy–clear air interface. *Journal of the Atmospheric Sciences*, 71(7):2564–2580, 2014.

[2] Marc Rautenhaus, Michael Böttinger, Stephan Siemen, Robert Hoffman, Robert M Kirby, Mahsa Mirzargar, Niklas Röber, and Rüdiger Westermann. Visualization in meteorology—a survey of techniques and tools for data analysis tasks. *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3268–3296, 2017.

[3] Figure 2.1 basic structure of netcdf file. https://simulatingcomplexity.wordpress.com/2016/09/21/.

[4] Russ Rew and Glenn Davis. Netcdf: an interface for scientific data access. *IEEE computer graphics and applications*, 10(4):76–82, 1990.

[5] Figure 2.2 data model of netcdf file. https://www.unidata.ucar.edu/software/netcdf/papers/nc4_conventions.html.

[6] netcdf4. https://unidata.github.io/netcdf4-python/netCDF4/index.html.

[7] xarray. http://xarray.pydata.org/en/stable/.

[8] Grib file format. https://en.wikipedia.org/wiki/GRIB.

[9] wradlib. https://docs.wradlib.org/en/stable/reference.html.

[10] Vtk file format user's guide. https://www.kitware.com/products/books/VTKUsersGuide.pdf.

[11] Sionlib. http://www.fzjuelich.de/ias/jsc/EN/Expertise/Support/Software/SIONlib/node.html.

[12] M Böttinger, K Meier-Fleischer, and C Ulmen. Tutorial: Interactive 3d visualization in earth system research with avizo green 8.0. *DKRZ/KlimaCampus Hamburg*, 2013.

[13] James Ahrens, Berk Geveci, and Charles Law. Paraview: An end-user tool for large data visualization. *The visualization handbook*, 717, 2005.

# Appendices

# Appendix A

# Plagiarism Report

F_13_AB_Report

ORIGINALITY REPORT

| 11% | 6% | 6% | 9% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Ramrao Adik Institute of Technology<br>Student Paper | 3% |
| 2 | www.aiktcdspace.org:8080<br>Internet Source | 2% |
| 3 | Sachin Bende, Rajashree Shedge. "Dealing with Small Files Problem in Hadoop Distributed File System", Procedia Computer Science, 2016<br>Publication | 1% |
| 4 | Submitted to K. J. Somaiya College of Engineering Vidyavihar, Mumbai<br>Student Paper | 1% |
| 5 | www.slideshare.net<br>Internet Source | 1% |
| 6 | www.ijser.org<br>Internet Source | 1% |
| 7 | www.clivar2004.org<br>Internet Source | <1% |

# Appendix B

# Paper Publication

Shubham Belgaonkar, Bhushan Bamble, Dr. Amit Barve, Dr. Bipin Kumar published ”Visualization of Droplet Dynamics in Cloud” in ITM Web Conf. Volume 32, 2020 International Conference on Automation, Computing and Communication 2020 Article Number-01007, Section-Automation, Published online: 29 July 2020
DOI: https://doi.org/10.1051/itmconf/20203203033

# Visualization of Droplet Dynamics in Cloud

*Shubham* Belgaonkar[1, *], *Bhushan* Bamble[2, **], *Amit* Barve[3, ***], *and Bipin* Kumar[4,****]

[1,2,3] Computer Department, Ramrao Adik Institute of Technology Navi Mumbai, India
[4] HPCS, Indian Institute of Tropical Meteorology, Pune, India

**Abstract.** Data visualization uses charts, graphs and maps to illustrate some information. Visualizing data helps us understand information faster. The study of droplet dynamics is a critical part of cloud physics and includes studying droplet properties. The aim of this work is to visualize the droplet dynamics obtained from DNS (Direct Numerical Simulation) data due to evaporation and condensation of the droplets. This simulation contains coupled Eulerian and Lagrangian frames. Animation is created for both Eulerian grid data and Lagrangian droplet movement. Scientific visualization provides a way to analyze these turbulent properties in a part of a cloud and learn about the nature of droplets and mixing process in such highly turbulent areas.

