Influence of tree density and terrain slope on ground point density in LiDAR point clouds: a simulation-based study with Helios++

Juan José González-Quiñones [1], Laurent Polidori [2], Francisco Javier Ariza-López [3], Manuel Antonio Ureña-Cámara [4], Juan Francisco Reinoso-Gordo [5]

[1, 5] Department of Architectural and Engineering Graphic Expression. University of Granada Spain - juanjosegonzalez@ugr.es, jreinoso@ugr.es
[2] Universidade Federal do Pará, Instituto de Geociências, Programa de Pós-Graduação em Geologia e Geoquímica, Belém - Pará, Brazil. - laurent.polidori@ufpa.br
[3, 4] Department of Cartographic, Geodesic and Photogrammetry Eng. University of Jaén. Spain. - fjariza@ujaen.es, maurena@ujaen.es

Keywords: LiDAR, Helios++, simulations, canopy, DTM, UAV.

Abstract

There is still no reliable solution for obtaining Digital Terrain Models (DTM) under the canopy in areas with difficult access on foot. The difficulty lies in the configuration of the environment. This configuration does not allow access on foot to all areas of the terrain. In addition, the dense leaf litter complicates any photogrammetry technique. Currently, LiDAR is a technique that could help solve this problem. The best solution to this problem requires many samples over hostile environment and LiDAR flight simulation can be a good option to approximate the problem. For this reason, this research carries out a series of LiDAR flight simulations, using Helios++, in different types of scenes under the canopy with the aim of analysing the influence of the density of trees and the mean terrain slope over the number of points that bounce off the terrain. Additionally, this research enables the scientific community to approach the problem and guide future research based on the obtained results.

1. Introduction

Obtaining Digital Terrain Models (DTM) under the canopy in forested areas is a challenging issue. Indeed, the leaf density prevents the use of traditional techniques like aerial photogrammetry. The only two techniques that could provide some suitable results are long wavelength synthetic aperture radar (SAR) and airborne LiDAR.

Synthetic aperture radar has been applied to digital elevation modelling using different methods, the most promising one being interferometry, implemented in well-known missions like SRTM. (Farr et al. 2007) However, most radar imaging systems use X- or C- band radar, i.e., short wavelength which cannot penetrate through the forest volume, so that the resulting elevation is at the canopy level and does not provide a DTM. Longer wavelengths have been tested. Indeed, L-band (15-30 cm) and P-band (30-100 cm) wavelengths have a deeper penetration into the canopy and can reach the ground. Wide areas have been surveyed by airborne P-band radar in the Brazilian Amazon (Correia 2011), and the upcoming ESA Biomass mission will use a P-band radar onboard a satellite for the first time for the mapping of above-ground biomass and so-called “secondary products”, including a quasi-global DTM even in forested áreas (Le Toan et al. 2011).

Airborne LiDAR is recognized as an efficient way to obtain a DTM under canopy, due to the fact that part of the laser pulses reach the ground through openings in the foliage, which is sufficient to build a lower envelope by interpolating the lowest points and to create a DTM (Kraus y Pfeifer, 1998). Liu et al. (2015) studied the error of the DTM obtained from a theoretical and practical point of view. Sterenczak et al. (2013) studied the accuracy of DTM in pine forests with a pixel size of 0.5, 1, 2, 3, 4, 5 and 10 meters, obtaining a minimum error of 0.02 cm for the DTM with a pixel size of 0.5 m and a maximum error of 0.23 cm for the DTM with a cell size of 10 m.

Following the previous studies, a common feeling is that Lidar technology seems to be able to help solve the problem, but its use needs to be further investigated. The main difficulties to bring about these types of studies are the cost of the systems, the difficulty in validating the data in situ and the need for segmentation (terrain or tree) of the points obtained in situ. Thus, through LiDAR flight simulations with the Helios++ development (Winiwarter et al., 2022) could reduce monetary and time costs, offering a greater capacity to segment the points on the terrain.

