A Structured Query Language Approach for processing Smartphone-based LiDAR of Understory Vegetation

Rory Pittman, Baoxin Hu

Dept. of Earth and Space Science and Engineering, York University, 4700 Keele Street, Toronto ON M3J 1P3 Canada
(rpittman, baoxin)@yorku.ca

Keywords: Terrestrial LiDAR, Data Definition Language, Understory Vegetation, Point Cloud Data, Rasterization

Abstract

LiDAR sensors incorporated within modern smartphone and tablet devices enable relatively quick and inexpensive collection of ground-based LiDAR data applicable for ground truth mapping as needed for modelling understory vegetation. However, this LiDAR data often requires conversion and processing prior to research use. This study presents a workflow with algorithms utilizing structured query language (SQL) to efficiently process detailed rasterized features from LiDAR data collected by an iPhone Pro Max via the ForestScanner app. After transformation of the LiDAR data, SQL has been employed to voxelize the LiDAR data from which rasterized features have been derived. Various cell sizes for voxels and subsequent pixels have been investigated, leading to a recommended spatial resolution of 0.05 m for cell size dimension. SQL provides precise control for advanced querying to process ground-based LiDAR data for vegetational modelling applications.

1. Introduction

Recently there has been increased interest for understory mapping of vegetation canopies, in connection to biodiversity studies (Bartels and Macdonald, 2023), carbon modelling (Dirnböck et al., 2020) as well as fuel loads for forest fire (Gopalakrishnan et al., 2025). Understory mapping can constitute as ground truth with vegetation modelling and is relevant for tree species classification due to associations and interactions between tree species and understory plants (Barbier et al., 2008). In addition, vegetation and soil properties are interconnected (Minasny et al., 2019) for the context of digital soil mapping. However, obtaining remote sensing data directly for understory vegetation from beneath overstory canopies can be challenging. Airborne LiDAR surveys may not capture adequate coverage of understory canopies, including for single photon LiDAR (Räty et al., 2022; Boretti, 2023). Terrestrial LiDAR enables the collection of detailed information encapsulating more precise structural attributes (Boretti, 2023) of understory vegetation. However, high costs of purpose-built mobile and handheld LiDAR devices hinder their applicability for research studies.

Advances with LiDAR sensors incorporated into smartphones and tablet computers have facilitated the retrieval of terrestrial LiDAR data for mapping vegetation (Tatsumi et al., 2022). Specifically, the lower costs of smartphones and tablets in comparison to conventional terrestrial LiDAR scanners have motivated their adoption for research use. Due to memory and power constraints as well as limitations with the range of retrievals (Tatsumi et al., 2022), data that has been acquired by smartphone LiDAR per recording is usually confined to small plots totalling less than a few hectares. Even so, as with other terrestrial LiDAR data (Chio, 2022), point densities can be large as tens of thousands of points can be collected per second. Consequently, although there is great potential for the retrieval of terrestrial LiDAR by smartphone or tablet devices, in practice this data is limited to smaller areas with correspondingly detailed but massive file sizes. Nonetheless, for research purposes this data can be integrated with data attained by sensors placed on UAV devices such as drones for observing overstory vegetation of the canopy.

Processing of LiDAR data acquired by handheld devices is not necessarily a straightforward task. Specifically, consistent methodology for resolving understory point cloud LiDAR data into various features should be implemented. This involves standard preprocessing, followed by specialized procedures for generating features. Preprocessing can entail the conversion of LiDAR files into alternative workable formats, including transformation of geocoordinates and adjustments with respect to the coordinate origin. Regarding processing, rasterized features can be constructed from LiDAR data (Takeshige et al., 2025), with topographic features typically generated from a digital terrain model (DTM) (Franklin, 2020). For airborne LiDAR collected for a canopy, features corresponding to a DTM and a digital surface model (DSM) are commonly of interest, from which a canopy height model (CHM) can be calculated. However, LiDAR from smartphone devices limited in range to a few meters is unlikely to capture retrievals for a full forested canopy layer. Therefore, features conforming to different attributes of understory vegetation should be derived. For data integration purposes, rasterized features are preferable as attributes resolved from other remote sensing technologies commonly correspond to 2-D rasterized products.

This study introduces a workflow for preparing LiDAR point cloud data as acquired from a handheld smartphone device, with algorithm code using structured query language (SQL) for generating features. Preprocessing entails transforming of LiDAR data into georeferenced point clouds, followed by processing into voxels from which pixels for rasterized features as of usage for remote sensing applications are subsequently derived. SQL presented in this study is adaptable to deriving rasterized features from terrestrial LiDAR point cloud data in general. The steps for preparing the LiDAR data and generating rasterized features are detailed in the following sections.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-237-2026 | © Author(s) 2026. CC BY 4.0 License.

2. Methodology

2.1 Data and Preprocessing

LiDAR point cloud data was retrieved for the understory of various forested sites situated within the southern transition zone of the boreal forest near the community of Chapleau in Ontario, Canada, during August 2025. The study area containing these locations ranged from 47.5° N to 48.3° N, and from 83° W to 84° W. The understory LiDAR data was collected for plots corresponding separately to either of jack pine (*Pinus banksiana*), balsam fir (*Abies balsamea*), trembling aspen (*Populus tremuloides*), white birch (*Betula papyrifera*) or eastern white pine (*Pinus strobus*). Understory vegetation for these plots varied from nonvascular plants such as mosses at ground level, to herbaceous perennial plants, both deciduous and evergreen shrubs, and younger tree seedlings, generally of no higher than 1 m in height.