***Keywords—DNS (Direct Numerical Simulation), NetCDF (Network Common Data Form), VTK (Visualization Toolkit), HPC (High Performance Computing)***

## 1 Introduction

Weather forecasting centers manage a large amount of meteorological data. Such data must be interpreted within short period of time. Scientific visualization can be a great help in supporting meteorological researchers and scientists. Decision-making depends on data that comes in a large quantity that it cannot be understood without some level of abstraction, such as a visual one. There are various types of meteorological data namely rainfall data, temperature data, radiation data, radar data and many more. This paper involves examining various droplets' properties. It is difficult to examine the properties by just looking at DNS data. Small scale simulation is one method in understanding large data quickly. The turbulent entrainment and mixing of dry air with moist or cloudy air influences the cloud's lifespan by changing the properties. Such modifications are responsible for the droplet dynamics. This mixing process is quite complex, and it is difficult to study it during field observations. However, it is not clear how the size of the droplets is influenced by the mixing process and has a major impact on the spectrum of the droplets. This illustrates how all of the droplets will evaporate by the same number, or the subset of the droplets evaporates entirely, while the other droplets remain intact in the cloud turbulence. The types of mixing processes of the droplets are known as homogeneous mixing and in-homogeneous mixing [1]. This work uses ParaView for visualization because of its ability to analyze extremely large datasets rather than using expensive commercial software. ParaView is an open-source application for the analysis of data and visualization of the same. It can be used on supercomputers to process large data and also on laptops for smaller data. It can quickly build visualization pipelines for the analysis of data using both qualitative and quantitative techniques.

## 2 Literature Survey

The data formats along with the tools involved in the visualization are given below.

### 2.1. NetCDF File Format

NetCDF is the popular data format defined by Unidata program at the University Corporation for Atmospheric Research (UCAR), used widely in climatic, meteorological and oceanographic applications such as climate change and forecasting of weather etc. It is a data format that is created to support the array-oriented data [2]. All NetCDF formats contain a header which describing the layout of the rest of the file, particularly the data arrays. The NetCDF model of data abstraction creates scientific data set as a group of named multidimensional variables along with their coordinate systems [3]. Every NetCDF file has three components: dimensions which are used to define shape of data, variables to store the array values, and attributes to

[*] e-mail: shubhamb756@gmail.com
[**] e-mail: bhushan.bamble@gmail.com
[***] e-mail: barve.amit@gmail.com
[****] e-mail: bipink@tropmet.res.in

provide a title or this history of the data set. They help understand the data more clearly.

### 2.2 VTK (Visualization Toolkit)

The Visualization Toolkit offers numerous writer object with the ability to read and write standard file formats along with their own file formats. The purpose for creating other data file format is to provide a consistent data representation scheme for a number of dataset forms and to provide an easy method for transmitting data [4]. There are two file formats available in VTK, Legacy and Serial. These formats are less flexible as compared to XML file formats. While these formats are much more complicated than the Legacy format. By just providing a simple XML description any VTK source or filter can be added.

### 2.3 ParaView

ParaView is one of the most preferred open-source scientific visualization tools for the visualization. This software is used in many communities because of its easy to use nature to analyze and visualize large data sets. ParaView uses three steps to visualizing data: reading or loading the data files, filtering the necessary details, and then rendering the images for the given data. Reading the data involves loading the data files which creates pipelines for the loaded data, filtering the data involves creating arrays or using arrays which are necessary for the visualization and rendering involves generating and saving the images for the data set. ParaView has Python, C, C++ and Fortran bindings. In this work python shell is used for the visualization process.

## 3 Problem Statement

The aim of this paper is to propose a procedure for the visualization of the droplet dynamics in the cloud obtained from the DNS (Direct Numerical Simulation) data in an open-source software application, rather than expensive commercial alternatives to help the scientists and researchers to examine the properties of the droplets in cloud turbulence.