Tree density and terrain slope are two variables to be taken into account when developing DTM from point clouds obtained by aerial LiDAR under the canopy; it is known the research by (Radhie et al., 2015) highlights the influence of tree density and terrain slope on the acquisition of Digital Terrain Models (DTM) in a forested area near an urban zone. Also, the study by Cățeanu and Ciubotaru (2021) concludes that one of the parameters in the reliability of DTMs is the density of points in the point cloud from which the DTM has been created; so that a DTM of the same area is more reliable if it is made up of more points that belong to the terrain than another that is made up of fewer (Ariza-López et al., 2018).

So, the objective of our research is to analyse, through a series of LiDAR simulations using Helios++, the influence of tree density and terrain slope on the number of terrain points per hectare obtained.

2. Material and method

The hardware employed in this research included an Intel i9-12900KF processor, an NVIDIA RTK 3090 graphics card, and 64 GB of RAM. For software, the tools used were QGIS, Cloud Compare, and Helios ++. QGIS was utilized for handling the DTMs, Cloud Compare was employed for processing and visualizing the point clouds, and Helios++ was used to simulate aerial LiDAR scans.

The workflow of this research (Figure 1) has four steps: i) experiment design, ii) data preparation for the generation of synthetic forest scenes, iii) generation of the synthetic forest scenes as input to Helios++, iv) run the simulations, v) analysis of results.

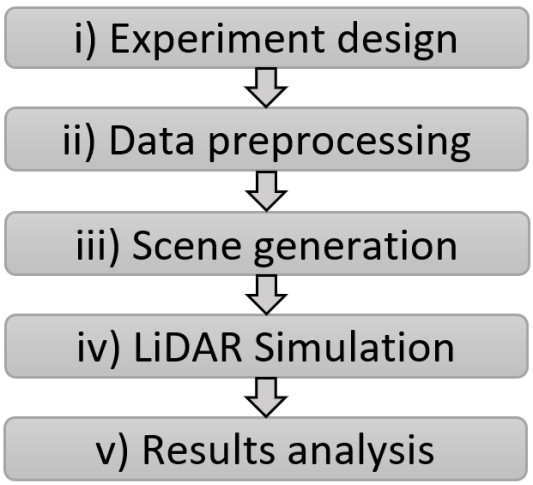


Figure 1. Research workflow

In the experiment design i), two explanatory variables and one response variable were used. The two explanatory variables are the tree density and the average slope of the terrain. The response variable was the number of points that reflect off the terrain per hectare.

The explanatory variable called tree density has three density levels. For this, a high density (Hd) of 700 trees/ha has been assumed using the equation introduced by Gadow and Kotze (2014). Then, a medium density (Md) of 500 trees/ha and a low density (Ld) of 300 trees/ha have been established.

Concerning the explanatory variable called average terrain slope, three DTMs have been taken, matching each level according (Reinoso-Gordo, 2010). Thus, the terrain that has been classified as high slope (Hs) has an average slope of 29.5°, the terrain with a medium slope (Ms) has an average slope of 16.6° and the terrain with a low slope (Ls) has an average slope of 3.5°. To calculate the average terrain slope, a slope raster was initially generated using the Horn algorithm (1981) integrated within QGIS. Subsequently, the mean slope was determined by averaging the values of all pixels within the raster.

All possible combinations have been made with the two explanatory variables and each combination has been repeated three times to ensure redundancy in the experiment (Table 1). Thus, 3 average terrain slopes x 3 tree densities x 3 samples for each combination = 27 simulations have been carried out.

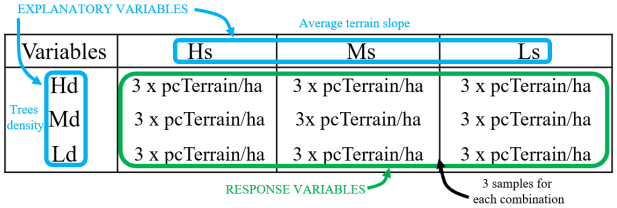


Table 1. Experiment design

The data preparation step ii) aims at generating a synthetic forest scene. The scene is the virtual landscape and it is composed of several trees and one terrain.