Retrievals for these plots were collected by a LiDAR sensor built into an iPhone 14 Pro Max using ForestScanner software (Tatsumi et al., 2022). In terms of size, these files were generally limited to approximately 10 million points and a few hundred MBs. The LiDAR data was obtained within a range of 5 m or so from the LiDAR sensor, acquired over a duration of 10 minutes or less. Afterwards, the LiDAR files were exported from the ForestScanner application in polygon (PLY) format. To convert to point clouds in LAS (LASer) format, these files were loaded in CloudCompare software. Within CloudCompare, the *Y* and *Z* attributes were interchanged so to conform with *Y* as northing coordinate and *Z* as elevation. For each plot, the data were then exported as point clouds in both LAS as well as text (TXT) format. Within these corresponding data formats, each line corresponded to one LiDAR point, with the horizontal and vertical coordinates recorded. These LAS were then read into Python using laspy (Laspy Developers, 2025), or as CSV (comma-separated values) or text input for the text format files.

An overview of the attributes recorded in the LiDAR data as exported into the LAS format is displayed in Table 1. The components for the easting (*X*), northing (*Y*) and elevation (*Z*) were all recorded with an origin corresponding to where the LiDAR data started being collected. Note at these origin points that a Trimble R1 GNSS receiver was utilized to attain origin geocoordinates accurate to within 2 m uncertainty for the eastings and northings. Red-green-blue (RGB) colour components (*Rf*, *Gf*, *Bf*) were also recorded within the LiDAR data, corresponding to additive colouring components for visualizing the point cloud imagery. Each colour component was rescaled as a frequency between 0 and 1, with a higher number denoting a greater intensity.

X	*Y*	*Z*	*Rf*	*Gf*	*Bf*
0.5624	-0.4902	-1.5744	0.55	0.51	0.50
0.5507	-0.4802	-1.5739	0.35	0.30	0.28
0.5493	-0.4672	-1.5679	0.32	0.27	0.25
⋮	⋮	⋮	⋮	⋮	⋮

Table 1. Fields for LiDAR point cloud data.

The LiDAR point cloud data was then transformed to the desired coordinate projection of NAD 1983 UTM Zone 17N. This LAS point cloud data was recorded with coordinates in units of meters [m], but with positive *X* and positive *Y* in opposite directions from the UTM Zone 17N projection; a multiplication by -1 was applied for each of these directions. In general, a rotation factor accounting for vertical and azimuth angles can be applied (Zeng et al., 2014; Chio, 2022). The geocoordinates recorded by the Trimble R1 GNSS receiver were used in converting the spatial coordinates in the local references to NAD 1983 UTM Zone 17N coordinates.

An example of understory LiDAR point cloud data as obtained for a plot corresponding to a mature eastern white pine is shown in Figure 1. The point cloud images here are presented as viewed in CloudCompare software. This plot was chosen for explication purposes due to the ease with identifying the location for the mature white pine tree. The left subfigure depicts the LiDAR point cloud as before transformation looking at ground level with features viewed from different distances. A photo of the plot at the ground level looking northward is shown in the middle subfigure. In the right subfigure is presented a top view looking downwards at the LiDAR point cloud, after coordinate transformation with the north direction along the top of the subfigure. For the top view profile, a scale bar for the extent of 1 m has been provided, which approximates the magnitude of the diameter of the tree trunk near the ground. Note that for orientation ascertainment purposes for this plot, that LiDAR data was not attained immediately to northeast of the mature pine tree.

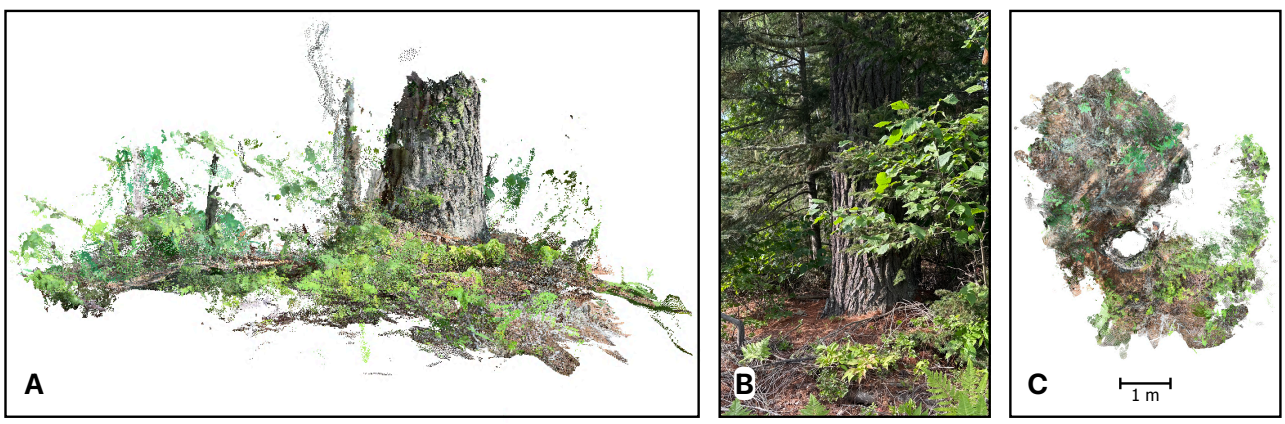

Figure 1. RGB colouration LiDAR point cloud data for a plot consisting of understory around a mature eastern white pine tree; A: Side profile of plot before coordinate correction; B: Corresponding photo of plot for reference, looking north; C: Top view profile of plot after coordinate correction.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-237-2026 | © Author(s) 2026. CC BY 4.0 License.

2.2 Processing

The coordinate transformed LiDAR point cloud data was processed using Python. Firstly, the general profile of the understory LiDAR data by elevation for each plot was computed, so to infer the height level above ground where the bulk of the LiDAR data was retrieved. This information supported in indicating height levels where rasterization of LiDAR data was of interest. A depictional overview of the approach for processing the LiDAR point cloud data into rasterized features is presented in Figure 2.

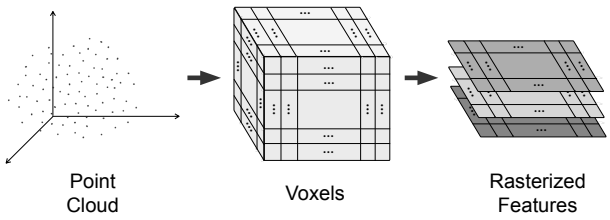

Figure 2. Overview of processing LiDAR point cloud data into voxels, followed by rasterized features.