## 4 Data Flow Diagram

The input data consists of two datasets - Eulerian NetCDF files and Lagrangian SION files. As far as NetCDF files are concerned, there is an inbuilt NetCDF reader in ParaView software. For SION files, there are no readers as such in ParaView. So, to load SION data, SION files have to be converted into VTK format which is the default format for ParaView. Then it is to be further converted into vtkUnstructuredGrid format. The dataset forms a topologically irregular set of points and cells.
After the conversion to VTK, we extract the mixing ratio variable from the NetCDF files and create new series of NetCDF files which are smaller in size. This allows ParaView to process the files faster. Now, both the VTK and extracted NetCDF files and loaded in ParaView and visualization of the data is obtained at various time steps. The purpose of ParaView is to analyze extremely large data sets. Now these images are used for the animation purpose.

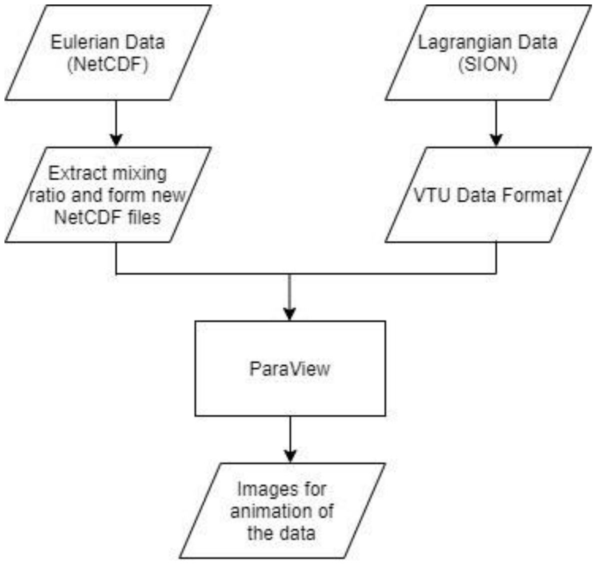


**Fig. 1.** Data flow Diagram

## 5 System Design

The DNS data stored in SION format could be converted to ASCII files using SIONLib [6]. ASCII files are also referred as text files that can be read easily. They often contain letters, numbers and some line separators such as comma or spaces. ASCII files may be larger in file size but the speed of writing and reading is much faster. They are efficient, but one has to specify the file's content before reading or writing the files.
So, to open and visualize the data a program must be used. There are many bindings available in all the popular languages to access these text files.
The procedure for the visualization of droplets dynamics includes extraction of the positions of the marker and radius of the droplets from the ASCII files, then preparation of data using Python script which will create series of VTK files that can be coupled with the NetCDF Eulerian grid data files and imported into open-source scientific visualization software - ParaView.
ParaView uses three steps to visualizing data: reading or loading the data files, filtering the necessary details, and then rendering the images for the given data. Data must be read in an acceptable file format into ParaView. Next, the data must be analyzed in different ways available and is to be filtered with the number of filters available. Finally, the visible image from the data is rendered and the images at different time steps can be stored. Now, these images are used to form an animation video using python script.

