CE676A: Laser Scanning and Photogrammetry
LiDAR Data Quality Assessment Project

Rithvik L
231030045
Civil Engineering
rithvikl23@iitk.ac.in

---

1 Introduction

In the realm of LiDAR (Light Detection and Ranging) data analysis, understanding the quality of LiDAR data is fundamental. Aerial LiDAR technology has emerged as a powerful tool for acquiring high-resolution, three-dimensional data over large areas with unparalleled precision and accuracy. However, the utility of LiDAR data critically depends on its quality, accuracy, and reliability. Assessing the quality of aerial LiDAR data is imperative to ensure its suitability for intended applications and to minimize errors in subsequent analyses and decision-making processes. This quality assessment encompasses a comprehensive evaluation of various parameters that characterize the accuracy, precision, and completeness of the LiDAR dataset. In the case of IIT Kanpur, the quality assessment parameters outlined provide a structured framework for evaluating the reliability and usability of the acquired LiDAR data.

Understanding the number and characteristics of LiDAR returns is essential for assessing data completeness and penetration capabilities, which are crucial for capturing features across different land cover types. Evaluating the overlap between multiple flight lines enables the identification of discrepancies and inconsistencies in data acquisition, which may affect point density and accuracy. Assessing the absolute vertical accuracy of LiDAR-derived elevations through comparison with ground truth data collected using GNSS instruments ensures the reliability of elevation measurements for terrain modelling and elevation analysis. Determining the absolute planimetric accuracy of LiDAR data involves verifying the positional accuracy of features in the horizontal plane, which is critical for mapping and spatial analysis applications. Evaluating the relative accuracy within swath and in overlap areas helps identify variations in point cloud density and positioning accuracy, which may arise due to flight geometry or data processing methods. Assessing the point cloud's nominal pulse spacing (NPS) and data density provides insights into data quality and coverage, influencing the precision and detail of derived products. Identifying and evaluating data voids or gaps in the LiDAR coverage is essential for understanding data completeness and identifying areas where additional data collection or interpolation may be required. Analysing the spatial distribution of LiDAR points helps assess the uniformity of point density and identify spatial patterns or anomalies that may affect data quality and interpretation.

2 Objectives

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

1. LiDAR returns – the number of returns in the data.
2. LiDAR overlap - the maximum, minimum, and average overlap in the two flight lines.
3. Altimetric (vertical) accuracy – the absolute vertical accuracy.
4. Planimetric (horizontal) accuracy – the absolute planimetric accuracy.
5. Relative accuracy – the relative accuracy within swath and in overlap area.
6. NPS and data density – the values of NPS and data density.
7. Data voids – identifying data voids that are not acceptable.
8. Spatial distribution (clustering and uniformity) – identifying if the data can be qualified as uniformly distributed.

3 Data

1. LiDAR Data: We are being provided LiDAR point cloud as a .las file of IIT Kanpur. The data are of two flights with a certain overlap between them.
2. Ground-Truth Data: This data is collected using a high-precision receiver like the R10 receiver. It serves as a reference or truth against which the accuracy of the flight data is evaluated. It consists of X, Y, Z coordinates representing known accurate positions and care should be taken so that it has the same horizontal and vertical datum as LiDAR data.