The LiDAR point cloud data was processed by summarizing the data to the confines of voxels. Condensing the LiDAR point cloud data to voxels provided facilitation and consistency for the derivation of rasterized features, by firstly summarizing data to 3-D cubes and then eventually 2-D pixels. This approach of computing voxels reduced the data to more manageable units instead of millions of points, as well as enabling the computation of a variety of rasterized features for different height levels. The voxelization also mitigated the effects of imprecision with measurements, as each voxel was generally constituted from multiple LiDAR points per confines. Various cell size dimension ranging from 0.001 m to 0.5 m for the voxels were investigated, with computation times recorded. For this analysis, the dimensions of the voxels ultimately used were 0.05 m. A study (Tatsumi et al., 2022) using the ForestScanner application reported root mean square errors (RMSEs) for diameter at breast height (DBH) measurements with scanned trees using iPhone LiDAR of less than 3 cm. As well as due to motions arising from light wind in the understory, it was decided that voxels with side dimensions of 0.05 m provided for appropriate precision. Regarding summarization for the LiDAR points into voxels, elevation counts were summed. Alternatively, averages of the RGB intensities were tabulated for each voxel.

Features corresponding to both height and point density were computed. As the iPhone LiDAR also acquired RGB colouration attributes, a covariate relating to green intensity was considered, as this supported in differentiating tree trunks from understory foliage. A basic understory height model (UHM) was derived as the difference between the maximum point and minimum point per 0.05 m by 0.05 m ground pixel. As the ground-based LiDAR was generally obtained from above the understory vegetation, it was surmised that the LiDAR data mapped the top of the understory vegetation, here corresponding to an understory surface model (USM), more accurately than the ground surface beneath the vegetation. Due to the presence of leaf litter debris and denser moss vegetation on the ground, 0.05 m spatial resolution can be insufficient for obtaining a representative DTM. Consequently, an adjusted UHM was also calculated based upon a DTM constituting a 0.2 m by 0.2 m pixel at a 5% quantile level of LiDAR points near the ground. It was presumed that within a larger pixel it was more likely for having LiDAR points corresponding to the ground surface for calculating the minimum, resulting in a purportedly more accurate DTM. Slopes within the plots were minimal, so 0.2 m by 0.2 m resolution for the ground DTM was deemed sufficient. For the other features, the point densities and green intensities were calculated over all recorded understory height levels. Tree trunk locations were greatly indicated by the point densities, whereas the green intensities assisted with distinguishing plants and leaf foliage from tree trunks or larger branches.

Structured query language (SQL) (Teate, 2021) was employed for processing the LiDAR data and computing features. This comprised of the efficient calculation of voxels and rasterized attributes at a pixel level using standard SQL aggregation functions. The SQL used for generating voxels, as well as rasterized features of a point density, a green intensity relating to the green scaling of RGB colouring, an UHM and an USM is depicted in Figure 3. Here, the round function was exploited in setting the confines of the voxels and subsequent pixels by generating the lower coordinate bounds, based on an aggregation via the group by clause. As indicated in Figure 3, the spatial resolution for each cell size dimension was set to 0.05 m. This was done for each voxel and pixel populated with LiDAR data, with the resulting pixels tabulated as lists. SQLite (Gaffney et al., 2022) within Python was utilized for these computations. Aggregation functions within SQL were applied for the covariate calculations. A summation function was used for calculating the point density, with greater counts corresponding to a greater point density per pixel. The average of the green frequency values was computed for the green intensity. As for features relating to height, the maximum of the elevation measurement was utilized for the USM, with this basic UHM calculated as the difference between the maximum elevation and minimum elevation per pixel selection.

```
CREATE TABLE Plot_Voxels AS
SELECT ROUND(ROUND(X/0.05 - 0.05,0)*0.05,2) AS X_BASE,
    ROUND(ROUND(Y/0.05 - 0.05,0)*0.05,2) AS Y_BASE,
    ROUND(ROUND(Z/0.05 - 0.05,0)*0.05,2) AS Z_BASE,
    COUNT(Z) AS PT_COUNT, AVG(Gf) AS GREEN_AVG
FROM Plot_LiDAR_Reprojected
GROUP BY X_BASE, Y_BASE, Z_BASE;

CREATE TABLE Plot_Covariates AS
SELECT X_BASE, Y_BASE, SUM(PT_COUNT) AS PT_DENSITY,
    AVG(GREEN_AVG) AS GREEN_INTENSITY,
    MAX(Z_BASE) - MIN(Z_BASE) AS UHM,
    MAX(Z_BASE) AS USM
FROM Plot_Voxels
GROUP BY X_BASE, Y_BASE;
```

Figure 3. SQL used for generating voxels and rasterized features of point density, green intensity, a UHM and USM.

For computing the adjusted UHM, firstly a DTM was calculated with a larger spatial resolution of 0.2 m. To compute this DTM, more sophisticated SQL was warranted. The SQL for computing the larger base voxels and accumulation counts by the vertical voxel levels are presented in Figure 4. Within each larger voxel base of 0.2 m by 0.2 m, the 5% quantile level of LiDAR points from the ground was selected to correspond to the DTM. After generating the voxels with bases of 0.2 m by 0.2 m, but still with vertical dimensions of 0.05 m, an intermediate table was created to sum the LiDAR point counts vertically per voxel level. This was achieved by summing the points over the voxel bases using the partition by clause. Doing this computation allowed for querying voxel heights corresponding to any desired quantile level.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-237-2026 | © Author(s) 2026. CC BY 4.0 License.

```
CREATE TABLE Plot_Voxels_DTM AS
SELECT ROUND(ROUND(X/0.2 - 0.2,0)*0.2,1) AS X_BASE_DTM,
    ROUND(ROUND(Y/0.2 - 0.2,0)*0.2,1) AS Y_BASE_DTM,
    ROUND(ROUND(Z/0.05 - 0.05,0)*0.05,2) AS Z_BASE,
    COUNT(Z) AS PT_COUNT
FROM Plot_LiDAR_Reprojected
GROUP BY X_BASE_DTM, Y_BASE_DTM, Z_BASE;

CREATE TABLE Plot_Z_Counts_DTM AS
SELECT X_BASE_DTM, Y_BASE_DTM, PT_COUNT,
    SUM(PT_COUNT) OVER (PARTITION BY X_BASE_DTM, Y_BASE_DTM
        ORDER BY Z_BASE) AS ACCUM_COUNTS
FROM Plot_Voxels_DTM
GROUP BY X_BASE_DTM, Y_BASE_DTM, Z_BASE;
```

Figure 4. SQL for calculating retrieval counts per voxel with larger bases to use for digital terrain model (DTM).

