# Laser Scanning and Photogrammetry

# CE – 676

## PROJECT REPORT

Name: Manoneet Anil Gawali

Roll no.: 231030035

Email: manoneetag23@iitk.ac.in

Course Instructor

Dr. Bharat Lohani

Civil Engineering Department
Geoinformatics
IIT kanpur

# 1 Objective

To assess the quality of aerial LiDAR data of IIT Kanpur, the following quality parameters must be included:

• LiDAR returns- How many returns are there in the data?

• LiDAR overlap- What is the maximum, minimum, and average overlap in the two flight lines?

• Altimetric (vertical) accuracy – What is absolute vertical accuracy?

• Planimetric (horizontal) accuracy – What is absolute planimetric accuracy?

• Relative Accuracy-What is relative accuracy within swath and in overlap area?

• NPS and data density-What are the values of NPS and data density?

• Data voids- Are there data voids that are not acceptable?

• Spatial distribution (clustering and uniformity)-Does data qualify as uniformly distributed in planimetry?

# 2 Theory

## 2.1 LiDAR returns:

The number of returns in LiDAR data can vary depending on the sensor and the data collection parameters. Typically, LiDAR systems can record multiple returns per pulse, including the first return (which hits the highest object), intermediate returns (hits on objects between the highest and ground level), and the last return (hits on the ground). This multiple return capability allows LiDAR data to capture information about vegetation, buildings, and terrain surfaces more accurately.

To determine the number of returns in your specific LiDAR dataset, you would need to inspect the data or consult the metadata provided with the dataset. The metadata should contain information about the number and type of returns recorded by the LiDAR sensor during data collection.

## 2.2 LiDAR overlap:

LiDAR overlap refers to the extent to which adjacent LiDAR flight lines or swaths overlap each other. Overlap is an important aspect of LiDAR data acquisition because it affects data quality and accuracy, particularly in areas with complex terrain or dense vegetation.

## 2.3 Altimetric (Vertical ) Accuracy:

Absolute vertical accuracy, also known as altimetric accuracy, refers to the measure of how closely LiDAR-derived elevation values or points represent the true elevation of the Earth's surface. It specifically focuses on the accuracy of elevation data in the vertical dimension, usually measured in units such as meters or feet. Absolute vertical accuracy assesses how well LiDAR data points align with the actual elevations of features on the ground or surface. It is essential for applications where precise elevation information is critical, such as floodplain mapping, terrain modeling, and infrastructure design.

### 2.4 Planimetric (horizontal) accuracy:

Planimetric accuracy, also known as horizontal accuracy, refers to the measure of how accurately geospatial data represents features on the Earth's surface in terms of their horizontal positions. Unlike vertical accuracy, which focuses on elevation differences, planimetric accuracy deals with the accuracy of geographic locations in the horizontal plane. Planimetric accuracy assesses the positional accuracy of geographic features in maps, GIS (Geographic Information System) data, satellite imagery, and other geospatial datasets. It refers to how closely the coordinates of features in these datasets match their actual locations on the Earth's surface.

### 2.5 Relative accuracy:

Relative accuracy within swath refers to the consistency and precision of LiDAR data points within a single flight line or swath. It assesses how closely the data points align with each other horizontally within the same swath. Relative accuracy within swath is evaluated by comparing the distances between adjacent LiDAR points within the swath. Ideally, points that represent the same object or feature should have consistent distances between them, indicating high relative accuracy within the swath.

Both relative accuracy within swath and in overlap area are critical aspects of LiDAR data quality assessment. They ensure that the horizontal positioning of data points is consistent and reliable, which is essential for various applications such as terrain modeling, vegetation analysis, flood mapping, and infrastructure planning. Evaluating and maintaining relative accuracy throughout the LiDAR dataset processing workflow is fundamental to achieving high-quality geospatial data products.

### 2.6 Data voids- Are there data voids that are not acceptable:

Yes, data voids in LiDAR data are generally considered unacceptable, especially in applications where complete and accurate coverage of the surveyed area is crucial. A data void refers to areas within a LiDAR dataset where no or insufficient data points are present, resulting in gaps or missing information in the geospatial representation of the terrain or objects.

Data voids lead to incomplete information about the surveyed area, which can hinder accurate analysis, modeling, and decision-making processes. Missing data can result in inaccurate terrain models, vegetation assessments, floodplain mapping, and other geospatial applications.

### 2.7 Spatial distribution (clustering and uniformity)

spatial distribution encompasses the arrangement and pattern of objects or data points in space, including clustering (grouping tendencies) and uniformity (evenness or regularity). Analyzing spatial distribution patterns provides valuable insights for a wide range of disciplines and applications.

# 3 Methodology

1. Read the reference material and course notes provided. Use the methods learned in class to solve the problems in this project.
2. You are free to use the software of your choice, or you can also write code to realize the output.
3. Determine the horizontal and vertical datum employed in the data files and mention the same.
4. For some of the quality measures, you will have to go to the field with the GNSS instrument. Use appropriate GNSS method to collect field data and also ensure that the horizontal and vertical datum for field data are same as that of given LiDAR data.
5. The data files cover part of IIT Kanpur and also some areas out of IIT Kanpur. The fieldwork will be conducted only inside IIT Kanpur. The choice of the place where field work will be completed is to be decided by the group. GI Laboratory will provide you the necessary instruments. Inform them in advance, so batteries etc. are ready. It is better to complete the fieldwork on the days of the laboratory or Saturdays. Make your arrangements after discussion with TAs and Laboratory Staff.

## 3.1 Steps followed for No. of returns:

We find the number of returns by of LiDAR data by saving the point cloud data into .csv, this is done when we save the point cloud data file after importing to CloudCompare. When we open this csv file in Excel we can see that the columns of Easting, Northing, Height, Number of returns, Edge value, R,G,B, Intensity, etc. So, we try to find the unique number in Number of Return column by writing a short MATLAB code.

## 3.2 Steps followed for Non-Vegetated Accuracy (NVA):

For Finding NVA we first selected 3 data sites which can be seen in LiDAR data and can be surveyed and are convenient, So Basketball court, GH-1 lawns and Tennis court, As data is pretty old so we need to select the survey sites accordingly. Now as we know the Co-ordinates of the LiDAR points so we will then find the Co-ordinates of Ground Survey. Hence we did GNSS survey using RTK method and find the Co-ordinates and imported both files LiDAR point cloud and .csv file obtained from RTK survey.

Steps to take observations from Trimble R10 Receiver in RTK mode:

i. Place the GNSS receiver in a location on the jobsite where equal range in all directions provides full coverage of the site. This is more important on larger jobsites, where the broadcast range of the base station radio may limit the operations of the system.
ii. Place the GNSS antenna in a location that has a clear line of sight to the sky in all directions. Do not place the antenna near vertical obstructions such as buildings, deep cuttings, site vehicles, towers, or tree canopy.
iii. Place the GNSS and radio antennas as high as practical. This minimizes multipath from the surrounding area and enables the radio to broadcast to the maximum distance Importance of cut-off angle in GNSS Mission planning.
iv. Level the receiver with help of bubble level attached with the Tripod Instrument
v. Turn on the receiver and TSC3 or Trimble Tablet. Run Trimble Access.

vi. In the General Survey menu: Tap Instrument
vii. On the receiver by pressing power button
viii. Connect the TSC3 and R10 receiver and connect TSC3 to wi-fi.
ix. On TSC 3 Click on “New Project” and Give Name for Our project file.
x. Give Job file Name and Set UTM 44 North Zone.Set datum as WGS -84.
xi. Set Project height as 0.
xii. Set co-ordinates as Grid. And click on Store.
xiii. Now Go to Settings
xiv. Click on instrument > GNSS function > Connections > Bluetooth > connect to GNSS rover
xv. Now Go back and Select > Rover mode > click Accept.
xvi. Now again click on Menu> Settings > Survey style > select “RTK” and Go to Rover option.
xvii. In rover option > select “Survey styles” > set antenna height of 2m & Elevation mask 10 degrees.
xviii. In GNSS Signal tracking Options Select Satellites which we want data (GPS, GLONASS, Galileo, Bei Dou, etc) > click on “Store”
xix. Go to main menu > Settings bar and click > Measure > “ RTK ” > Measure points.
xx. Now In measure point, Give Point name, Code. Antenna height can be changes from here too.
xxi. Now select measure to start measuring data of Point. And Reverse timer will start that our data capturing is started.
xxii. After Capturing data > click on “Store” to store observations.
xxiii. Now we can capture other points too with same clicking on measure. No need to adjust settings again which we have setup for rover.

**The Outline we followed is:**

a. Identify and define the areas in your LiDAR dataset that are non-vegetated.
b. Obtain ground truth data for these non-vegetated areas. Here we used RTK GNSS survey.
c. Extract LiDAR data corresponding to the non-vegetated areas defined earlier.
d. Process the LiDAR data to obtain elevation values for these areas.
e. Compute the differences between the LiDAR-derived elevation values and the ground truth values.
f. Calculate accuracy metrics such as Root Mean Square Error (RMSE)
g. Use the accuracy metrics calculated above to determine the NVA for your LiDAR data.
h. Lower NVA values indicate higher accuracy in non-vegetated areas, while higher NVA values suggest potential issues or inaccuracies that may need attention.

1. Import the LiDAR point cloud and visualize the non-vegetated site, Basketball court of PE ground, IIT Kanpur can be visualized.