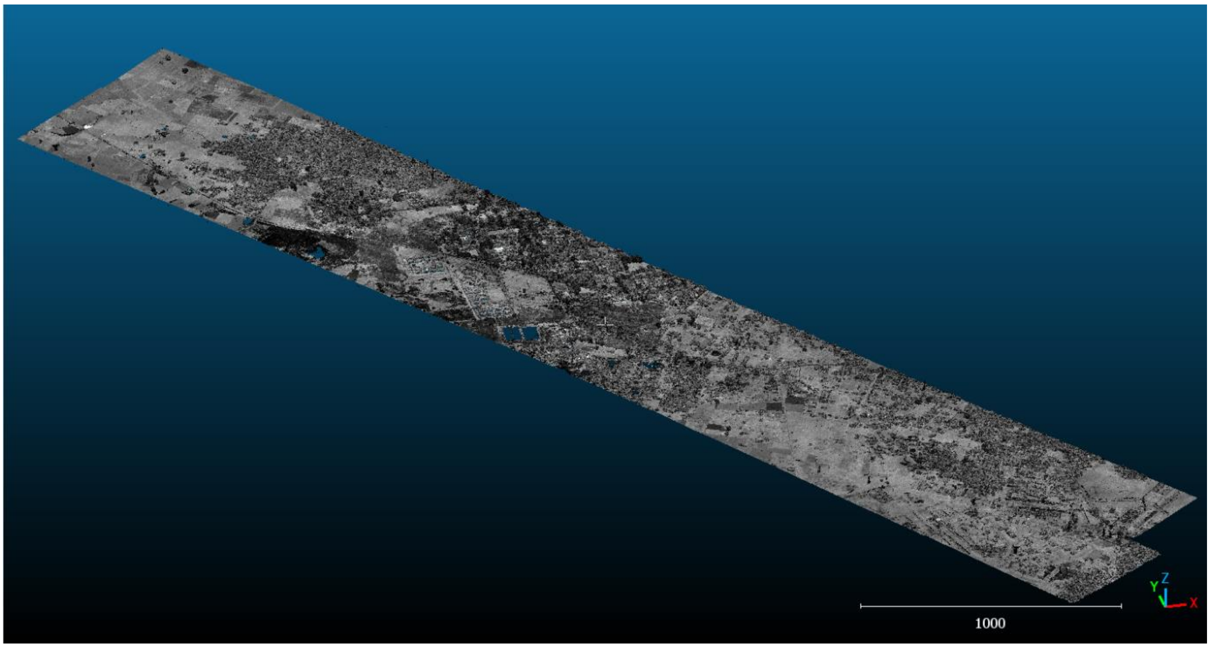


Figure1: Given data of the two flight lines visualized in cloudcompare based on their intensities.

4 Methodology

1. LiDAR returns- number of returns in the data.

   The objective of analysing LiDAR returns is to determine the quantity and distribution of returns within the dataset. This analysis helps in understanding the vertical structure of the environment. Differentiating between different return types (e.g., first, last, and intermediate returns) provides information about surface composition and structure.

   - The given data was inputted into lasinfo tool in QGIS.
   - This will result in a text file that contains all the information about the file.
   - The maximum number of returns in the data is highlighted below:

```
reporting minimum and maximum for all LAS point record entries ...
  X             421638683  425103247
  Y             929212962  933725973
  Z                 50766     233722
  intensity             0        255
  return_number         1          3
  number_of_returns     1          3
  edge_of_flight_line 0            1
  scan_direction_flag 0            1
  classification        1         12
  scan_angle_rank     -17         15
  user_data             0          0
  point_source_ID       0        205
number of first returns:        3199307
number of intermediate returns: 1652
number of last returns:         3199137
number of single returns:       3146327
overview over number of returns of given pulse: 3146327 102654 4788 0 0 0 0
histogram of classification of points:
          589634  unclassified (1)
         2664135  overlap (12)
```

Figure2: Information of the .las file yielded by QGIS tool lasinfo displayed in .txt format.

2. LiDAR overlap - the maximum, minimum, and average overlap in the two flight lines.

   When conducting LiDAR surveys, it's essential to ensure adequate overlap between adjacent flight lines. The primary objective of analysing LiDAR overlap is to assess the extent of redundancy between adjacent flight lines. Specifically, this involves determining the maximum, minimum, and average overlap values within the dataset.

   - Import las file into cloud-compare. Segment the overlap area into number of patches. Calculate the area of each patch.
   - Analyse the areas of each patch and where the maximum area is observed, that will be the maximum overlap.

- The minimum area refers to the minimum overlap. The average of all the areas will give us the average overlap of the two flight lines.
- We could observe manually that the maximum overlap is at the middle of the overlap and the minimum overlap is at the ends of the overlap.