Once the accumulation of points per level were calculated, the DTM was computed as the minimum height level where the accumulated point count corresponded to at least 5% of the total point count per voxel column, as the 0.05 m spatial resolution for the vertical extents of the voxels facilitated precision to within 0.05 m spatial resolution as used for the voxels. For this calculation, a where clause with a nested query was utilized, while also using a group by clause to aggregate for the minimum height level satisfying the where clause. Afterwards, this more representative DTM was used for the computation of an adjusted UHM, with the DTM constituted as the minimum elevation measurement that was subtracted from the USM. Here, a case clause was used for calculating the adjusted UHM. If the DTM was greater than the USM for a pixel, which in theory was possible as this DTM corresponded to the 5% quantile level above the ground for a larger area, then the UHM was assigned a value of zero. The SQL for computing the DTM and adjusted UHM is shown in Figure 5.

```
CREATE TABLE Plot_DTM AS
SELECT X_BASE_DTM, Y_BASE_DTM, MIN(Z_BASE) AS DTM
FROM Plot_Z_Counts_DTM
WHERE ACCUM_COUNTS >= (SELECT 0.05 * MAX(ACCUM_COUNTS)
    FROM Plot_Z_Counts_DTM GROUP BY X_BASE_DTM, Y_BASE_DTM)
GROUP BY X_BASE_DTM, Y_BASE_DTM;

CREATE TABLE Plot_UHM_DTM AS
SELECT DISTINCT a.X_BASE, b.X_BASE_DTM, a.Y_BASE, b.Y_BASE_DTM,
    a.USM, b.DTM, CASE WHEN a.USM >= b.DTM THEN (a.USM - b.DTM)
        ELSE 0 END AS ADJUSTED_UHM
FROM Plot_Covariates AS a FULL OUTER JOIN Plot_DTM AS b
    ON ROUND(ROUND(a.X_BASE/0.2 - 0.2,0)*0.2,2) = b.X_BASE_DTM AND
        ROUND(ROUND(a.Y_BASE/0.2 - 0.2,0)*0.2,2) = b.Y_BASE_DTM
ORDER BY a.X_BASE, b.X_BASE_DTM, a.Y_BASE, b.Y_BASE_DTM;
```

Figure 5. SQL for generating a larger base digital terrain model (DTM) and the adjusted understory height model (UHM).

The computations using SQL resulted in tables consisting of entries with each row corresponding to a different pixel for easting and northing bounds. For creating the rasterized features, the lists containing the pixels were restacked into 2-D arrays. It was crucial that all pixel coordinate combinations were represented, particularly pixels with no data values so that the restacked 2-D arrays conformed to imagery. Pandas (McKinney, 2010) within Python was employed for these restacking operations.

The output for each rasterized feature was a text file with each data point corresponding to a pixel, with a no data specification denoted for pixels with missing data. Header information containing the number of rows and columns, the respective lowest bounds for the northings and eastings coordinates, the cell size and the no data indicator were each specified. These files were then each saved in ASC (ASCII-encoded text) format and then loaded in QGIS software which was used for visualizing the rasterized features.

3. Results

LiDAR data for a variety of forested sites were processed using the SQL algorithms, but for the sake of brevity the results are primarily displayed for a plot corresponding to a mature eastern white pine. For the mature eastern white pine plot, the profile of understory LiDAR points is shown in Figure 6. In this plot, LiDAR counts were calculated as a line histogram chart along the vertical axis for elevation, with each bin of the histogram corresponding to 0.05 m resolution with respect to the elevation. This plot conformed to a relatively small area of less than 25 m^2, but with 7,936,933 LiDAR points. There was only a slight elevation difference within the site, with a recorded ground elevation of approximately 452 m for the mature eastern white pine tree. Inspecting this plot, the higher points generally corresponded to the tree trunk structures, with lower points to understory plants. For this plot, the understory plants consisted of ferns and deciduous shrubs. As observed, the profile counts were maximal at an elevation just over 452 m, corresponding to understory vegetation. Most the of the LiDAR data was obtained within 1 m or so of the ground.

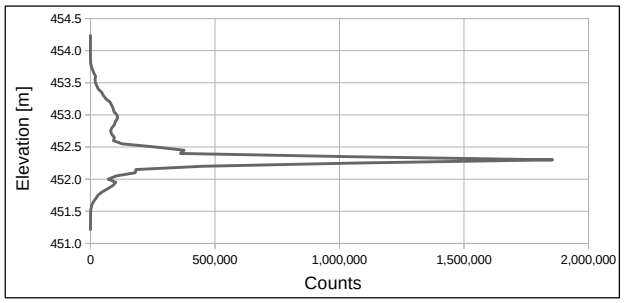

Figure 6. Vertical profile of LiDAR counts by elevation for plot with mature eastern white pine as shown in Figure 1.