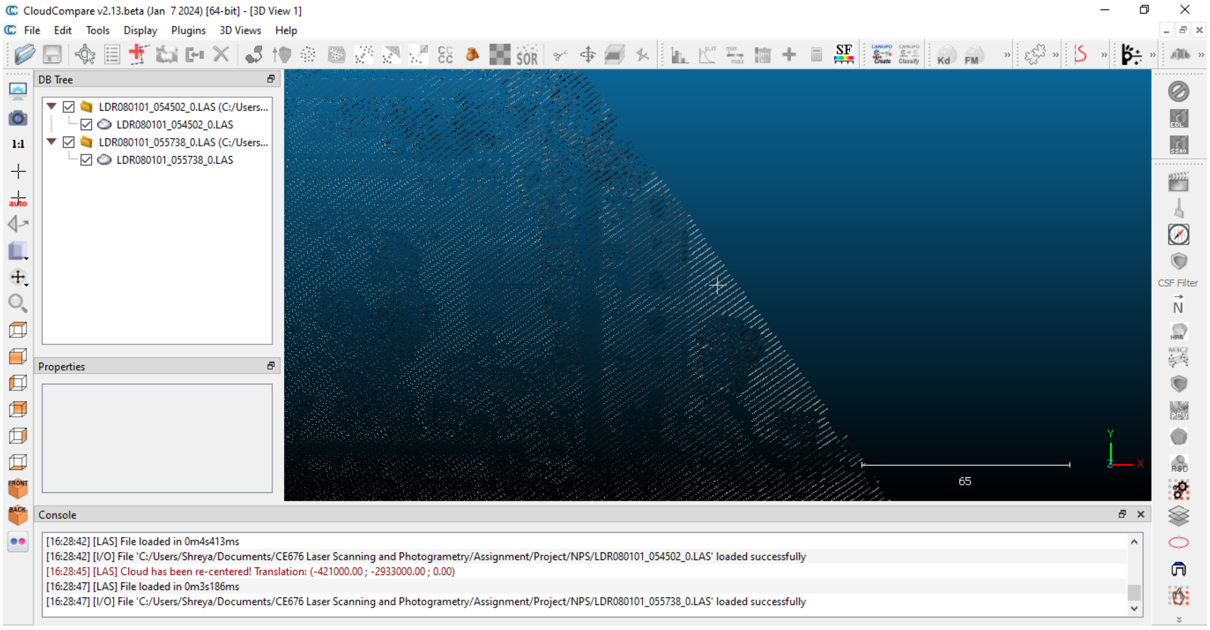



*Figure 1: Visualizing Imported Point cloud data, Basket ball court can be seen.*

2. Segment the Area of Interest

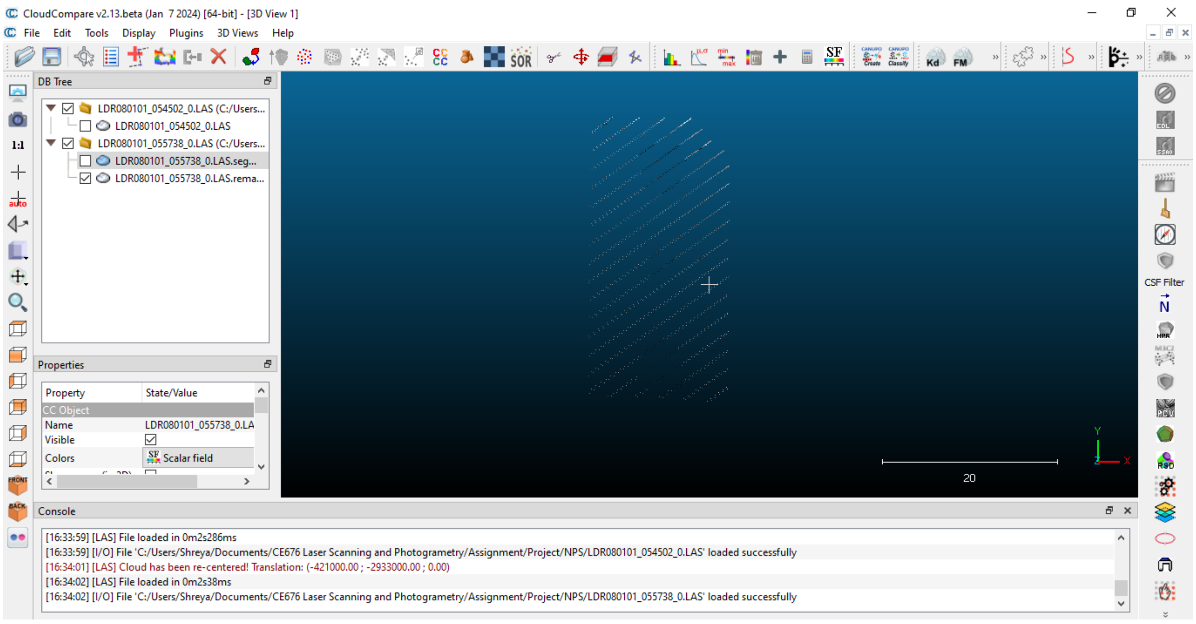



*Figure 2: Segment the area of Interest.*

3. Overlay the GCPs collected on the Point cloud data in Cloudcompare.

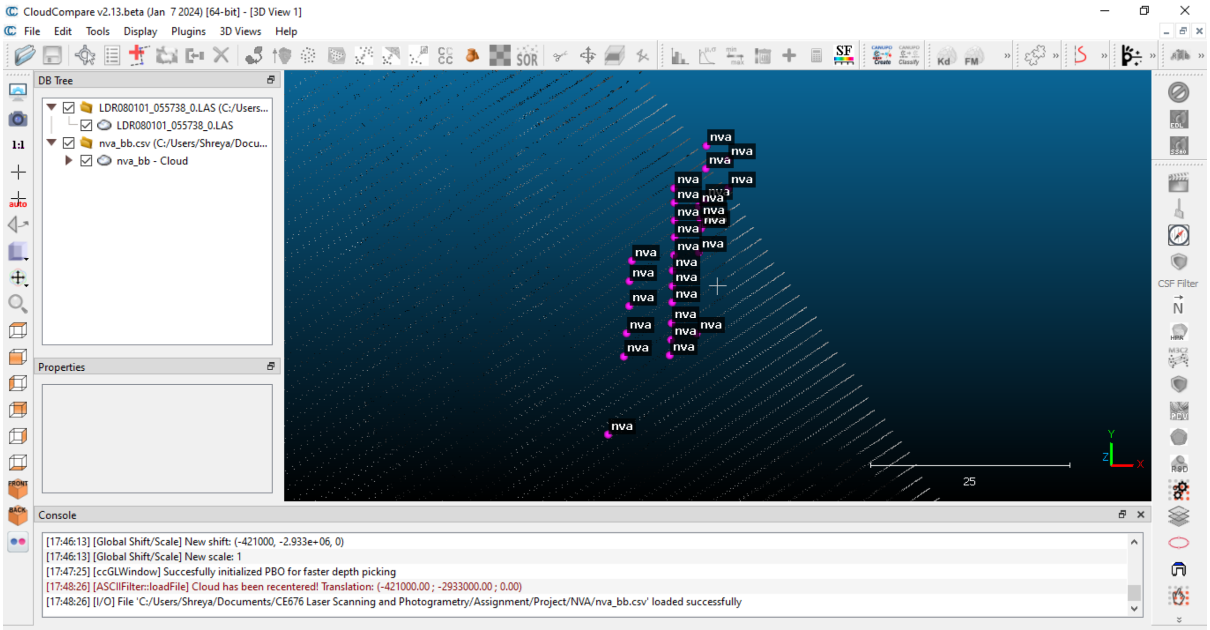
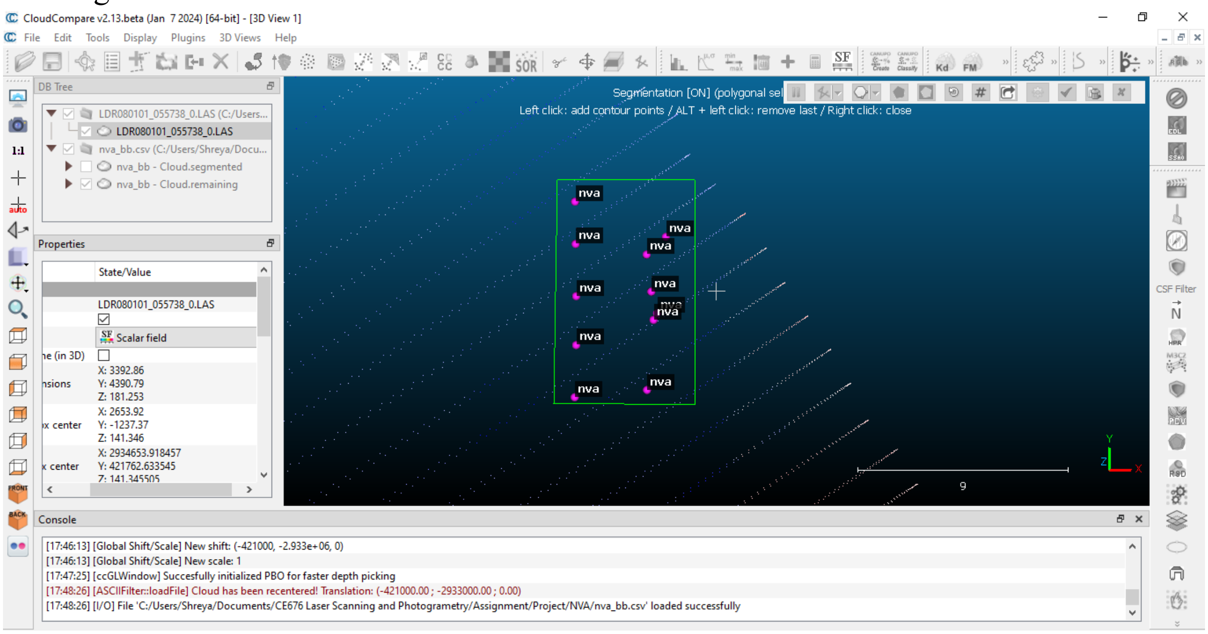