The tree database made available to the scientific community by Weiser et al., in (2022) has been used. The database contains point clouds of 1491 trees of 22 different tree species from aerial and terrestrial LiDAR scans. The point cloud database is referenced in ETRS89 / UTM 32 N coordinates; EPSG:25832, ellipsoidal height, GRS80. Among all the tree species contained in the database, *Pinus sylvestris* has been chosen due to its presence in numerous forests in the northern hemisphere. The database contains 158 scanned *Pinus sylvestris*. Each *Pinus sylvestris* is composed of several point clouds from aerial and terrestrial scans. So, it has been necessary to merge the aerial and terrestrial scans of each *Pinus sylvestris* into a single point cloud, thus obtaining 158 point clouds; one for each *Pinus sylvestris*. Then, the reference system of the point cloud of each tree has been converted into a local reference system in which Z = 0 m is the Z of the lowest point and X = 0 m and Y = 0 m is the geometric centre of the points whose Z ranges between 1 and 10 meters, which is an approximation of the trunk axis, since most of these points belong to the trunk (Figure 2). Once a point cloud has been obtained for each of the 158 *Pinus sylvestris* in a previously defined local reference system, it has been necessary to down sample each point cloud so that the RAM does not saturate during the simulation process.

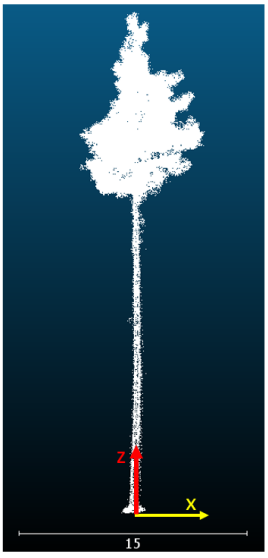


Figure 2. Example of a point cloud of a *Pinus sylvestris* in its local reference system

Regarding the terrain, the database of Instituto Geográfico Nacional of Spain has been used. From this database, several sheets of the Granada area have been taken with a resolution of 2 m x 2 m. Then, these sheets have been cut in such a way that 3 DTMs have been obtained with a spatial dimension of 200 m x 200 m each and with the average slope equal to that shown in the first step, i) experiment design, of this workflow.

The next step iii) is the generation of synthetic forest scenes as input for Helios++. The script scene-writer.py from the pyhelios package, made available by the Helios++ developers on their GitHub, has been used to generate a file in .xml format. This file defines the scene that Helios++ uses as input to carry out the simulation. The previous script has been integrated into another script called random-forest.py, developed by the authors of this research, for the creation of synthetic forest scenes for Helios++ (Figure 3). Specifically, random-forest.py receives as input a DTM in .tif format, and a forest density value (trees/hectare) and generates a scene file in .xml format ready to simulate. The .xml file generated contains a synthetic forest created by randomly selecting n coordinates from the TIFF file (based on the input density) and placing a different *Pinus sylvestris* at each coordinate, randomly chosen from one of the 158 *Pinus sylvestris* prepared in step ii). Taking into account the experiment design, it was necessary to generate 27 different scenes.

As regards the simulations iv), 27 have been carried out; one for each scene. All simulations have been run with the Livox Mid-70, at a constant height of 100 m above the average altitude of the terrain, at a constant speed of 5 m/s and using two straight and parallel flight paths with a separation of 50 m from the edge of the ground and a separation of 100 m between them.

The v) results were analysed using both quantitative and qualitative techniques. The quantitative analysis was carried out using an ANOVA and the qualitative analysis was carried out on the basis of a graph showing the position of the trees and the scanned points belonging to the terrain.