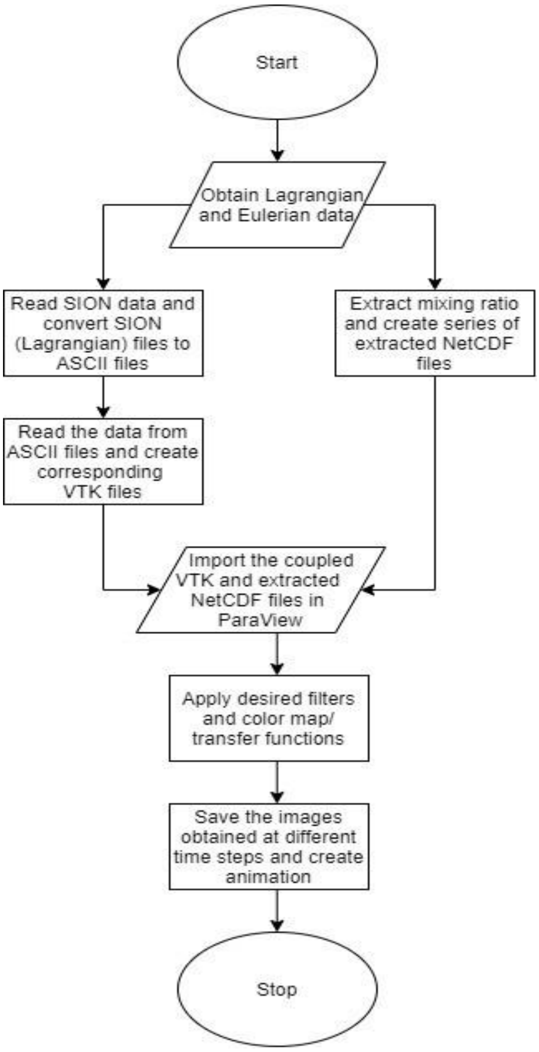


**Fig. 2.** Flowchart

## 6 Implementation

Extract the mixing ratio variable from the Eulerian data files and create series of NetCDF files containing only mixing ratio data as ParaView can process smaller files faster. Get the Lagrangian Data files converted to ASCII text files by extracting the marker positions and radius of the droplets. Load these coupled files in ParaView by creating NetCDF Reader for extracted Eulerian data and XMLUnstructuredGrid Reader for Lagrangian data. Set scalar coloring for both and create render view with best possible resolution. Results for this paper have 2555 x 1376 pixels. Change representation type apply color maps/ transfer functions until desired visualization is obtained. Edit color legend/bar for the datasets respectively and choose the scale based on minimum and maximum values in data. Save the Screenshots of the desired visualization and run above steps in loop for each extracted Eulerian file and Lagrangian file. Use python program to animate the saved images to create an animation video.