The rasterized features of a point density regarding counts, green intensity, UHM, and an adjusted UHM were each computed and are displayed in Figure 7. Observing the subfigures in Figure 7, the location of the mature eastern white pine tree is evident. However, the locations of the tree trunks, here for this large eastern white pine as well as a smaller nearby coniferous tree, are most apparent by the higher density counts in Figure 7-A. These density counts ranged into the tens of thousands of points per 0.05 m by 0.05 m pixel for some tree bark, with lower density counts for the ground. Note that the gaps within these figures correspond to where no LiDAR retrievals were collected nor able to be obtained, such as within the trunk of the mature eastern white pine tree. The green intensity was able to somewhat differentiate between areas with more understory cover, when compared to the RGB composite LiDAR point cloud visualization in Figure 1-C. For the pixels, greater green intensity corresponded to higher values closer to 1. Heights within the understory were discerned by the UHM and adjusted UHM, with the tallest heights corresponding to the tree trunks. Detailing for some understory plants are observed, with generally slightly larger height measurements for the adjusted UHM as this used a DTM based over a larger pixel size. The adjusted UHM was smoother than the basic UHM for some of the lower vegetation in for some understory vegetation, with the larger 0.2 m by 0.2 m pixels based on the DTM apparent in the locations with sparser vegetation and height.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-237-2026 | © Author(s) 2026. CC BY 4.0 License.

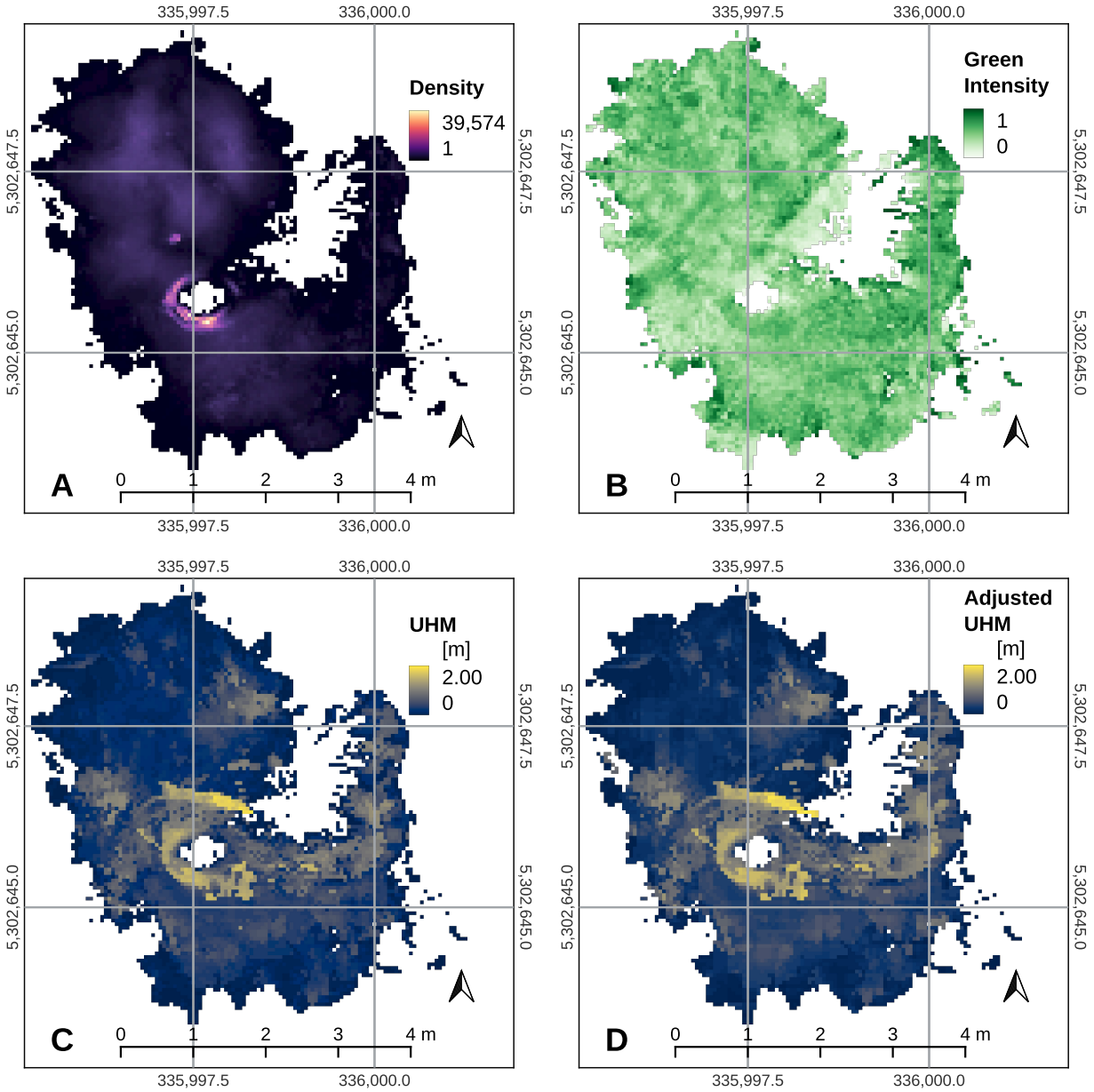

Figure 7. Rasterized features generated for a plot conforming to a mature eastern white pine; A: Density intensity; B: Green intensity; C: A basic understory height model (UHM); D: Adjusted UHM from considering a coarser digital terrain model (DTM). Eastings [m] and northings [m] are noted for coordinate grids in the NAD 1983 UTM Zone 17N projection.

For comparison purposes, the SQL for generating these features was reran by varying the corresponding voxel and pixel cell size, with the resulting computation times indicated in Table 2. Code was executed 5 times per specified cell size dimension, with standard deviations calculated and denoted after the plus and minus sign within the table. Note for these computations that the resolution for the DTM utilized for the adjusted UHM was fixed at spatial resolutions of 0.2 m, except for 0.5 m when the voxel cell size dimension was upscaled to 0.5 m. As noticed, computation times increased substantially as voxel sizes and corresponding pixel sizes decreased to less than 0.05 m. However, further increasing the voxel and pixel cell size dimensions over 0.05 m did not lead to material gains with lesser computation times. Therefore, selection of the spatial dimension of 0.05 m was sufficient for processing the voxels and pixels for the corresponding rasterized features.