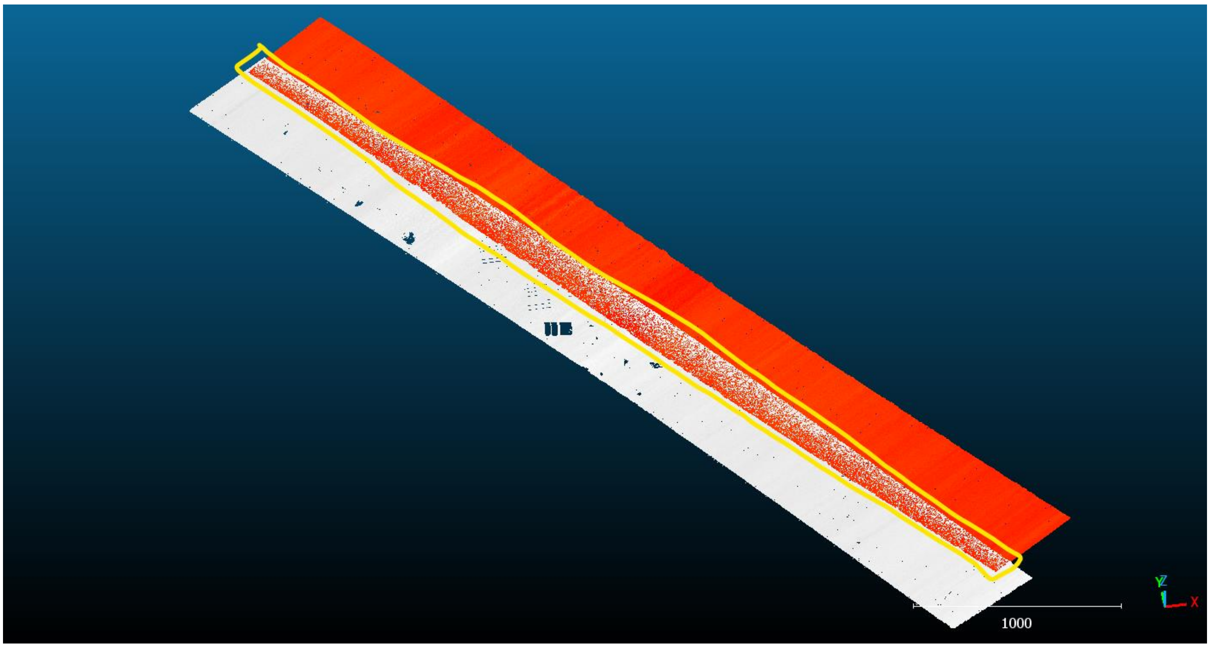


Figure3: Overlap between the two flight lines highlighted in yellow box.

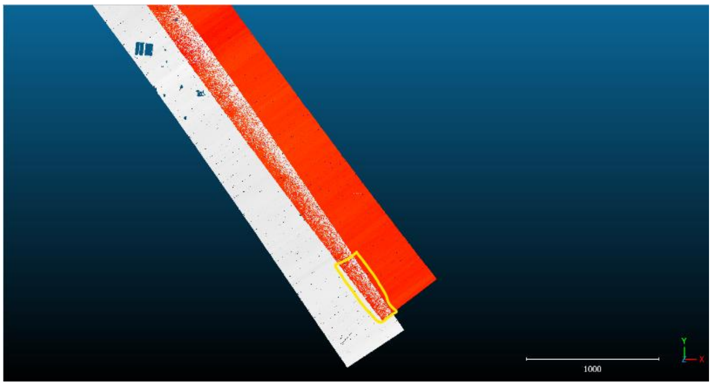


Figure4: Minimum overlap at the beginning of flight line.

- The minimum overlap was found out to be at the beginning and the end of the two flight lines.
- The minimum overlap was found out using compute surface area tool in cloudcompare.
- The minimum overlap was roughly found out to be 20%.