*Figure 3: Overlaying the GCPs on the Point Cloud Data.*

4. Segment the relevant GCP data.



5. Observe the Height differences over time.

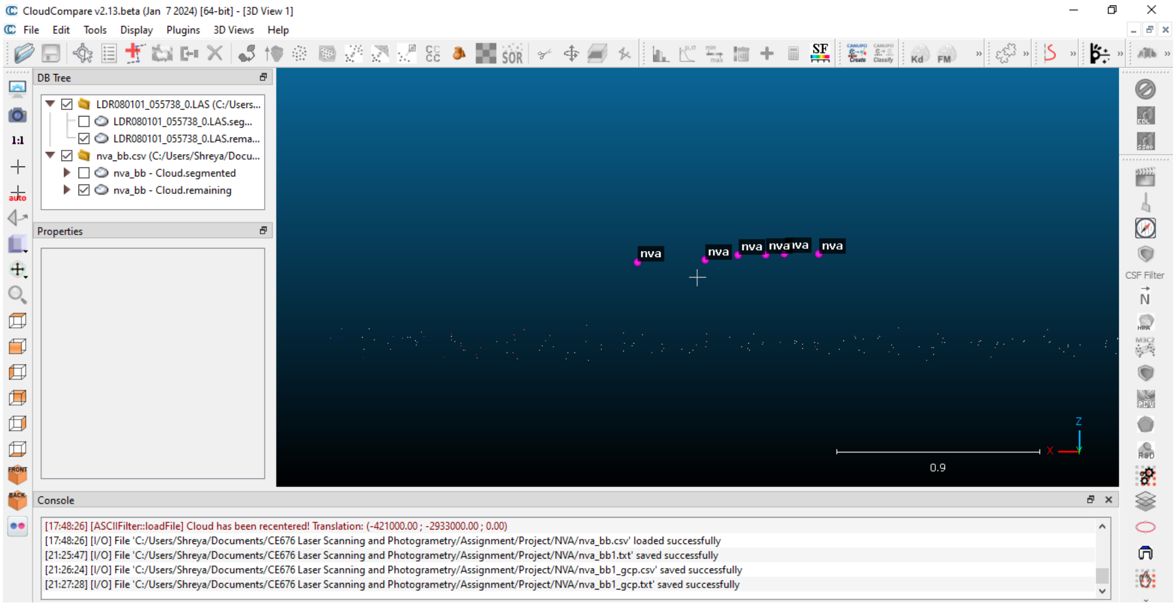
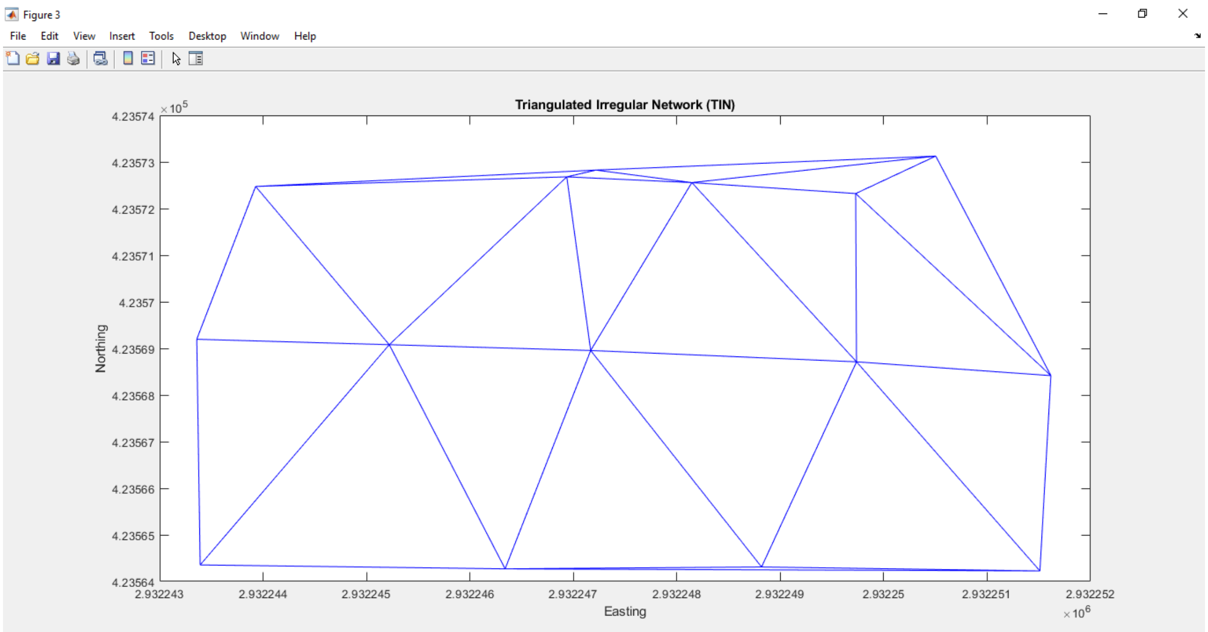



*Figure 4: Height Difference over time.*

6. Generated the TIN by TIN interpolation.



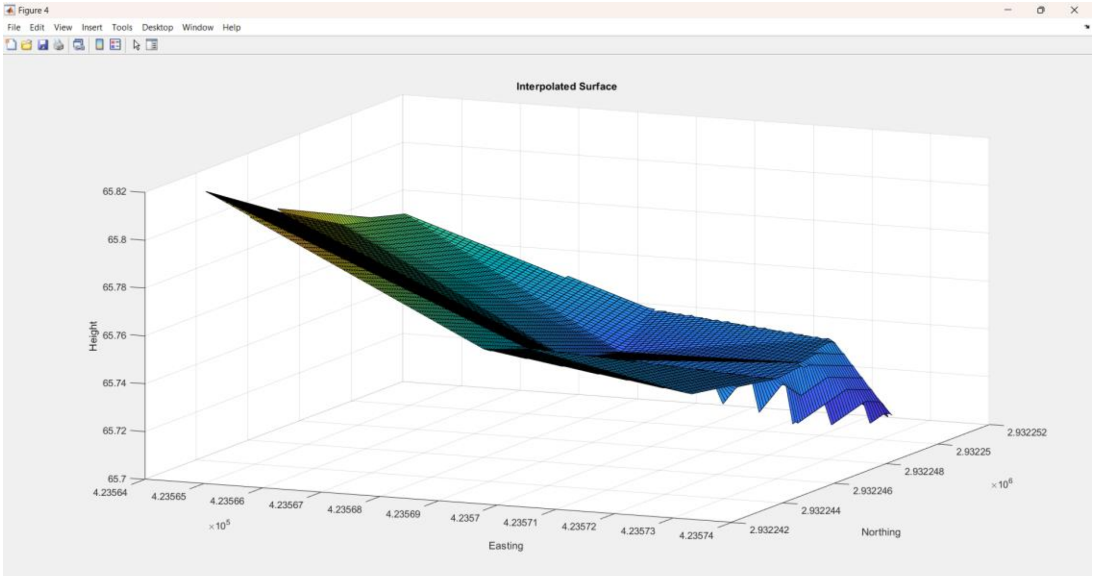



*Figure 5: Interpolated TIN generated by MATLAB.*

The surface interpolated shows idea about the range that points are in the values of 65.7 to 65.82.

7. Now from MATLAB we can also observe the Lidar point cloud points and GCPs representation.

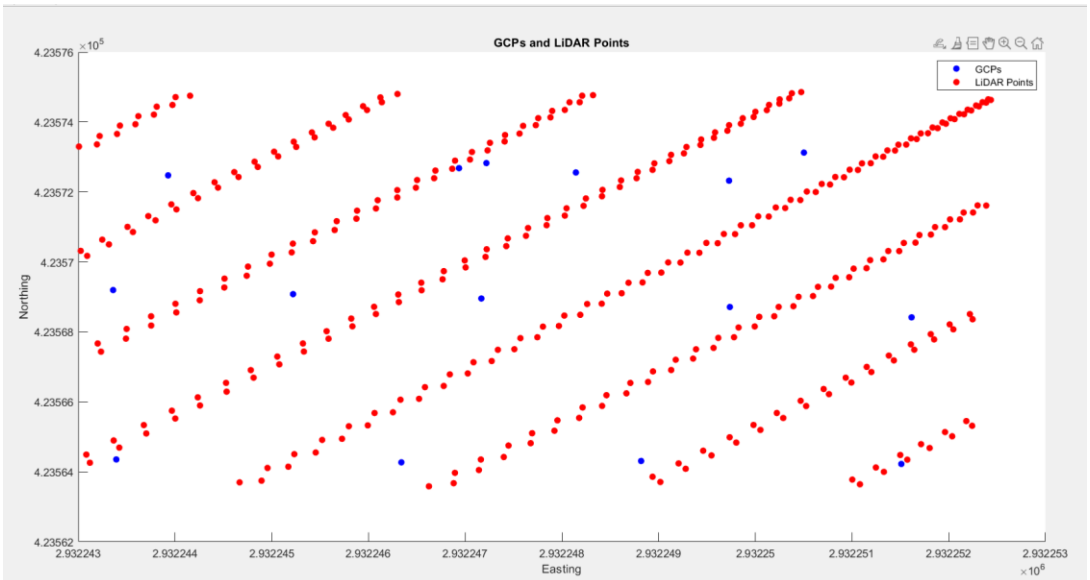



*Figure 6: Observing the lidar point cloud points and GCPS from MATLAB code.*

For height measurements difference between ground survey data and point cloud we generate a TIN interpolation and then calculate the RMSE value for this elevations. For this first we load the point cloud data and GCPs and Generate a TIN using GCPs co-ordinates of easting and Northing, Then using TIN we interpolate the heights, then we identify the Non-NaN height points from the corresponding data and calculate height difference of point cloud data and heights, and then we calculate the height difference and RMSE values.