Voxel and Pixel Cell Size Dimension [m]	Computation Time [s]
0.001	1036.005 ± 4.716
0.005	142.451 ± 1.534
0.01	60.476 ± 1.264
0.05	29.840 ± 0.364
0.1	28.260 ± 0.297
0.2	28.076 ± 0.253
0.5	27.962 ± 0.318

Table 2. Computation times for generating features for plot with eastern white pine as in Figure 1, varying voxel and pixel sizes. Average computation times with standard deviations are noted.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-237-2026 | © Author(s) 2026. CC BY 4.0 License.

Results of these analogous rasterized features for a plot consisting of mature jack pine trees are presented in Figure 8. As also seen in Figure 7, the density plots indicate the positions of the tree trunks by the higher density counts. Here, densities of at least a few thousand points corresponded to the tree trunks. Locations of logs and debris are apparent in the green intensity and UHM plots. To resolve better detailing with the understory, here the UHM plots only display within 0.5 m height above the ground surface. Also apparent are some branch structures of understory shrubs or younger secondary growth of trees.

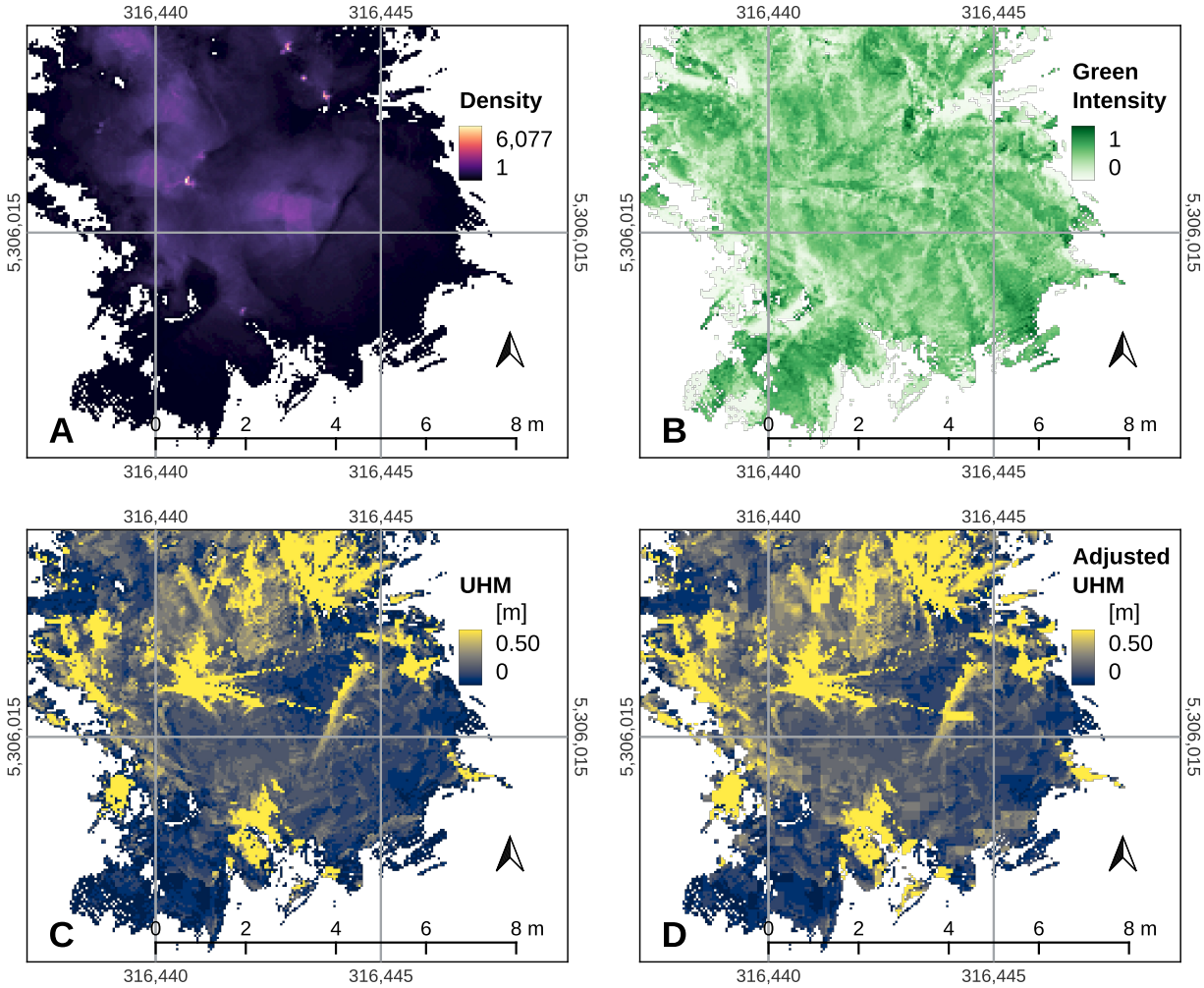

Figure 8. Rasterized features generated for a plot consisting of mature jack pine trees; A: Density intensity; B: Green intensity; C: A basic understory height model (UHM); D: Adjusted UHM from considering a coarser digital terrain model (DTM). Eastings [m] and northings [m] are noted for coordinate grids in the NAD 1983 UTM Zone 17N projection.

4. Discussion

LiDAR point cloud data was collected from understories corresponding to various tree species within boreal forest. However, LiDAR data was able to be retrieved more readily from understories of certain tree species. In general, trembling aspen understories presented the greatest difficulties with obtaining LiDAR data. Better success with retrieving LiDAR data was attained for conifer canopies, or where understory vegetation was not too dense with deciduous shrubs.