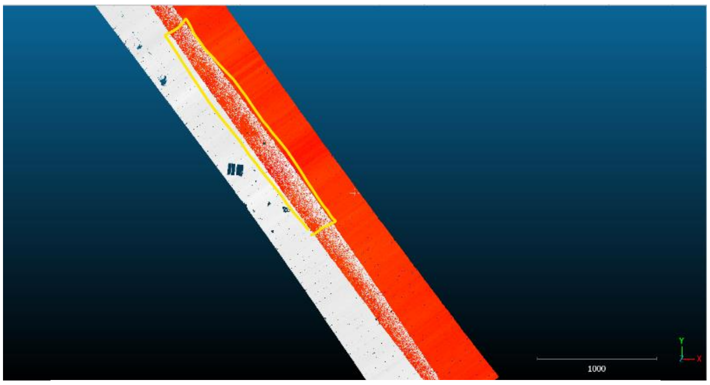
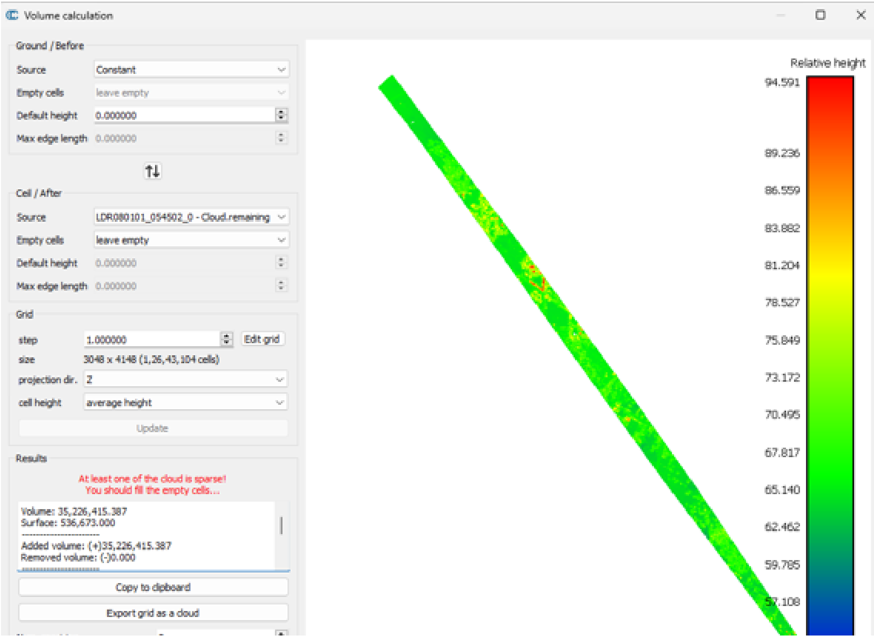


Figure5: Maximum overlap at the middle of flight

- The maximum overlap was found out to be at the middle of the intersection of the two flight lines.
- The maximum overlap was found out using compute surface area tool in cloudcompare.
- The maximum overlap was roughly found out to be 35%.



- The average overlap in the two flight lines computed by considering the entire overlap area and dividing the area by area of one flight line.

- The average overlap was found out to be roughly around 30%

Figure6: Entire overlap considered in the two flight lines for average overlap.

3. Altimetric (vertical) accuracy – the absolute vertical accuracy.

   Altimetric accuracy refers to the precision and reliability of elevation measurements derived from LiDAR point cloud data. The primary objective of analysing altimetric accuracy is to assess how closely LiDAR-derived elevations match the true elevations determined using GNSS instruments. This involves quantifying the vertical discrepancy between LiDAR measurements and ground truth data obtained in RTK mode.

   - Load LiDAR data and GCPs data: The code loads two CSV files containing LiDAR data and GCPs data, respectively.
   - Extract coordinates: It extracts X, Y, and Z coordinates from both the LiDAR data and the GCPs data.
   - Create Delaunay triangulation: It creates a Delaunay triangulation from the GCPs data. This triangulation is used to define the ground plane for each LiDAR point.
   - Calculate perpendicular distances: For each LiDAR point, the code iterates through each triangle formed by the GCPs and calculates the perpendicular distance from the LiDAR point to the plane defined by the triangle.
   - Calculate accuracy metrics: The code calculates the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) of the perpendicular distances. These metrics represent the vertical accuracy of the LiDAR points relative to the ground plane defined by the GCPs.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(h_i - h_{ref})^2}$$

Reporting Vertical Accuracy,

$$\text{Vertical accuracy} = 1.96 \times \text{RMSE}$$

4. Planimetric (horizontal) accuracy – the absolute planimetric accuracy

The primary objective of analysing planimetric accuracy is to evaluate how closely the horizontal positions derived from LiDAR data align with true ground positions. This involves quantifying the horizontal discrepancies between LiDAR-derived coordinates and ground truth data obtained from high-precision surveying methods.

- Define Flight and GCP Coordinates: The Flight LiDAR coordinates (x, y) and corresponding GCP coordinates (x1, y1) are defined. These coordinates represent the points captured by LiDAR and the ground truth points, respectively.
- Calculate Distances: For each LiDAR point, the Euclidean distance between its coordinates and the coordinates of the corresponding GCP is calculated. This distance represents the deviation between the LiDAR point and the ground truth.