Formula for RMSE calculation:

$$RMSE_{height} = \sqrt{\sum(Z_{Lidar}^2 - Z_{GCP}^2)/n}$$

Vertical accuracy for 95% confidence interval:

$$Vertical\ Accuracy_{height} = 1.96 * RMSE_{height}$$

Then all the other Patches will be processed and followed the same methodology for calculation of Vertical accuracy and RMSE.

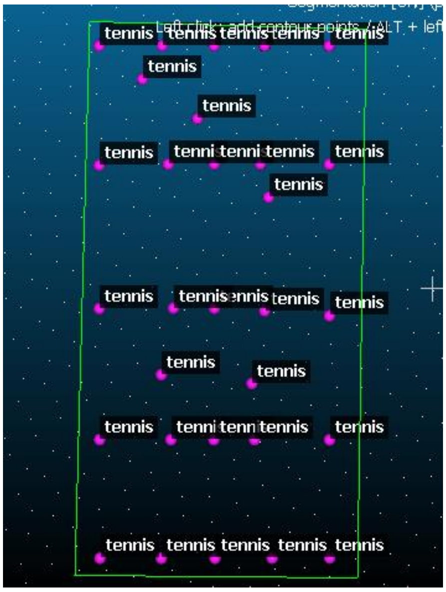


Figure 8: Overlay Point cloud and GCPs

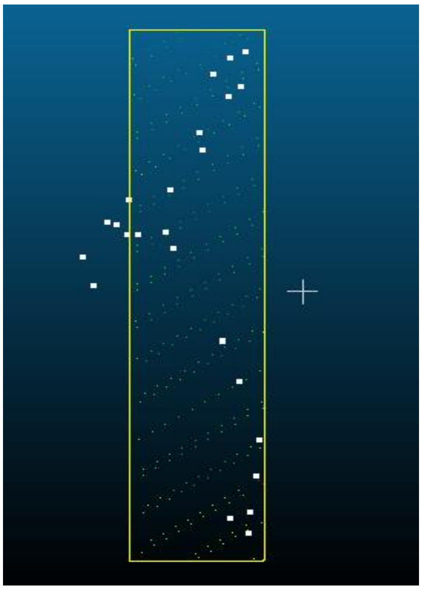


Figure 7: Tennis court area segmented data

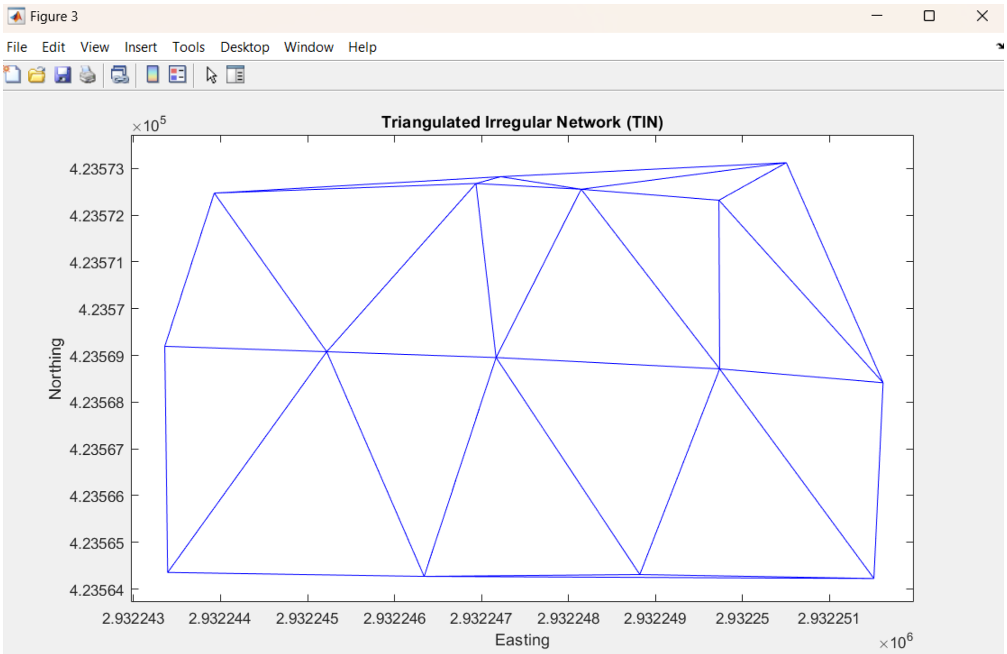



Figure 9: Generated TIN Network, visualised in MATLAB

## 3.3 Steps followed for Vegetated Vertical Accuracy:

Vegetated vertical accuracy is been calculated for vegetative area, basic idea being is that for the area covered in vegetation what will be accuracy range for lidar data having vegetation and the the points calculated from the RTK mode of survey at the same area of Interest. GCPs collected for the area were 30 points per data patch.

Accuracy of data generally would have been effected by the vegetation growth change due to years which includes change in height of trees, plants, grass, etc. We can generate DEM to accurately measure elevation of ground surface. Which leads to error propagation.