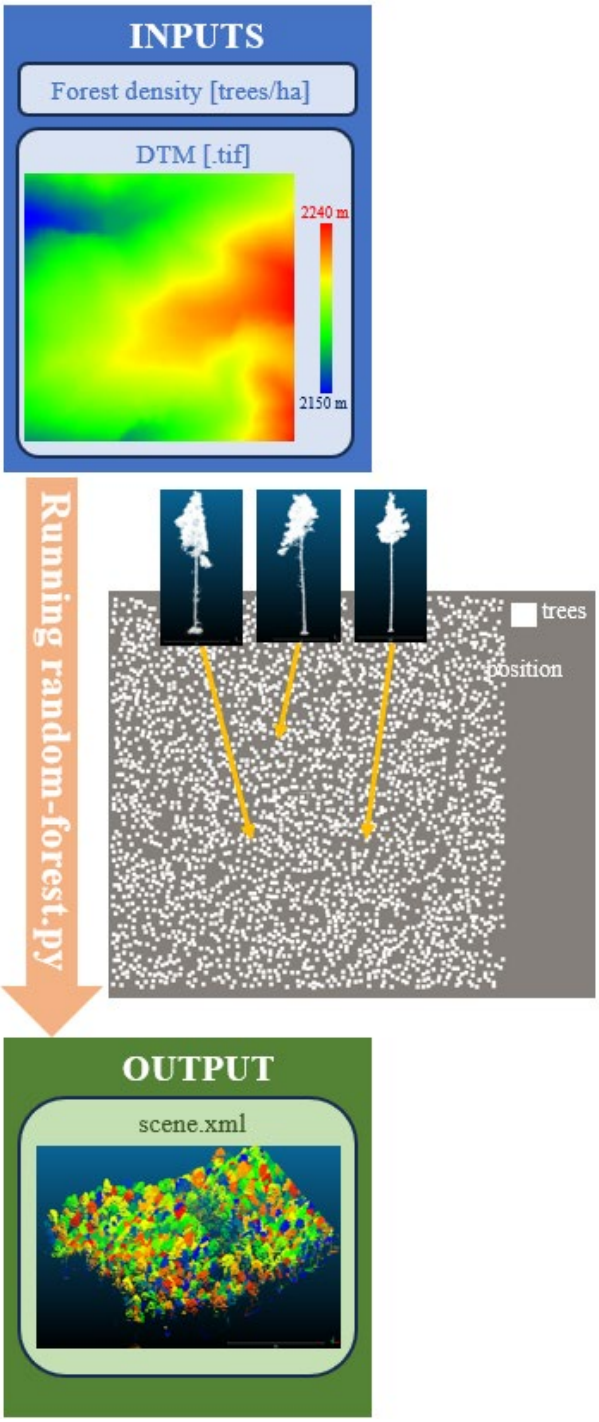


Figure 3. random-forest.py workflow

3. Results and discussion

The number of points resulting from the 27 simulations can be seen in Table 2, where each column indicates the total number of points (pcTotal), the number of points reaching the terrain (pcTerrain) and the number of points per hectare reaching the terrain (pcTerrain/ha) and where each row represents a combination of slope levels (high (sH), medium (sH) and low (sH)), density levels (high (dH), medium (dM) and low (dL)) and sample number (sam1, sam2, sam3).

The identification the points that have bounced off the terrain (pcTerrain) is immediate due to Helios++ assigns to each point the object in the scene that the laser bounced off to obtain it during the simulation process. In this way, it has been possible to segment the points belonging to the terrain from those belonging to the trees (Figure 4). This fact shows the segmentation potential of the tool used for the simulation: Helios++.