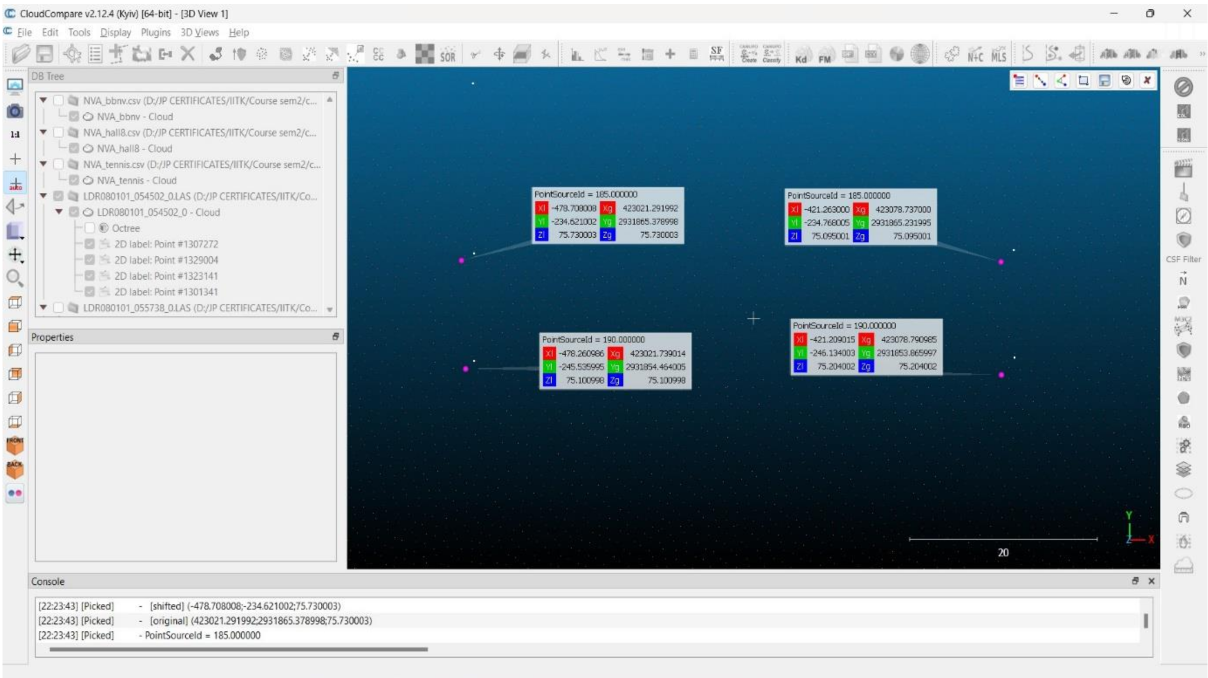


Figure7: Selecting points that is near to control point in CC.

- Calculate RMSE: The Root Mean Square Error (RMSE) is calculated using the formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i)^2}$$

Where d_i represents the distance between the i^{th} LiDAR point and its corresponding GCP, and n is the total number of points. LiDAR point and its corresponding GCP, and n is the total number of points.

5. Relative accuracy – the relative accuracy within swath and in overlap area.

   Relative accuracy assessment is a vital aspect of evaluating LiDAR data quality, particularly concerning the consistency and precision of measurements within and between flight lines. It focuses on understanding the variations and disparities in elevation and position measurements across the LiDAR dataset, both within individual swath areas and in overlap regions where adjacent flight lines intersect.
   The primary objective of analysing relative accuracy is to assess the consistency and reliability of LiDAR measurements within swath areas and overlap regions. This involves quantifying the degree of variation or discrepancy in elevation and position measurements relative to a reference standard, typically ground truth data or control points.

   - Import the two flight line data. Select a patch from any one of the flight line data. Make sure the patch is taken on a plane area.
   - Consider the Z values of the patch and export them into csv file.
   - Compute the standard deviation of Z values of the patch, which is the measure of relative accuracy.
   - Similarly, we computed relative accuracy for swath and overlap areas. The results are reported in the discussion section.