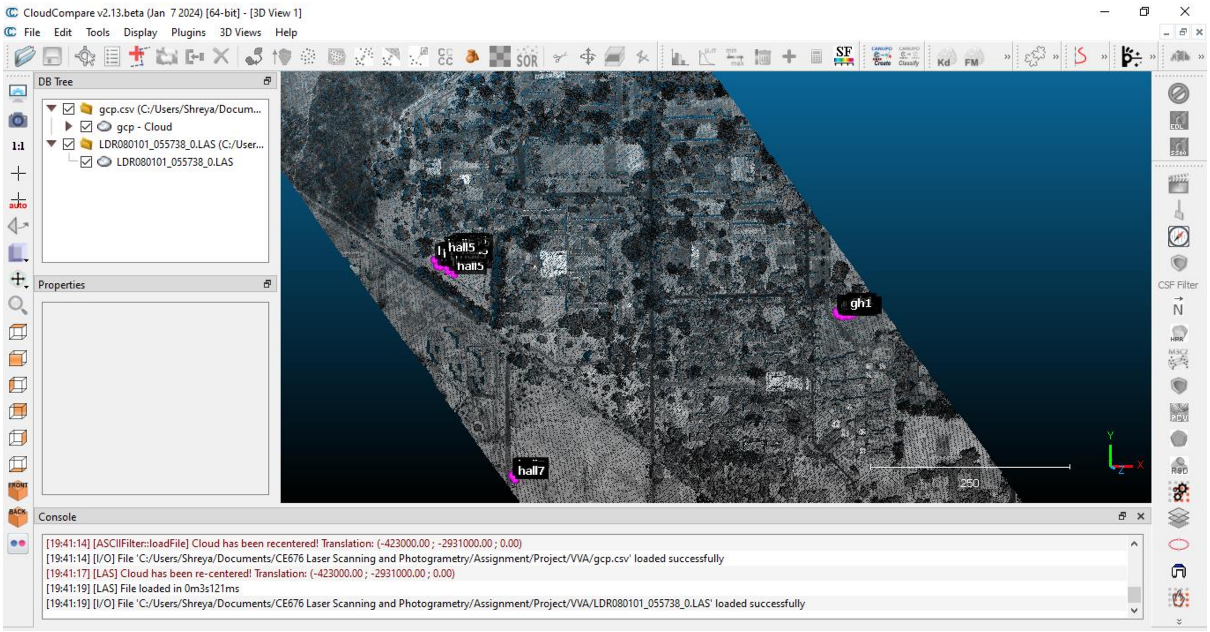



*Figure 10: Data overlayed in Cloudcompare.*

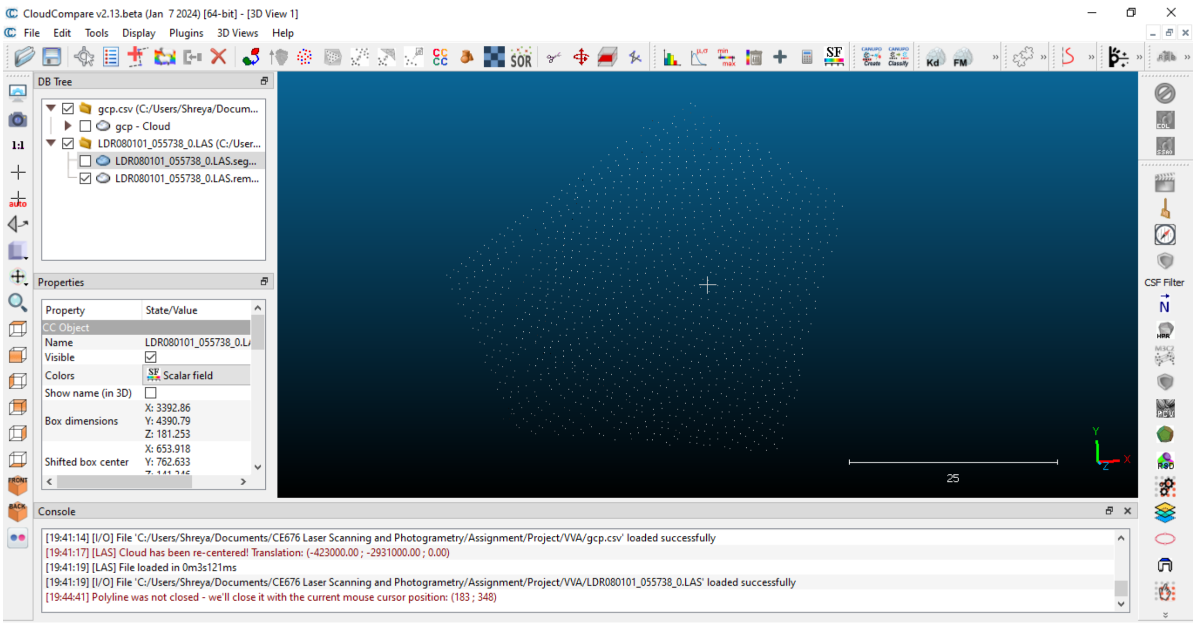



*Figure 11: Segmented data in Cloudcompare.*

1. reads elevation data from a CSV file.
2. Read GCPs, reads elevation data from another CSV file
3. Calculates the differences between the LiDAR elevation and the GCP elevation.
4. sorts the calculated differences in descending order. Sorting the differences allows for identifying the largest discrepancies between LiDAR and GCP elevations.
5. calculates the 95th percentile of the sorted differences. The 95th percentile represents the threshold below which 95% of the differences fall, providing a measure of the vertical accuracy (in this case, the discrepancies) at a specific confidence level.

$$RMSE_{height} = \frac{\sqrt{Z_{Lidar} - Z_{GCP}}}{n-1} = 3.605 \text{ at } 95\ \% \text{ Confidence interval}$$

## 3.4 Steps followed for Planimetric Accuracy:

Planimetric accuracy is a quality assessment for Lidar data accuracy, which refers to the horizontal accuracy of the LiDAR data points with respect to their real-world locations on the Earth's surface.

In airborne LiDAR data, the under-sampling characteristic refers to the fact that the spacing between LiDAR data points (point spacing) is typically larger than the laser beam's ground spot diameter. This can pose challenges when selecting and measuring ground control points (GCPs) for planimetric accuracy verification, as traditional GCP selection methods may not directly apply due to the sparse distribution of LiDAR points on the ground.

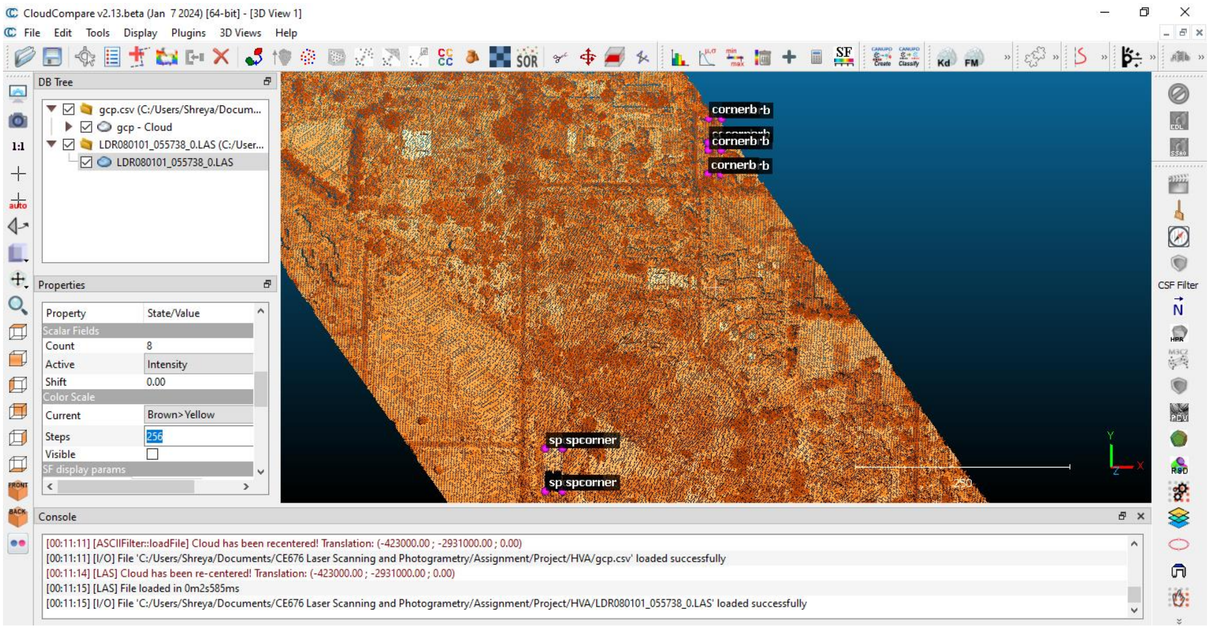



*Figure 12: Points overlayed on data.*

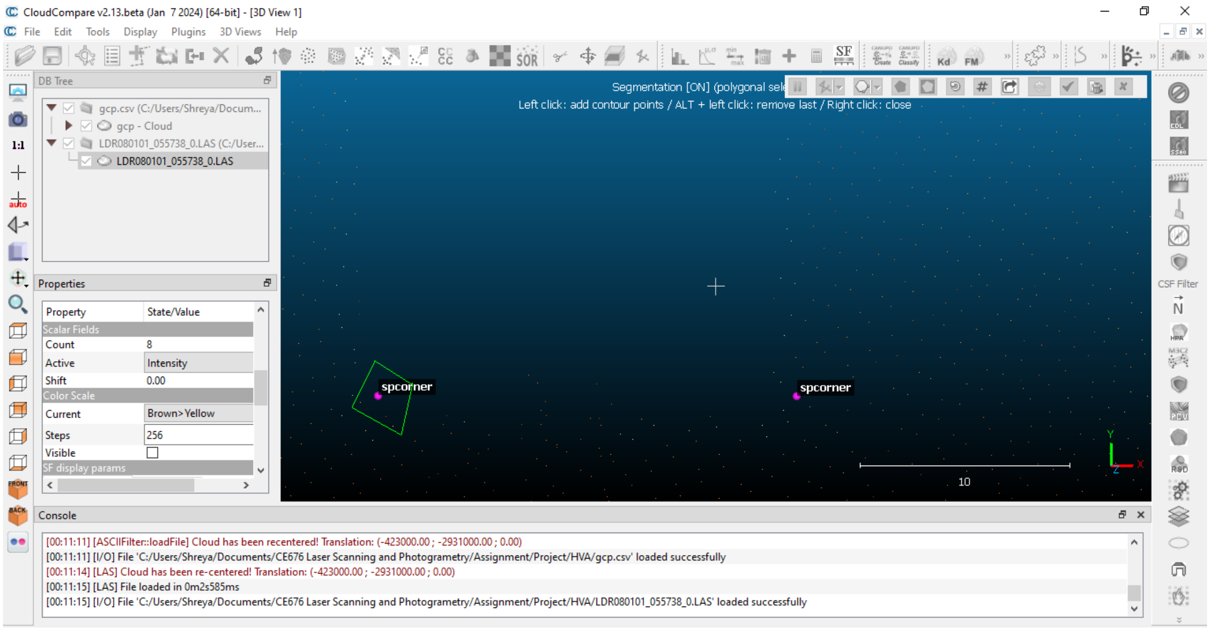



*Figure 13: Segmenting the Possible corners of data in CLoudCompare.*

for calculating the Root Mean Square Error (RMSE) and 95th percentile confidence intervals for planimetric accuracy verification between LiDAR data and Ground Control Points (GCPs) stored in CSV files:

1. Read the csv files of data and GCPs.
2. The northing and easting coordinates are extracted from both the LiDAR data and GCP data using indexing.
3. calculates the differences between corresponding northing and easting coordinates in the LiDAR and GCP datasets.
4. Mean squared errors are calculated by taking the mean of squared errors.
5. Root Mean Square Error (RMSE) is computed by taking the square root of the mean squared errors. RMSE provides a measure of the average magnitude of errors between LiDAR and GCP coordinates.
6. Assuming a normal distribution, the code estimates the 95th percentile confidence intervals for both northing and easting RMSE values using a confidence multiplier of 1.96
7. Calculate RMSE values for northing and easting, as well as the 95th percentile confidence intervals for northing and easting.

## 3.5 Steps followed for NPS and Data Density:

NPS (Nominal Point Spacing) in LiDAR refers to the average spacing between adjacent LiDAR data points in a point cloud. It represents the density of data points collected by the LiDAR system, typically measured in units such as meters or feet. The NPS is a crucial parameter in LiDAR data processing and analysis as it directly impacts the level of detail, accuracy, and information content in the generated point cloud.

1. Select the tets sites for plain ground so that we can get just a single return from the LiDAR data.
2. Now Segement the data of various sites.
3. Save the data in csv file format and Load the file for two flight lines.
4. Compute the distance between the Adjusent points for point cloud.
5. Now we need to calculate the average distance between adjusent points using moving average.
6. Now we need to arrange distances in descending for observing the maximum distance among the data arranged.
7. Now we compute the NPS by determining thethreshold for 5% of averages of Maximums values indicating NPS.
8. Now display the NPS for each flight line.