Due to the issue of gaps and limitations with plot size, understory LiDAR retrievals should be acquired for multiple plots in close proximity to one another. As an example, for a small study area of a few hectares that can be observed by drone imagery, understory LiDAR for dozens of plots should be obtained. Limitations with plot sizes can arise on account of data and processing constraints with the device utilized for collecting LiDAR. During the August 2025 field campaign, the maximum plot sizes attainable without capping out corresponded to around a few hundred m^2 and just over 10 million LiDAR points. Regarding UAVs, for drone sensors the optical imagery with respect to surface reflectance can be attained at 0.01 m (Zhang et al., 2022) upwards to around 0.3 m (Shin et al., 2019) spatial resolutions. Therefore, pixel cell sizes of 0.05 m for LiDAR-derived features suffice as these covariates can be resampled to larger cell sizes.

Part of the challenge of adapting detailed LiDAR data for small plots contends to data integration with other geospatial data. When data is retrieved over larger areas at coarser spatial resolutions, errors for geocoordinates is lesser of an issue. However, for plot sizes totalling less than a hundred m^2 with extents only in magnitude of 10 m per dimension, then

uncertainties with geospatial coordinates are critical. Furthermore, there can also be deviations pertaining to the bearing orientations. Many handheld GPS devices as used in field campaigns are accurate to within 8 m (Lee et al., 2023); R1 or L1 capable devices can attain accuracies within a couple of m (Lee et al., 2023; Tatsumi et al., 2022). In integrating imagery from different sources, it can be difficult or even impractical to align tree trunks with their corresponding overstory canopies. Generally, with handheld LiDAR data there can also be gaps within the LiDAR coverage. These issues can hinder the utility of understory LiDAR data in conjunction with other data attributes. Consequently, depending on the intended usage, processing of understory LiDAR data is a nontrivial task.

For comparison purposes, single photon LiDAR (SPL) point cloud data as retrieved by aerial survey was obtained for these sites from the Ontario Ministry of Natural Resources (OMNR). This SPL data for the plot consisting of a mature eastern white pine as seen in Figure 7 is shown in Figure 9 as looking in a northward direction. The approximate point density for the SPL data was 36 points per m^2. As seen within Figure 9, there was very little SPL coverage for understory vegetation beneath the mature eastern white pine tree. Most of these retrievals corresponded to the top canopy structure of the tree, or else the ground where little vegetation cover was present. For some of the shrubs to the immediate east of the tree, the vegetation was approximately a meter or so in height, which compares in magnitude to the UHMs in Figure 7.

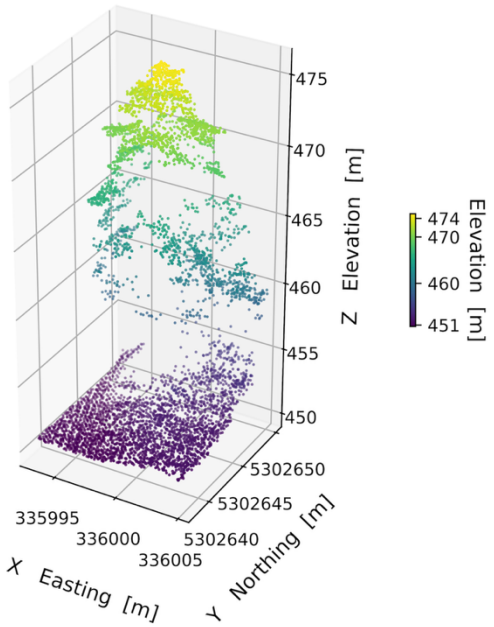

Figure 9. 3-D profile of SPL data by elevation for plot with mature eastern white pine. Eastings [m] and northings [m] are in the NAD 1983 UTM Zone 17N projection, with gridlines marked for every 5 m distance.

Recent studies have investigated iPad LiDAR collection within boreal forest in Ontario (Guenther et al., 2024a; Guenther et al., 2024b) for the estimation of tree DBH. These studies collected LiDAR data for small plots within radii of 11.28 m for vegetation sampling network (VSN) plots in northwestern Ontario. The accuracy of the DBH measurements generated from the iPad LiDAR had mean absolute error (MAE) of less than 2 cm for various sites corresponding to different boreal tree species composition and age stands (Guenther et al., 2024b). Another study (Tatsumi et al., 2022) reported root mean square errors (RMSEs) for DBH measurements of less than 3 cm using iPhone LiDAR. Therefore, these errors from iPhone LiDAR were within the precision of 0.05 m spatial resolutions that can be utilized for voxelization and rasterization of features.

The SQL code presented in the methodology section can be applicable for voxelization and deriving the according rasterized features in general from LiDAR point cloud data. For SPL or other LiDAR data of lower point densities, obviously larger spatial resolutions for the voxels and pixels should be utilized. For data integration purposes, a spatial resolution matching to that from multispectral imagery attained for the study area should be considered, providing there is sufficient point density per specified pixel size. The procedure described in this study for obtaining the DTM can be useful for calculating more representative DTMs from terrestrial LiDAR data.

Future work includes the linking of these features to those generated for overstory or midstory canopy vegetation. The connection between overstory attributes corresponding to tree species and that to understory vegetation with respect to biodiversity or carbon considerations will likely be investigated in future research.

5. Conclusions

A process incorporating SQL was developed to effectively generate rasterized features from LiDAR point cloud data collected via a smartphone. This LiDAR point cloud was summarized into voxels, from which rasterized features for a point density, green intensity and understory height models were subsequently computed. A spatial cell size resolution of 0.05 m was utilized for the voxels and pixels for the rasterized features, which adequately captured sufficient detail for the features so to match or exceed spatial resolution from features generated from sensors mounted on drones or higher resolution satellite imagery. SQL facilitated enhanced control and consistency with the voxelization and rasterization of features.

Acknowledgements

Funding for this research was provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Single photon LiDAR (SPL) for comparison purposes was obtained from the Ontario Ministry of Natural Resources (OMNR) and contains information licensed under the Open Government license – Ontario.

References

Barbier, S., Gosselin, F., Balandier, P., 2008: Influence of tree species on understory vegetation diversity and mechanisms involved—A critical review for temperate and boreal forests. *Forest Ecology and Management* 254(1), 1-15. doi.org/10.1016/j.foreco.2007.09.038.