6. NPS and data density - the values of nominal pulse spacing (NPS) and data density

   The primary objective of analysing NPS and data density is to quantify the number of LiDAR data points within a defined area and assess their spatial distribution. This involves calculating the density of LiDAR points per unit area and understanding how it impacts the quality and usability of the dataset. NPS and data density metrics provide critical information about the level of detail and coverage available in LiDAR point cloud data.

   - Read the LiDAR Data: Read the LiDAR data from the CSV file, assuming that it contains columns for x, y, and z coordinates.
   - Extract Coordinates: Extract the x, y, and z values from the LiDAR data.
   - Calculate Volume: Determine the area covered by the LiDAR data by calculating the difference between the maximum and minimum x and y coordinates. This represents the area over which the LiDAR points are distributed.

   - Calculate Data Density: Divide the total number of LiDAR points by the area to obtain the data density, which is expressed as points per square meter.

$$\text{Data Density} = \text{number of points/area}$$

- Calculate NPS: Apply the equation to compute the NPS value based on the data density.

$$\text{NPS} = \frac{1}{\sqrt{\text{Data Density}}}$$

7. Data voids – identifying data voids that are not acceptable.

   The primary objective of analysing data voids is to identify areas within the LiDAR point cloud data where information is missing or insufficient. This involves assessing the extent and distribution of data voids and determining whether they are acceptable or pose significant limitations to the dataset's usability.

   - QGIS tools such as the rectangle or polygon selection tool was used to delineate the area that has no points.
   - The Calculate Void Area tool in QGIS was utilized to identify areas within the point cloud dataset where no or insufficient data points are present.
   - If the area computed is more than $4*(\text{NPS})^2$ and if area identified is not a waterbody or a low reflectance region, the area is considered to be a void.

Figure8: Comparing voids correspondingly with Google Earth Pro images.

   - Exceptions
     Water bodies
     Low reflectance areas for NIR regions

   - The above images, google earth image (on the left) and lidar data (on the right) were compared wherever a void is encountered in the data. It appears that the void area either belongs to a waterbody or a low reflectance area for NIR region.

   Hence it can be concluded that the data has no void that is not acceptable.

8. Spatial distribution (clustering and uniformity) – identifying if the data qualifies as uniformly distributed in planimetry.

   The primary objective of spatial distribution analysis is to determine whether the LiDAR point cloud data exhibit a uniform distribution in planimetry. This involves assessing the clustering or dispersion of data points across the surveyed area and identifying any spatial patterns that may influence data quality and interpretation.

   Spatial distribution analysis using the lasgrid tool in QGIS enables the assessment of point cloud data to determine if they exhibit a uniform distribution in planimetry.

   - The LiDAR point cloud dataset (.las file) that has only the first return is considered and imported into QGIS.
   - The lasgrid tool in QGIS was used to rasterize the LiDAR point cloud data.
   - The parameters for rasterization cell size (resolution) which is 2*NPS was entered in the size of grid cell field.
   - The lasgrid tool was run and the following output was obtained.