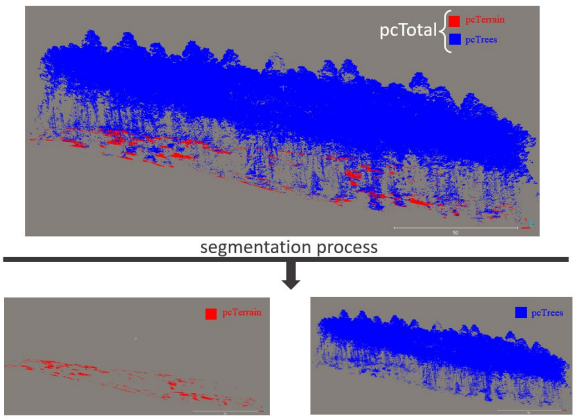


Figure 4. Segmentation process

Samples	pcTotal	pcTerrain	pcTerrain/ha
sH_dH_sam1	6732685	9183	2295.8
sH_dH_sam2	6726751	8220	2055
sH_dH_sam3	6747084	7325	1831.3
sH_dM_sam1	6645903	58776	14694
sH_dM_sam2	6638915	80031	20008
sH_dM_sam3	6613647	71785	17946
sH_dL_sam1	6423147	443215	110804
sH_dL_sam2	6466156	418259	104565
sH_dL_sam3	6469154	444203	111051
sM_dH_sam1	6896444	6239	1559.8
sM_dH_sam2	6885874	10523	2630,8
sM_dH_sam3	6917612	4500	1125
sM_dM_sam1	6786705	44014	11004
sM_dM_sam2	6750855	72239	18060
sM_dM_sam3	6785567	65874	16469
sM_dL_sam1	6529069	390888	97722
sM_dL_sam2	6538566	439251	109813
sM_dL_sam3	6556312	432498	108125
sL_dH_sam1	7060675	5012	1253
sL_dH_sam2	7062987	5019	1254.8
sL_dH_sam3	7087790	3223	805.75
sL_dM_sam1	6904974	65283	16321
sL_dM_sam2	6923814	44731	11183
sL_dM_sam3	6943194	55298	13825
sL_dL_sam1	6620911	337124	84281
sL_dL_sam2	6629302	382304	95576
sL_dL_sam3	6630984	375568	93892

Table 2. Simulation results number of points reaching the field in different slope-density scenes

An ANOVA analysis was performed to determine if there were differences between the means of the number of points that reached the terrain. To do this, the pcTerrain column was used, which resulted in a Two-way ANOVA balanced analysis with 3 replicates for a significance level of α = 5%. Table 3 shows the result of the ANOVA analysis, where each column follows the Neter et al., (1996) notation, i.e.:
- Sum sq. column: the sum of squares due to each source).
- df: degree of freedom.
- Mean Sq.: (Sum sq.)/df.
- F: F-statistic, which is the ratio of the mean squares.
- Prob>F: the p-value, is is the probability that the F- statistic can take a value larger than the computed test-statistic value.

Source	Sum Sq.	df	Mean Sq	F	p-value Prob>F
slope (s)	4.22e9	2	2.11e9	9.77	0.0013
dens (d)	8.47e11	2	4.23e11	1958.2	0
s-d (interac.)	4.40e9	4	1.10e9	5.09	0.0064
Error	3.89e9	18	2.16e8		
Total	8.59e11	26			

Table 3. Two-way ANOVA test: slope and density factors with three-levels partitioning

In Table 3, it can be deduced from the p-values that the different levels of both the slope and the density produce significantly different amounts of points on the terrain for a significance level of 5%. The difference that occurs due to the density factor was somewhat predictable, since the greater the number of trees, the greater the number of points that will collide with the trees and will not be able to reach the terrain. However, the slope of the terrain was, a priori, an unknown and needed a simulation study to estimate its behavior. Also, there is an interaction between the 9 combinations resulting from the different levels of the factors, which gives rise to a difference in the number of points that reach the terrain according to these 9 combinations.

Table 4 represents the influence of the slope level on pcTerrain and allows us to visually appreciate that the Low level has a mean pcTerrain value significantly different from the High and Medium levels. The circle indicates the mean value of the factor and the endpoints of each line show the limits of the corresponding confidence interval.