Data density refers to the amount of information captured and represented by data points within a specific area. it directly impacts the level of detail, accuracy, and usability of the captured data for various applications.

Methodology we use here is to calculates the data density using Thiessen polygons (also known as Voronoi polygons/ Delaunay triangulation) based on LiDAR data. Thiessen polygons partition a space into regions based on proximity to given points.

1. It loads LiDAR data from a CSV file.
2. Creates a Delaunay triangulation using the LiDAR data points.
3. Calculates the area of the Thiessen polygon for each LiDAR data point.
4. find its nearest neighbours (excluding itself) using the k-nearest neighbours (KNN) algorithm.
5. Calculate the area of the Thiessen polygon for each point based on its nearest neighbours.
6. Sorts these areas in descending order.
7. Determines the top 5% threshold for data density by taking the 95th percentile of the sorted areas.
8. Calculates the data density using the sorted areas, considering the top 5% threshold.

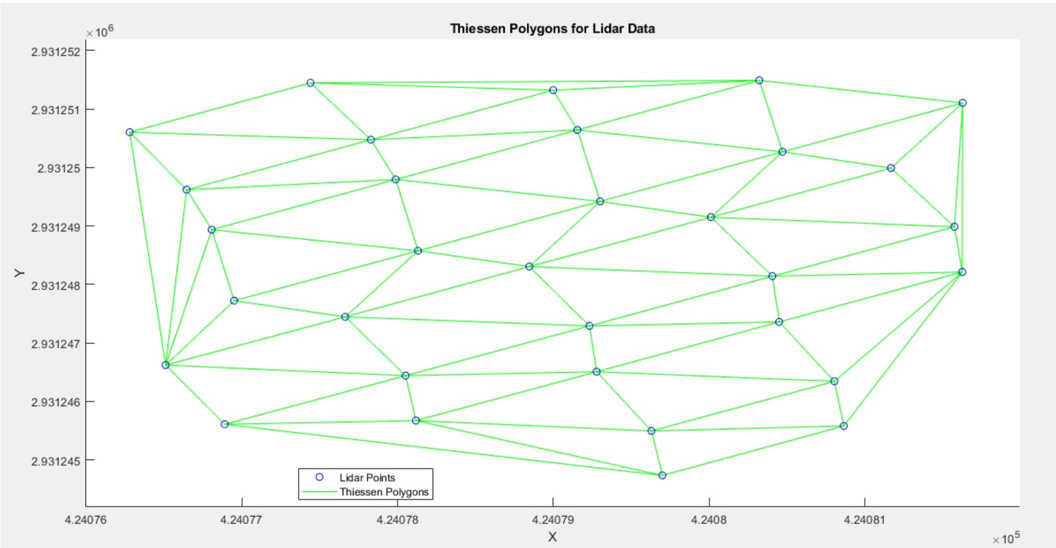



*Figure 14: Theisen polygon generated by MATLAB for point cloud.*

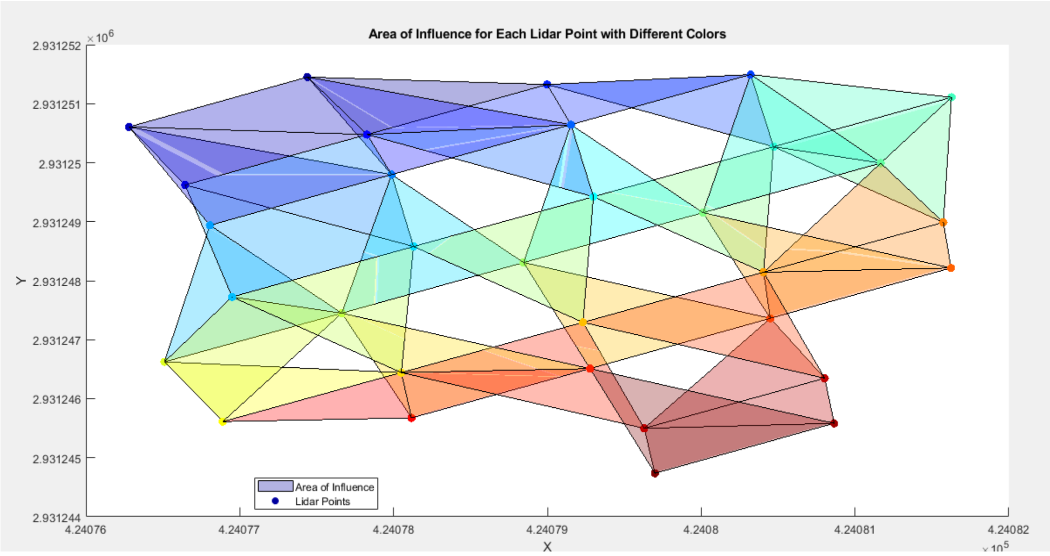



*Figure 15: Visualising for representation for area of Influence in MATLAB.*

## 3.6 Steps followed for Relative Accuracy:

Relative accuracy within the swath is a measure of how consistent and precise elevation measurements are within a single swath or flight line obtained from LiDAR data. It focuses on assessing how closely the elevation values for points within the same swath align with each other, indicating the level of reliability in the vertical information captured by the LiDAR sensor.

The process of determining relative accuracy typically involves selecting a test site within 90% of the swath's width, starting from the centre. This range is chosen because it covers a significant portion of the swath while avoiding potential edge effects that might occur near the boundaries. By focusing on a central portion, the analysis aims to capture the typical elevation variations present within the swath without being overly influenced by outliers or boundary conditions.

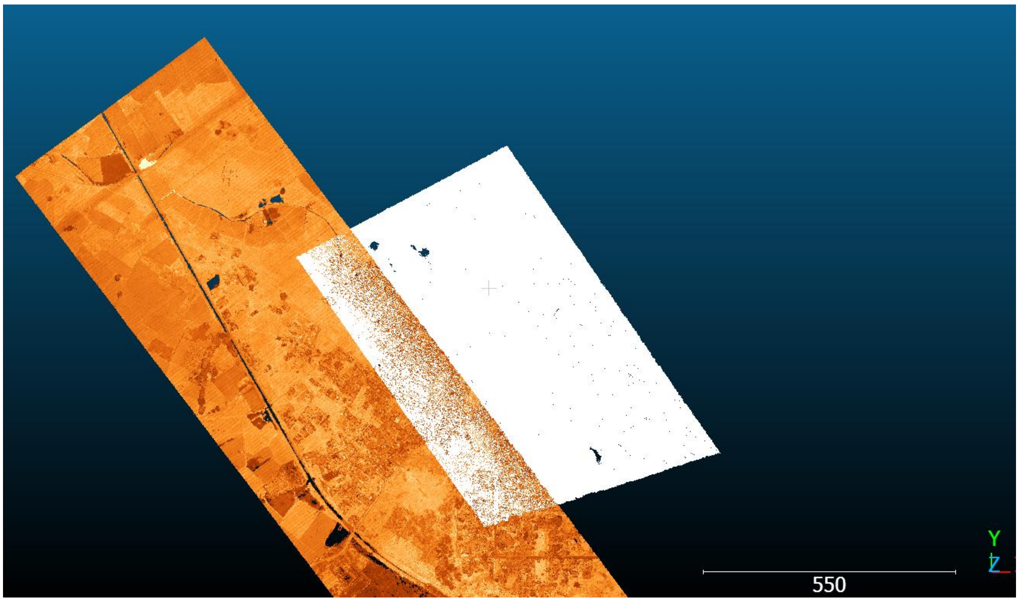



*Figure 16: Test site for flight line.*

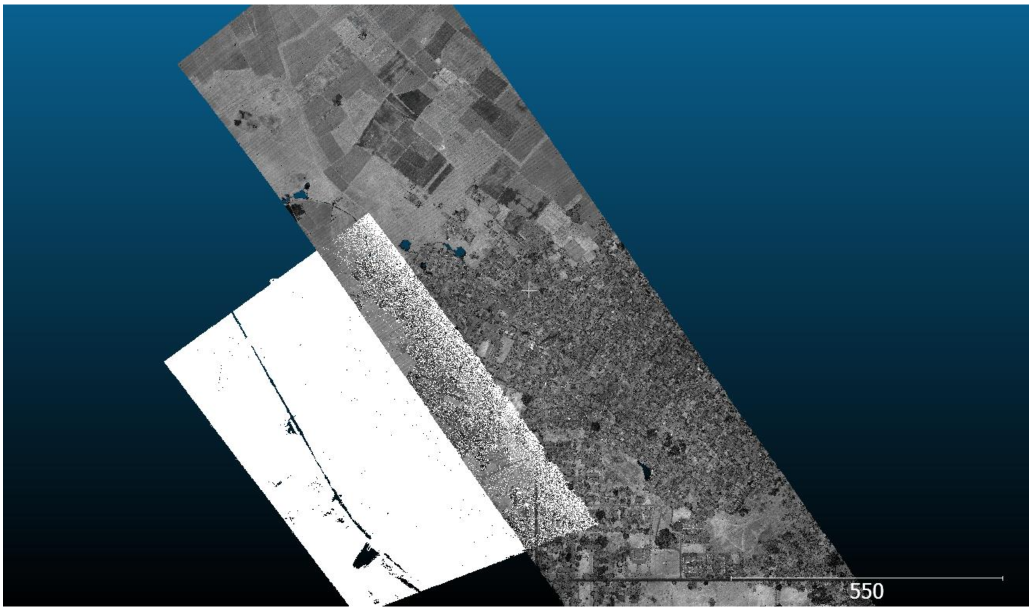



*Figure 17: Test site for flight line.*

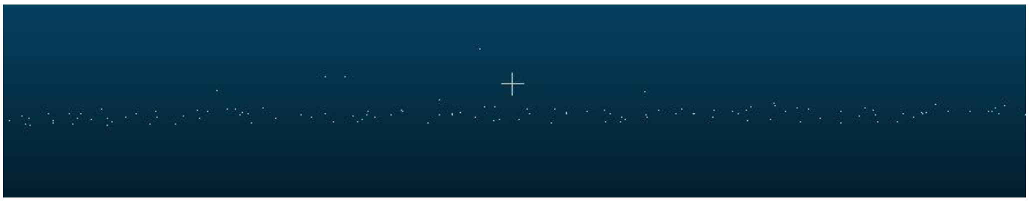


*Figure 18: Representation of relative accuracy of point cloud in CC*

1. loads a CSV file which likely contains LiDAR data points.
2. extracts the x, y, and z values from the data.
3. Calculate Relative accuracy as the standard deviation of the z-values multiplied by 100 to convert it to centimeters.
4. Similarly perform the relative accuracy for flight line 2.
5. elative accuracy can be estimated from the standard deviation of elevation (z) values. It also handles overlap regions between adjacent swaths to assess accuracy in those areas.