## 7 Results/ Screenshots

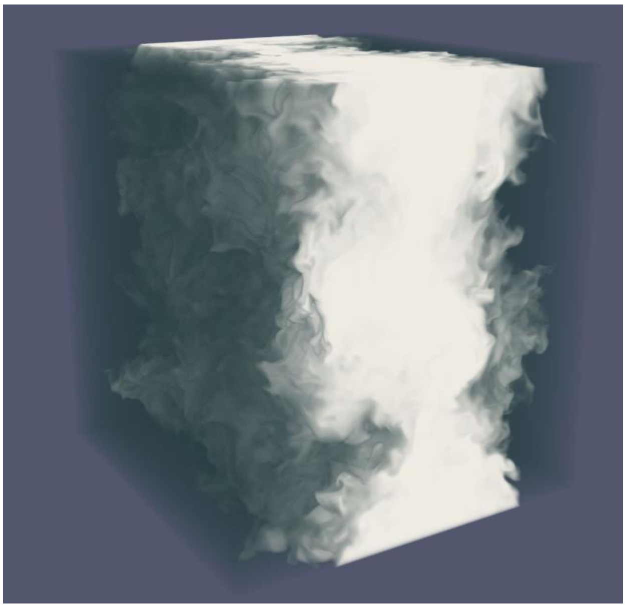

**Fig. 3.** Visualization of Mixing-Ratio of cloud at time step 20

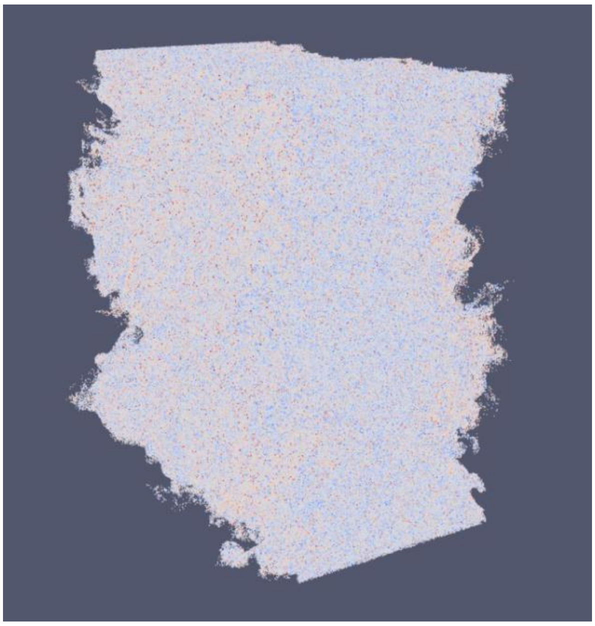

**Fig. 4.** Visualization of droplets based on the radius size at time step 20

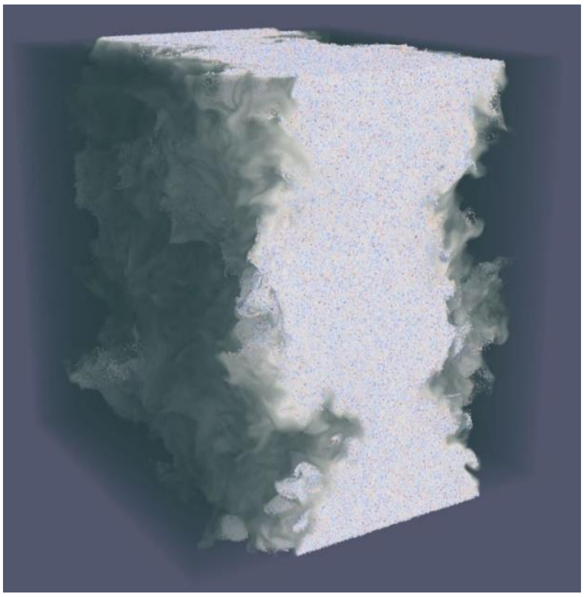

**Fig. 5.** Visualization of Coupled Eulerian and Lagrangian files at time step 20

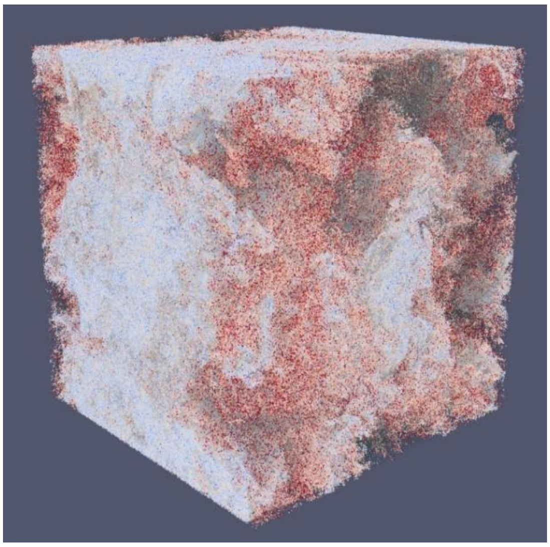

**Fig. 6.** Visualization of the Coupled Eulerian and Lagrangian files at time step 80

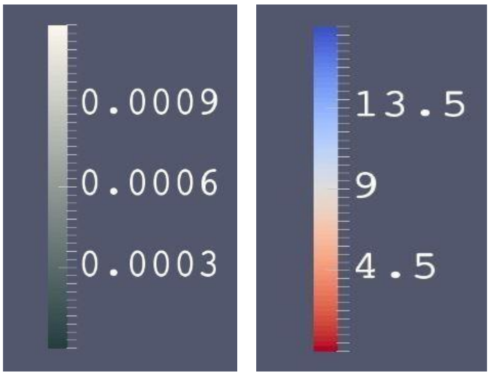


**Fig. 7.** Color maps/Transfer functions for mixing ratio and droplet radius respectively, Scale depends on the maximum and minimum values.

Shown in fig. 3, is the visualization of Eulerian grid data at time step 20 colored by Mixing-Ratio variable. Central cloud slab is visible in white. Fig 4 shows the visualization of Lagrangian data at time step 20 colored by radius variable. It shows the visualization of droplets based on their radius size. Blue droplets have larger radii and the droplets in red are with the smallest radii. Shown in fig. 5, is the visualization of coupled Eulerian grid data and Lagrangian data at time step 20. Both of these data files are imported simultaneously and visualized together. The droplets of different radii are visible in white cloud slab. Shown in fig. 6, is the visualization of Eulerian grid data and Lagrangian data together at time step 80. The droplets have started to evaporate and get smaller in size (red particles). Thus, the mixing process of the droplets in the cloud turbulence can be observed. In fig. 7, the color bar on the left represents the color of the cloud depending on the data and the color baron the right represents the droplet movement in the cloud.