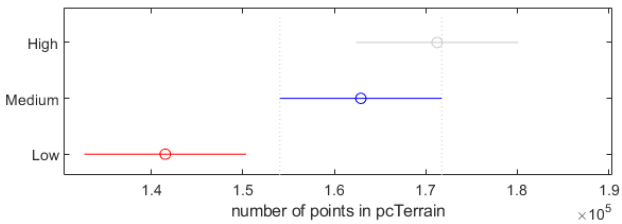


Figure 5. Significative differences in the number of points in pcTerrain taking account the slope levels

The representation of Figure 5 in terms of p-value can be seen in Table 4 which corroborates the graphical representation.

Slope Level A	Slope Level A	p-value
High	Medium	0.4676
High	Low	0.0012
Medium	Low	0.0168

Table 4. Slope p-value for High, medium and low levels

The interaction between the 9 combinations of the density and slope factor levels can be analyzed from the observation of Figure 6.

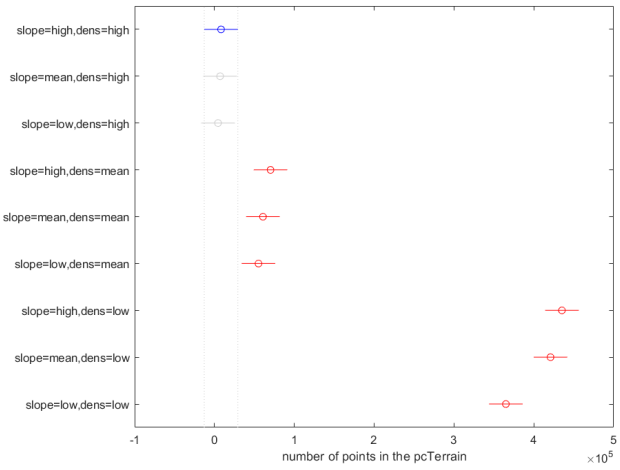
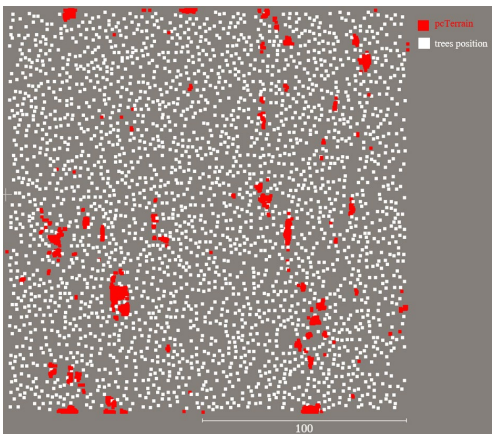
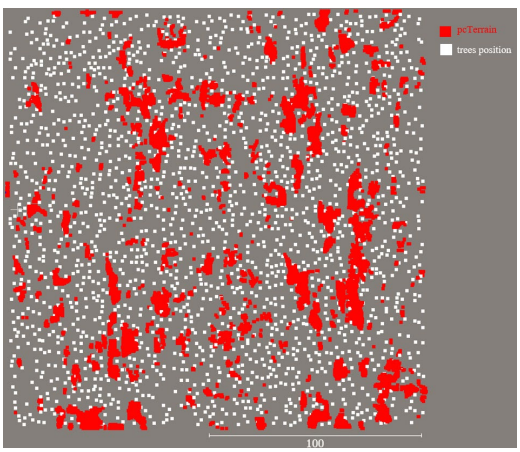


Figure 6. Significative differences in the number of points in pcTerrain taking account the levels combination from slope and density