## Now for Overlap area:

Relative accuracy in overlap zones assesses the consistency and alignment of elevation measurements between adjacent swathes or flight lines that overlap in coverage. When LiDAR data is collected in multiple passes or from different flight lines, there is typically an overlap region where data points from adjacent swaths cover the same area on the ground. This overlap is intentional and serves several purposes, including improving data quality and enabling the creation of a continuous surface model.

1. loads CSV files containing LiDAR data from different flight lines.
2. extracts the z-values (elevation values) from these datasets.
3. elative accuracy is calculated as the standard deviation of the z-values. This approach assumes that a smaller standard deviation indicates higher accuracy and consistency in elevation measurements.

4. calculate standard deviation is then converted to centimeters by multiplying by 100, assuming the z-values are in a unit
5. Now, loads CSV files representing data from adjacent swaths, focusing on the overlap zones where the swaths cover the same area
6. extracts the z-values from both swaths.
7. compare the elevation measurements in the overlap zone, the code trims the datasets to ensure they have the same size (matching the smaller dataset's size).
8. elevation differences between corresponding points in the overlap zone are calculated by subtracting z-values from one swath from those of the other swath.
9. standard deviation of these elevation differences is then computed to quantify the relative accuracy in the overlap zones.

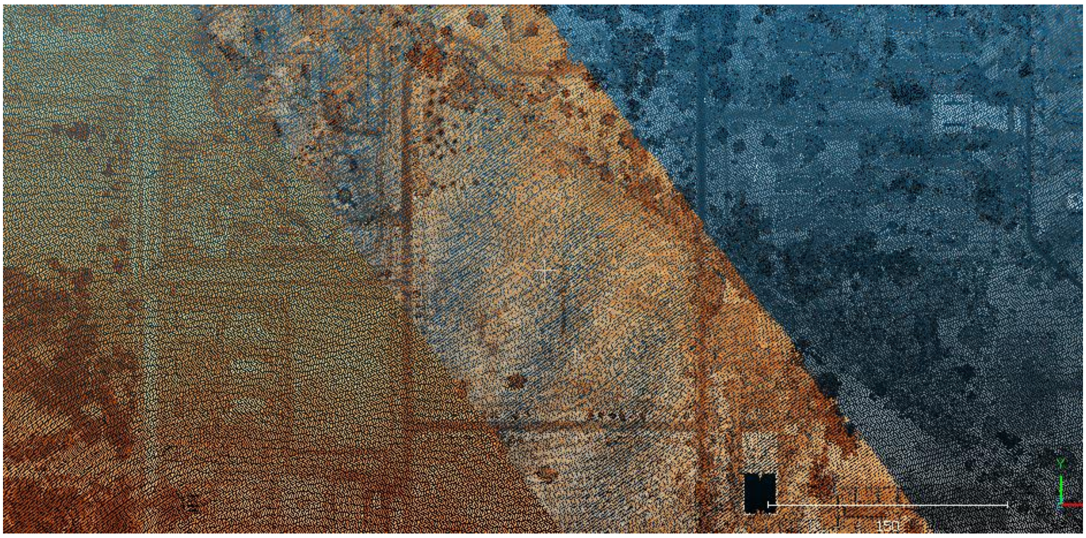


*Figure 19: Identifying the swath regions*

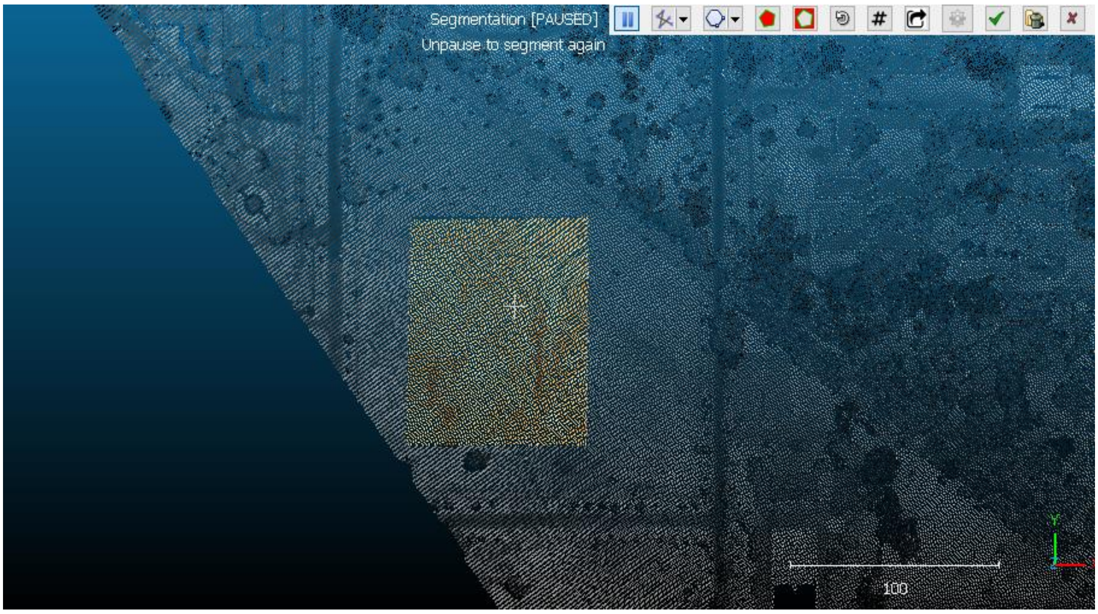



*Figure 20: Segmenting the patch.*

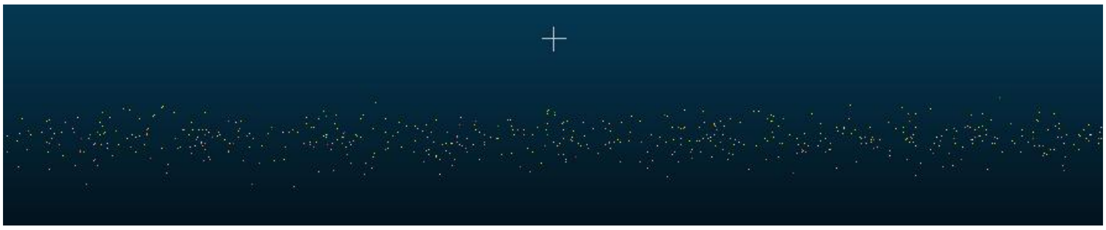


*Figure 21: Visualizing the swath in Cloudcompare*

## 3.7 Steps followed for Spatial Distribution

The spatial distribution of LiDAR data refers to how the data points are spread out across the surveyed area. It's about understanding the pattern and density of points. When we talk about uniform spatial distribution, it means that the points are evenly distributed across the area without significant clustering or gaps. On the other hand, a non-uniform distribution suggests variations in point density or clustering in certain areas, which could be due to various factors like terrain characteristics or sensor settings.

To study spatial distribution, a grid is often used as a reference. The grid size is typically determined by the NPS (Number of Points per Square). For example, if the grid is 2xNPS in size, it means each square in the grid covers an area that contains NPS LiDAR points.

To assess uniform distribution, a large batch of data is considered, covering around 90% of the survey area (swath). If 90% of the grid cells are populated with LiDAR points, it indicates a uniform spatial distribution. However, if less than 90% of the grid cells have LiDAR points, the distribution is considered non-uniform.

This method helps in quickly evaluating the overall spatial distribution pattern of LiDAR data, providing insights into data coverage and density across the surveyed area.