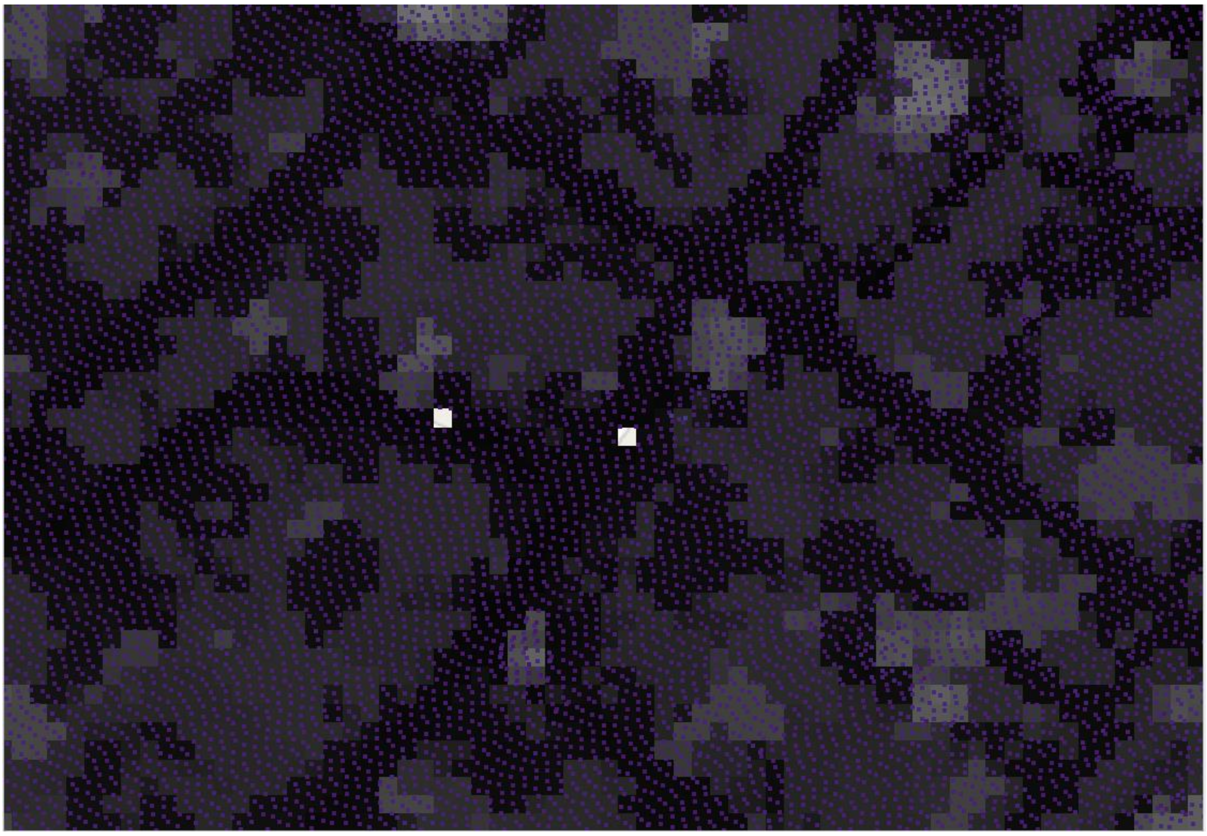


Figure9: LiDAR Points overlaid on grids to identify data distribution.

   - By visual inspection, it can be identified that more than 90% of the grids are filled with lidar points which justifies that the distribution of points is uniform.

5 Results and Discussion

1. The RMSE and vertical accuracy of different patches of both the vegetated and non-vegetated areas are displayed as below:

NVA area	RMSE	V. Acc
Tennis court	0.5	0.97
Hall8 road	0.2	0.38
Basketball court	0.13	0.25
VVA area	RMSE	V. Acc
Hall5 ground	0.59	1.17
GH1 ground	0.06	0.11

2. The horizontal accuracy of the given data by taking the the patch near the hall 8 buildings has been calculated as Horizontal Accuracy (RMSE): 1.39 meters.

3. The relative accuracy obtained for the lidar dataset is:

Area	Rel. Acc
Overlap	0.592688
Swath	0.061084

4. The data density and the NPS of the flight lidar data is obtained to be:
   Data density: 0.45 points per square meter.
   Nominal Pulse Spacing: 1.496849.

6 Conclusion

The assessment of quality parameters for aerial LiDAR data is essential for ensuring the accuracy, reliability, and usability of geospatial information across various sectors and applications. Through a systematic evaluation of parameters such as LiDAR returns, overlap, accuracy, data density, and spatial distribution, we have gained valuable insights into the strengths and limitations of LiDAR datasets.

Moving forward, continuous efforts in quality assurance and quality control are essential to maintain the integrity and usefulness of LiDAR data. Implementing best practices in data acquisition, processing, and validation will enhance the trustworthiness and applicability of

LiDAR-derived information, contributing to advancements in geospatial science and technology.

In summary, the assessment of quality parameters for aerial LiDAR data is a critical endeavour that underpins the success and impact of geospatial applications in various domains. Through rigorous quality assessment processes, we ensure the reliability and accuracy of LiDAR datasets, paving the way for innovation, informed decision-making, and sustainable development in our increasingly interconnected world.

7 References

- Lecture notes by Prof. B. Lohani.
- G Vosselman _HansGerd MaasAirborne and terrestrial laser scanningWhittles Publishing (2010)