International Cartographic Association
Association Cartographique Internationale

HexDTM - A method of thinning LiDAR point clouds using cartographic thematic data classification

Adam Inglot [a*], Paulo Raposo [b], Kristian Kozioł [c]

[a] *Faculty of Civil and Environmental Engineering, Gdańsk University of Technology - adam.inglot@pg.edu.pls,* [b] *Faculty ITC, University of Twente, - p.raposo@utwente.nl,* [c] *Department of Integrated Geodesy and Cartography, AGH University of Science and Technology - krystian.koziol@agh.edu.pl*

* Corresponding author

Keywords: LiDAR, classification, generalization

Abstract:

LiDAR is commonly used as the raw data source when high resolution digital terrain or digital surface models (DTMs, DSM) are generated. Depending on sensor specifications and height above ground, aerial LiDAR surveys produce point cloud data sets at varying surface densities. The density of a given LiDAR data set is often significantly higher than is necessary for the modeling purpose, and thus methods that thin or generalize a point cloud data set to a lower number of points—ideally without losing sampling precision—are beneficial in terms of storage, computation, and time savings.

We present a new method of thinning point clouds meant for use on aerial LiDAR .las or .laz files toward production of raster DTM models. Called HexDTM, our method discretizes the point cloud into a grid of hexagons or squares, where the cell width is determined in proportion to the digital elevation model (DEM) square pixel width the user intends to generate. Inside each cell, we consider only pre-classified ground points, and use a series of statistical data classification methods commonly used in thematic cartography to collapse these points into a smaller number than those present in the input data file. The data classification method chosen in each hexagon is determined by comparing tabular accuracy index (TAI) values (Jenks and Caspall, 1971, Armstrong et al., 2003) for each classification method and selecting the highest-scoring one.

The user provides a number of classes n (e.g., 5). Provided there are $n+1$ or more input ground points in the cell, the algorithm proceeds (it otherwise simply returns all ground points in the cell, un-thinned). Within each hexagon the ground-classed points are classified into n classes by their z values using each of the following methods:

- arithmetic series, ascending;
- arithmetic series, descending;
- equal interval;
- geometric series, ascending;
- geometric series, descending;
- Jenks natural breaks;
- quantiles.

TAI is calculated for each classification method. This coefficient, with values between 0.0 and 1.0, is a normalized disruption rate in the classification of numerical data (Carrão et al., 2014), given by

$$TAI = 1 - \frac{\sum_{j=1}^{k}\sum_{i=1}^{N}|z_{ij} - \overline{z}_j|}{\sum_{i=1}^{N}|z_i - \overline{z}|} \tag{1}$$

where

z_{ij} is the value of the next height within the class k,
z_i is the value of the next height,
$\overline{z}_j$ is the arithmetic mean of the height value in the jth grade,
$\overline{z}$ is the arithmetic mean of all values,
k is number of classes, and

https://doi.org/10.5194/ica-abs-7-61-2024

N_j is size of the class.

In each cell, the highest TAI is taken to indicate the best method of class division among those; i.e., the one we take to be the best representative of the terrain's z value distribution. The classes produced by this method progress to the next phase, where either a single or a few collapsed output points are determined for the cell, using one of four definitions:

1. The single artificial point at the mean (x,y,z) coordinate for the class with the most points;
2. The single input point closest to the mean (x,y,z) coordinate for the class with the most points;
3. One point in each of the n classes, at the mean (x,y,z) coordinate for each class; and
4. One point in each of the n classes, closest to the mean (x,y,z) coordinate for each class.

Figure 1 illustrates the overall process.

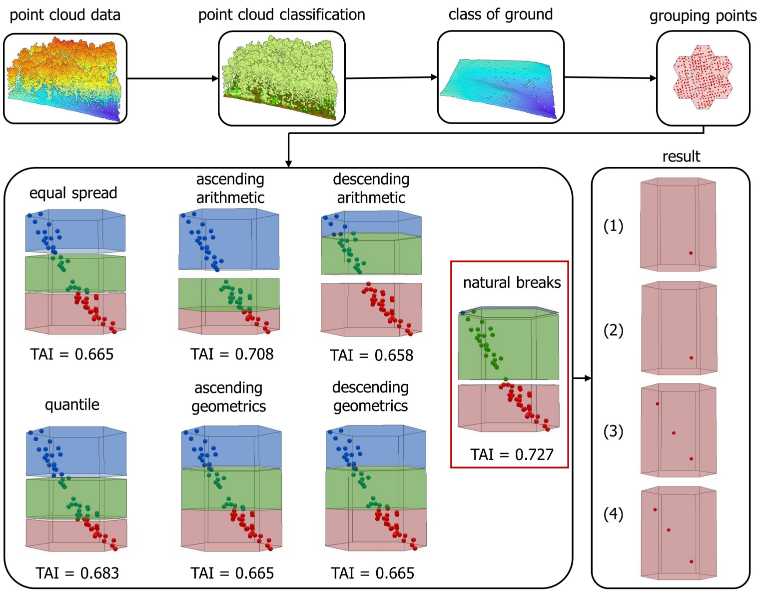

Figure 1. The functioning of the HexDTM algorithm. The example illustrated classifies the points into 3 classes (drawn in blue, green and pink). Point height z values have been exaggerated for illustration. The red box indicates that one classification method, natural breaks, has been chosen because it has the highest TAI value. The result box illustrates the four ways we have implemented the selection of the final output points.

The algorithm has been implemented in Python code, and is being evaluated across hexagon and square cells, various cell widths and class counts, and across the four above-described point selection methods, all applied to four input LiDAR datasets representing various terrain types. Results of these evaluations will be presented at the EuroCarto 2024 conference.

Acknowledgements

Adam Inglot acknowledges support from the Gdańsk University of Technology through the DEC-6/2021/IDUB/II.1b/AMERICIUM grant under the AMERICIUM — "Excellence Initiative - Research University" program.

References

Armstrong, M. P., Xiao, N. and Bennett, D. A., 2003. Using Genetic Algorithms to Create Multicriteria Class Intervals for Choropleth Maps. *Annals of the Association of American Geographers* 93(3), pp. 595–623.

Carrão, H., Singleton, A., Naumann, G., Barbosa, P. and Vogt, J. V., 2014. An optimized system for the classification of meteorological drought intensity with applications in drought frequency analysis. *Journal of Applied Meteorology and Climatology* 53(8), pp. 1943–1960.

Jenks, G. F. and Caspall, F. C., 1971. Error on Choroplethic Maps: Definition, Measurement, Reduction. *Annals of the Association of American Geographers* 61(2), pp. 217–244.

https://doi.org/10.5194/ica-abs-7-61-2024 |