Bartels, S.F., Macdonald, S.E., 2023: Dynamics and recovery of forest understory biodiversity over 17 years following varying levels of retention harvesting. *Journal of Applied Ecology* 60(4), 725-736. doi.org/ 10.1111/1365-2664.14366.

Boretti, A., 2024: A perspective on single-photon LiDAR systems. *Microwave and Optical Technology Letters* 66(1), 1-10. doi.org/10.1002/mop.33918.

Chio, S.-H., 2022: An Investigation on a Plane-Based Dynamic Calibration Method for the Handheld LiDAR Scanner. *Sensors* 22(1), 369. doi.org/10.3390/s22010369.

Dirnböck, T., Kraus, D., Grote, R., Klatt, S., Kobler, J., Schindlbacher, A., Seidl, R., Thom, D., Kiese, R., 2020: Substantial understory contribution to the C sink of a European temperate mountain forest landscape. *Landscape Ecology* 35(2), 483-499. doi.org/10.1007/s10980-019-00960-2.

Franklin, S.E., 2020: Interpretation and use of geomorphometry in remote sensing: a guide and review of integrated applications. *International Journal of Remote Sensing* 41(19), 7700-7733. doi.org/10.1080/01431161.2020.1792577.

Gaffney, K.P., Prammer, M., Brasfield, L., Hipp, D.R., Kennedy, D., Patel, J.M., 2022: SQLite: Past, Present, and Future. *Proceedings of the VLDB Endowment* 15(12), 3535-3547. doi.org/10.14778/3554821.3554842.

Gopalakrishnan, R., Korhonen, L., Maltamo, M., Adnan, S., Packalen, P., 2025: Prediction and mapping of boreal forest fire fuel loads using high-resolution satellite stereo imagery. *International Journal of Remote Sensing* 46(21), 8028-8050. doi.org/10.1080/01431161.2025.2562006.

Guenther, M., Heenkenda, M.K., Leblon, B., Morris, D., Freeburn, J., 2024a: Estimating tree diameter at breast height (DBH) using iPad pro LiDAR sensor in boreal forests. *Canadian Journal of Remote Sensing* 50(1), 2295470. doi.org/10.1080/07038992.2023.2295470.

Guenther, M., Heenkenda, M.K., Morris, D., Leblon, B., 2024b: Tree diameter at breast height (DBH) estimation using an iPad pro LiDAR scanner: a case study in boreal forest, Ontario Canada. *Forests* 15(1), 214. doi.org/10.3390/f15010214.

Laspy Developers, 2025. laspy: Python library for lidar LAS/LAZ IO. GitHub. github.com/laspy/laspy (14 November 2025).

Lee, T., Bettinger, P., Merry, K., Cieszewski, C., 2023: The effects of nearby trees on the positional accuracy of GNSS receivers in a forest environment. *PLoS ONE* 18(3): e0283090. doi.org/10.1371/journal.pone.0283090.

McKinney, W., 2010. Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 51–56.

Minasny, B., Berglund, Ö., Connolly, J., Hedley, C., de Vries, F., Gimona, A., Kempen, B., Kidd, D., Lilja, H., Malone, B., McBratney, A., Roudier, P., O'Rourke, S., Rudiyanto, Padarian, J., Poggio, L., ten Caten, A., Thompson, D., Tuve, C., Widyatmanti, W., 2019: Digital mapping of peatlands – A critical review. *Earth-Science Reviews* 196, 102870. doi.org/10.1016/j.earscirev.2019.05.014.

Räty, J., Varvia, P., Korhonen, L., Savolainen, P., Maltamo, M., Packalen, P., 2022: A Comparison of Linear-Mode and Single-Photon Airborne LiDAR in Species-Specific Forest Inventories. *IEEE Transactions on Geoscience and Remote Sensing* 60, 1-14. doi.org/10.1109/TGRS.2021.3060670.

Shin, J.-I., Seo, W.-W., Kim, T., Park, J., Woo, C.-S., 2019: Using UAV Multispectral Images for Classification of Forest Burn Severity—A Case Study of the 2019 Gangneung Forest Fire. *Forests* 10(11), 1025. doi.org/10.3390/f10111025.

Takeshige, R., Htoo, K.K., Onishi, M., Rahman, F.M., Hoshizaki, K., Ida, H., Ishihara, M.I., Itoh, A., Kaneko, T., Katayama, A., Kuramoto, S., Kurokawa, H., Maki, M., Masaka, K., Nakaji, T., Nakamura, M., Nishimura, N., Noguchi, M., Sakai, A., Takashima, A., Tashiro, N., Tokuchi, N., Yamagawa, H., Onoda, Y., 2025: High-resolution digital canopy height models, terrain models, ortho-mosaic photos, and canopy tree crown shapes derived from UAV-borne LiDAR at 22 tree census plots across Japanese natural forests. *Ecological Research* 40(4), 657-670. doi.org/10.1111/1440-1703.12555.

Tatsumi, S., Yamaguchi, K., Furuya, N., 2023: ForestScanner: A mobile application for measuring and mapping trees with LiDAR-equipped iPhone and iPad. *Methods in Ecology and Evolution* 14(7), 1603-1609. doi.org/10.1111/2041-210X.13900.

Teate, R.M., 2021: SQL for Data Scientists: A Beginner's Guide for Building Datasets for Analysis. John Wiley & Sons, Inc. doi.org/10.1002/9781119669388.

Zeng, T., Li, C.-X., Liu, Q.-H., Chen, X.-L., 2014: Tracking with nonlinear measurement model by coordinate rotation transformation. *Science China Technological Sciences* 57(12), 2396-2406. doi.org/10.1007/s11431-014-5694-y.

Zhang, T., Liu, Y., Li, J., 2022: Wheat Canopy Cover Estimation by Optimized Random Forest and UAV Multispectral imagery. *Journal of Physics: Conference Series* 2203(1), 12013. doi.org/10.1088/1742-6596/2203/1/012013.