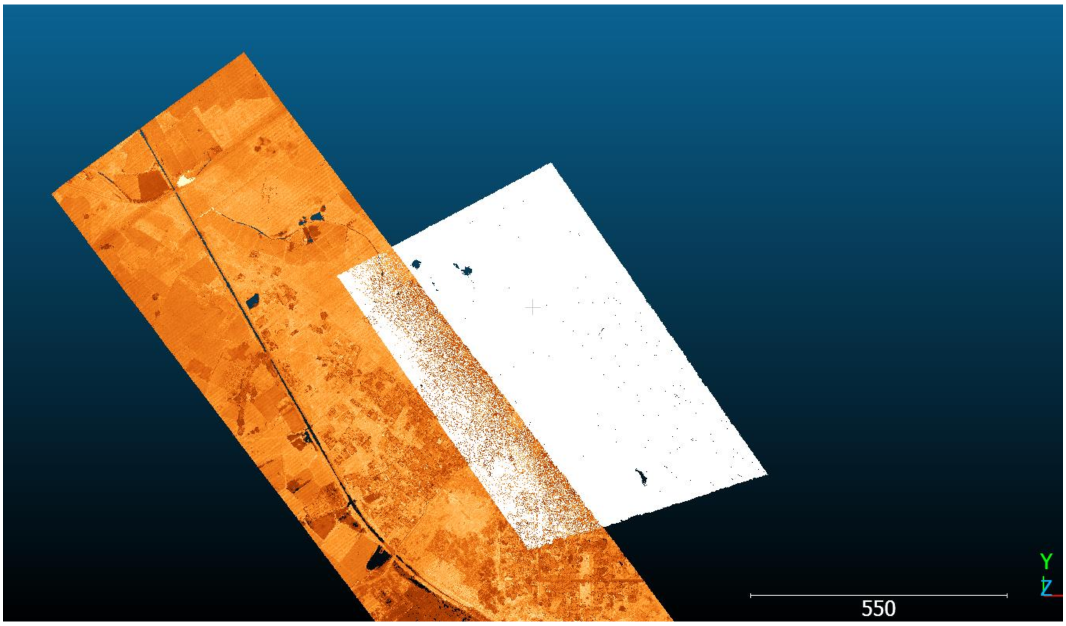



*Figure 22: Segmented test site for distribution.*

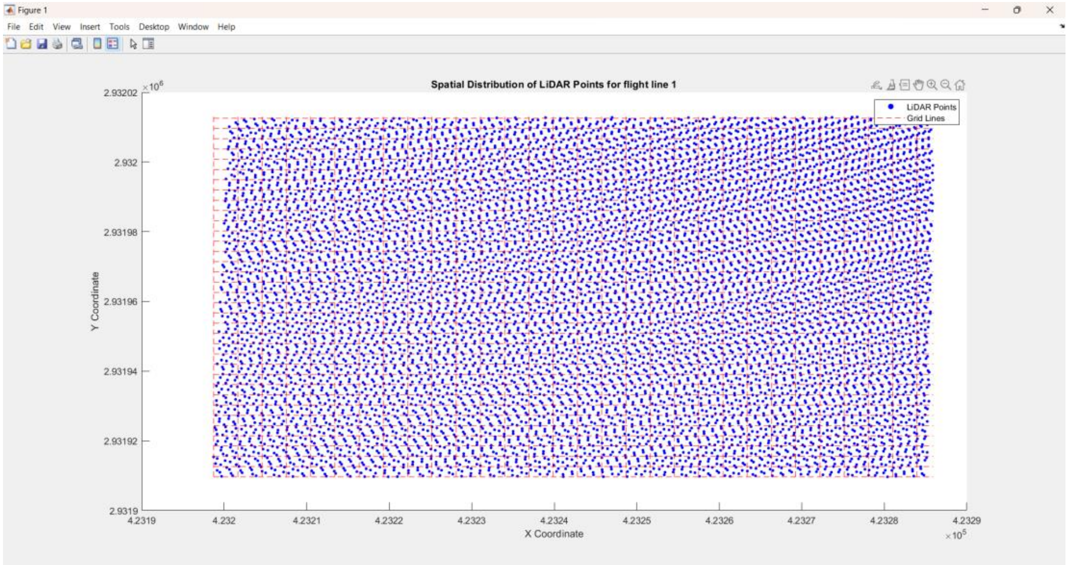



*Figure 23: Spatial Distribution of point in Flight line .*

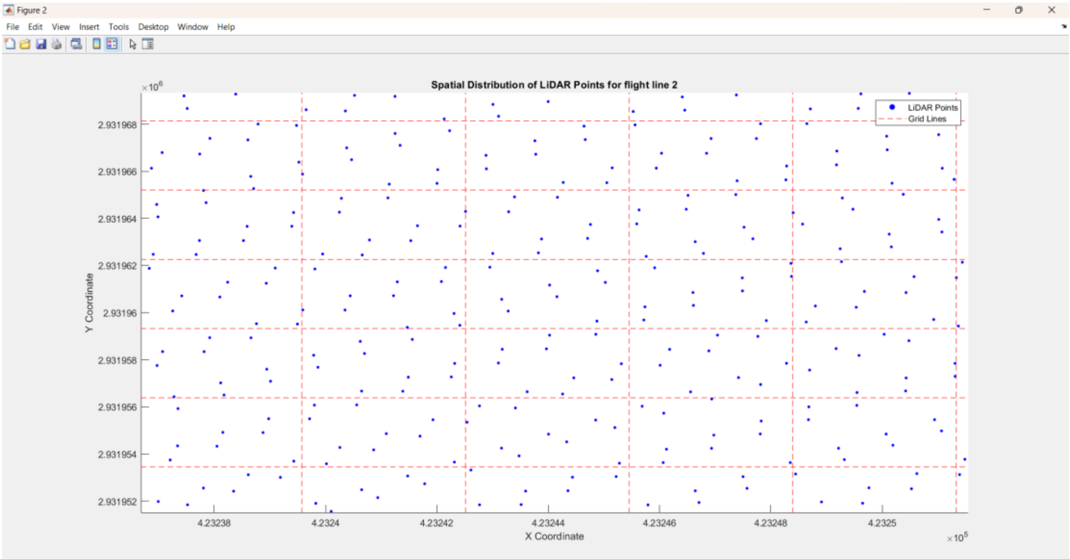



*Figure 24: Spatial distribution of lidar data of flight line 2 (Zoomed).*

1. Load LiDAR data from a CSV file for each flight line.
2. Extract the x, y coordinates, and z values from the loaded data.
3. Calculate the grid size based on NPS. The grid size is set as twice the NPS value.
4. Create the grid using the defined grid size and coordinate extents
5. count the number of LiDAR points falling within each grid cell.
6. Calculates the total number of grid cells based on the grid size and the number of cells in each dimension.
7. number of grid cells containing at least one LiDAR point is counted and the percentage of filled grid cells is computed.
8. Based on the percentage of filled grid cells, determines if the spatial distribution is uniform or non-uniform. If the percentage exceeds or equals 90%, it considers the distribution uniform; otherwise, it's considered non-uniform.
9. generates a scatter plot of LiDAR points for each flight line, with grid lines overlaid to visualize the grid cells. This scatter plot is shown in above.

# 4 Results

The following are the results obtained for the given data:

i. Number of returns
The number of returns per pulse for the given data is 3.

ii. Overlap percentage
Maximum overlay: 17.5903 %
Minimum overlay: 19.2237%
Average overlay: 18.407%

iii. Vertical accuracy
Non vegetated vertical accuracy for the data is 46.7 cm.
Vegetated vertical accuracy for the data is 360.5 cm.

iv. Horizontal accuracy
Horizontal accuracy for 95% confidence Northing : 0.87
Horizontal accuracy for 95% confidence Easting : 1.50

v. Relative Accuracy
Flight line 1: 3.7683 cm
Flight line 2: 8.0516 cm
Overlap regions: 19.1502 cm

vi. NPS
Nominal pulse spacing for Flight line1 : 1.4696 m
Nominal pulse spacing for Flight line2 : 1.505 m

vii. Data density
Data density for the given point cloud: 0.744 points/ $m^2$

viii. Data voids
Acceptable, as at least one data point is present in 90% of grid cell.

ix. Spatial Distribution
Uniform as both flight lines have at least one data point in 90% of the grid cells, indicating consistent coverage.

# 5 Discussions

a. Determine the horizontal and vertical datum employed in the data files and mention the same.
   - Horizontal datum gives a reference for measuring position in terms of co-ordinates, here, it is used as WGS-84, UTM zone 44 North projection.
   - While for vertical datum is use for reference for providing height measure. Here we have used as WGS-84 Ellipsoidal height.

b. RMSE Analysis
   Obtained RMSE for NVA is 46.7 cm, it falls into Class C according to ASPRS standards. This means the elevation data has a moderate level of accuracy, which is suitable for general mapping and land cover classification applications where moderate precision is acceptable.

# 6 Conclusion

This Projects gives a valuable insight into the factors affecting data quality and following learnings can be obtained by computing the results.

The number of returns per pulse indicates the density of information captured by LiDAR, influencing the quality and detail of the data. Overlap percentage and nominal pulse spacing contribute to data redundancy and resolution, affecting the precision and accuracy of elevation models or point cloud analyses. Vertical accuracy measures the reliability of elevation data, highlighting variations in accuracy between vegetated and non-vegetated areas. Horizontal accuracy indicates the precision of spatial positioning, crucial for mapping and geospatial applications. Relative accuracy assesses the consistency and alignment between data sources, essential for integrating datasets or comparing different surveys.Spatial distribution reflects the uniformity and completeness of data coverage, influencing the reliability and comprehensiveness of analyses and interpretations. Data density affects the level of detail and information available for analysis, impacting the resolution and accuracy of derived outputs. Detecting and addressing data voids ensures that analyses are based on complete and representative datasets, reducing potential biases or inaccuracies.

# 7 References

[1] H. G. M. George Vosselman, Airborne & Terrestial Laser scanning, USA: CRC Press, 2010.

[2] J. S. &. Toth, Topographic Laser Ranging & Scanning, Florida: CRC Press, 2009.

[3] A. 1, “LiDAR Accuracy An Airborne Perspective,” Airborne Corporation, California , 1998.

[4] C. P. Humbolt.edu, “Accuracy Metrices,” [Online]. Available: https://gsp.humboldt.edu/olm/Courses/GSP_216/lessons/accuracy/metrics.