To summarize, fig. 3 to fig. 6 represent the visualizations at different time steps using the two-color bars to analyze the droplet movement. The blue droplets are the ones with larger radius which evaporate in the cloud turbulence and become smaller in size as seen in the final step (red particles) as shown in fig 6. These results show the mixing process of the droplets in cloud due to evaporation and condensation.

## 7 Conclusion

In meteorological data analysis, visualization has emerged as a great tool to understand the data more quickly and efficiently. The results are communicated between climate researchers and scientific community. A picture is worth more than a thousand lines of numbers or words.
This visualization of droplet evolution provides the opportunity to peek inside the cloud and observe properties of individual droplets. This investigation is not possible with field observations or with traditional data analysis.

## Acknowledgement

The simulation has been carried out on a HPC facility provided by Ministry of Earth Science, Government of India at Indian Institute of Tropical Meteorology, Pune. A sincere thanks to the Institute and the scientists who contributed in this project and let us use the facility at the institute. Also, thanks to Matt Rehme, software engineer at National Centre of Atmospheric Research, who helped with his valuable inputs in the visualization with ParaView.

## References


[1] B. Kumar, J. Schumacher, R. A. Shaw, Lagrangian mixing dynamics at the cloudy-clear air interface, Journal of Atmospheric Sciences 71 (7) (2014)

[2] Marc Rautenhaus, Michael Bottinger, Stephan Siemen, Robert Hoffman, Robert M Kirby, Mahsa Mirzargar, Niklas Rober, and Rudiger Westermann: Visualization in meteorology—A Survey of Techniques and Tools for Data Analysis Tasks (2017)

[3] Russ Rew and Glenn Davis: NetCDF - An Interface for Scientific Data, (1990)

[4] VTK file format: URLhttps://www.kitware.com/products/books/VTKUsersGuide.pdf

[5] S.G. Parker, D.M. Weinstein, and C.R. Johnson, “The SCIRun Computational Steering Software System,” Modern Software Tools in Scientific Computing, E. Arge, A.M. Brauset, and H.P. Langtangen, pg. 5-44, (1997)

[6] Sionlib: URLhttp:/www.fzjuelich.de/ias/jsc/EN/Expertise/Support/Software/SIONlib/node.html

# Acknowledgement

We take this opportunity to express our profound gratitude and deep regards to our guide **Dr. Amit Barve** for his exemplary guidance, monitoring and constant encouragement throughout the completion of this report. We are truly grateful to his efforts to improve our understanding towards various concepts and technical skills required in the project. The blessing, help and guidance given by his time to time shall carry us a long way in the journey of life on which we are about to embark.

We take this privilege to express our sincere thanks to **Dr. Mukesh D. Patil, Principal, RAIT** for providing the much necessary facilities. We are also thankful to **Dr. Leena Ragha**, Head of Department of Computer Engineering, Project Co-ordinator **Mrs. Smita Bharne** and Project Co-ordinator **Mrs. Bhavana Alte**, Department of Computer Engineering, RAIT, Nerul Navi Mumbai for their generous support.

We are also grateful to **Indian Institute of Tropical Meteorology(IITM), Pune** who gave us an opportunity to perceive project internship at IITM under Datavisualization Domain and provide us necessary facilities. We are thankful to **Dr. Bipin Kumar** , **Dr. Rajib Chattopadhyay** and **Mr. Manmeet Singh** for the encouragement and support that they have extended. Their continuous guidance and direction for the project made this possible. Also, thanks to **Matt Rehme**, software engineer at National Centre of Atmospheric Research, USA, who helped with his valuable inputs in the visualization with ParaView.

Last but not the least we would also like to thank all those who have directly or indirectly helped us in completion of this thesis.

**Mr. Bhushan Bamble**

**Mr. Shubham Belgaonkar**