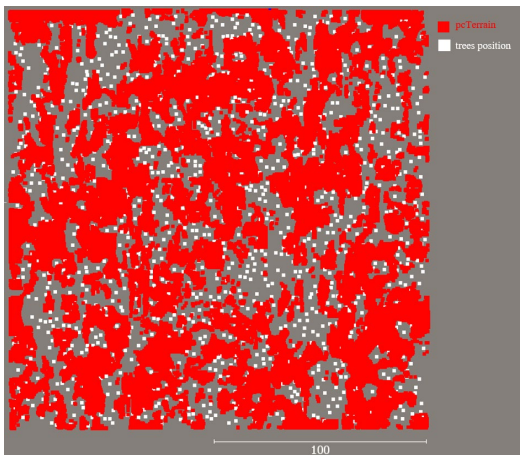
Figure 6 shows what was already anticipated about the difference in mean value of pcTerrain depending on the density levels, while in slope, only at the Low level it seems to produce a significantly different value in pcTerrain. These results seem to indicate that a more in-depth study should be made on the influence of the slope on pcTerrain, particularly to find the limits of the confidence interval in which the difference is significant.



a) Plan view of the pcTerrain and trees position for the Medium slope and High density simulation, sample 1



b) Plan view of the pcTerrain and trees position for the Medium slope and Medium density simulation, sample 1



c) Plan view of the pcTerrain and trees position for the Medium slope and Low density simulation, sample 1

Figure 7. pcTerrain (point belonging to the terrain, in red) from simulation case and trees position (in white)

Another result that is considered relevant to discuss is the geographical distribution of the points that bounce off the terrain and their relationship with the position of the trees in the forest. To this end, Figure 7 shows in red the point cloud resulting from the points that bounce off the terrain and in white the positions of the trees that make up the forest. In addition, the result of a sample of High density (Figure 7.a.), Medium density (Figure 7.b) and Low density (Figure 7.c.) is shown.

It can be observed that the points belonging to the terrain are grouped in some positions, forming kinds of spots. In addition, it can be observed that these groupings occupy the positions where there are no trees, or at least, in the case of a continuous canopy, in the areas furthest from the trunks, which are less dense and therefore are crossed by more laser pulses. Based on the above, it seems reasonable that an optimal flight path could be drawn such that it passes through the maximum possible distance from all trees, in order to get holes in the canopy that allows to reach the terrain ground; in addition, said path could have numerous vertices and the use of a multirotor would be necessary since it is a UAV that allows for quick changes of direction and sense instead of a fixed wing airplane. Undoubtedly this is a clear change in flight strategy compared to the usual, which is a sweep along parallel tracks. This change requires some prior knowledge of the configuration of the tree mass on which the lidar capture is going to be carried out.

4. Conclusions

As expected at the beginning of this study, tree density significantly influences the number of points obtained on the ground after a LiDAR flight.

Regarding the mean slope of the terrain, the greater the mean terrain slope, the more the laser bounces off the terrain. In addition, the ANOVA analysis has shown that there are significant differences between the Low slope with respect to the other two slopes and has also shown that there are no significant differences between the High and Medium slopes of the terrain. This indicates that further analysis is needed.

Points belonging to the terrain as a result of the simulation are arranged in spots and are in the areas furthest from the trees.

This fact may lead to the idea that there may be an optimal flight path with numerous vertices and/or changes of direction that could be achieved with a multirotor UAV. This is proposed for future research.

The LiDAR simulation tool called Helios++ is useful for performing this type of analysis due to its versatility when interpreting scenes composed by objects from different nature and because it allows segmentation work based on the bounce of the laser with each object defined in the scene.

5. Acknowledgements

Grant PID2022-138835NB-I00 funded by the Spanish MICIU/AEI/10.13039/501100011033 and by “ERDF/EU".

6. References

Ariza-López, F.J., Chicaiza-Mora, E.G., Mesa-Mingorance, J.L., Cai, J., Reinoso-Gordo, J.F., 2018. DEMs: An approach to users and uses from the quality perspective. *International Journal of Spatial Data Infrastructures Research* 13, 131-171. doi.org/10.2902/1725-0463.2018.13.art12

Cățeanu, M., Ciubotaru, A., 2021. The Effect of LiDAR Sampling Density on DTM Accuracy for Areas with Heavy Forest Cover. *Forests* 12, 265. doi.org/10.3390/F12030265.

Correia, A.H., 2011. Metodologias e resultados preliminares do projeto Radiografia da Amazônia. *Anais XV Simp. Brasileiro de Sensoriamento Remoto*, INPE, Curitiba, p.8083-8090 (http://marte.sid.inpe.br/col/dpi.inpe.br/marte/2011/06.27.19.46/doc/p1032.pdf). Accessed on 19 July 2024

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The Shuttle Radar Topography Mission. Review of Geophysics, 45, RG2004.

Gadow, K. V., Kotze, H., 2014. Tree survival and maximum density of planted forests – Observations from South African spacing studies. *Forest Ecosystems* 1, 1-9. doi.org/10.1186/S40663-014-0021-4

Hojo, A., Takagi, K., Avtar, R., Tadono, T., Nakamura, F., 2020. Synthesis of L-Band SAR and Forest Heights Derived from TanDEM-X DEM and 3 Digital Terrain Models for Biomass Mapping. *Remote Sensing* 12, 349. doi.org/10.3390/RS12030349

Horn, B.K.P., 1981. Hill Shading and the Reflectance Map. *Proceedings of the IEEE* 69, 14-47. doi.org/10.1109/PROC.1981.11918

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and remote Sensing* 53, 193-203.

Le Toan,T., Quegan, S., Davidson, M.W.J., Balzter, H., Paillou, P., Plummer, S., Rocca, F., Saatchi, S., Shugart, H., Ulander, L., 2011. The BIOMASS mission: mapping global forest biomass to better understand the terrestrial carbon cycle. *Remote Sensing of Environment,* 115: 2850-2860.

Liu, X.H., Hu, H., Hu, P., 2015. Accuracy Assessment of LiDAR-Derived Digital Elevation Models Based on Approximation Theory. *Remote Sensing* 7, 7062-7079. doi.org/10.3390/RS70607062

Neter, J., Kutner, M., Nachtsheim, C., Wasserman, W., 1996. Applied linear statistical models.

Radhie, M., Salleh, M., Ismail, Z., Zulkarnain, M., Rahman, A., 2015. Accuracy assessment of lidar-derived digital terrain model (DTM) with different slope and canopy cover in tropical forest region. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-2/W2. doi.org/10.5194/isprsannals-II-2-W2-183-2015

Reinoso-Gordo, J.F., 2010. A priori horizontal displacement (HD) estimation of hydrological features when versioned DEMs are used. *Journal of Hydrology* 384, 130-141. doi.org/10.1016/J.JHYDROL.2010.01.017

Sterenczak, K., Zasada, M., Forestry, M.B., 2013. The accuracy assessment of DTM generated from LIDAR data for forest area–a case study for scots pine stands in Poland. *Baltic Forestry*.

Tokunaga, M., 1997. DTM Accuracy derived from interferometry SAR, en: ACRS.

Weiser, H., Schäfer, J., Winiwarter, L., Krašovec, N., Fassnacht, F.E., Höfle, B., 2022. Individual tree point clouds and tree measurements from multi-platform laser scanning in German forests. *Earth System Science Data* 14, 2989-3012. doi.org/10.5194/ESSD-14-2989-2022

Winiwarter, L., Esmorís Pena, A.M., Weiser, H., Anders, K., Martínez Sánchez, J., Searle, M., Höfle, B., 2022. Virtual laser scanning with HELIOS++: A novel take on ray tracing-based simulation of topographic full-waveform 3D laser scanning. *Remote Sensing of Environment* 269, 112772. doi.org/10.1016/J.RSE.2021.112772

Xu, K, Zhao, L, Chen, E, Li, K, Liu, D, Li, T, Li, Z, Fan, Y, Xu, Kunpeng, Zhao, Lei, Chen, Erxue, Li, Kun, Liu, Dacheng, Li, Tao, Li, Zengyuan, Fan, Yaxiong, 2022. Forest Height Estimation Approach Combining P-Band and X-Band Interferometric SAR Data. *Remote Sensing* 14, 3070. doi.org/10.3390/RS14133070