DIMENZIÓK
Matematikai Közlemények
XII. kötet, 2024 33–41

doi:10.20312/dim.2024.04

Individual tree detection and spatial distribution analysis without reference data[1]

Tóth Zsolt*
Soproni Egyetem, Faipari Mérnöki és Kreatívipari Kar, Alaptudományi Intézet
toth.zsolt@uni-sopron.hu, 0000-0003-0999-784X

Farkas Péter
Soproni Egyetem, Faipari Mérnöki és Kreatívipari Kar, Cziráki József Faanyagtudomány és Technológiák Doktori Iskola
farkaspeter@phd.uni-sopron.hu, 0009-0000-9150-8472

Novotni Adrienn
Soproni Egyetem, Faipari Mérnöki és Kreatívipari Kar, Cziráki József Faanyagtudomány és Technológiák Doktori Iskola
novotni.adrienn@uni-sopron.hu, 0000-0003-1451-0709

ÖSSZEFOGLALÓ.
Két 100x100 méteres területet, egy erdőt és egy vegyes ültetvényt vizsgáltunk LiDAR ponthamaz alapján. A Progressive Morphological Filtering (PMF) és a Local Maximum Filtering (LMF) módszerek 257 fát azonosítottak az erdőben (módszertől függően) részben véletlenszerű, részben szabályos, 47-et a vegyes területen klaszteres eloszlással. A sűrűségmérések és legközelebbi szomszédok távolságának azonosítása G- és K-statisztikával. Monte Carlo-módszerrel és kvadrátpróbával jelentős különbségeket mutattak a fák eloszlásában, s jól jelezték a területek jellege közötti eltéréseket. Az alkalmazott módszer alkalmas az erdőterületek anomáliáinak azonosítására is.

ABSTRACT. This study processed two 100x100m areas from LiDAR dataset: a forest and a mixed forest-plantation. Progressive Morphological Filtering (PMF) and Local Maximum Filtering (LMF) methods identified 257 trees in the forest and 47 in the mixed area, showing (depending on the method) partly random, partly regular spacing in the forest and clustering in mixed areas. The density assessments and nearest-neighbour evaluations with G statistic, K statistic, Monte Carlo method, and quadrat tests revealed a significant difference in tree distribution, highlighting the effectiveness of these methods for detecting spatial patterns in diverse forest environments, too.

1 Introduction

The accurate identification of individual trees from LiDAR (ALS) measurements is a critical issue in forestry geoinformatics and statistics. However, the applicability of the frequently used

[1] HUNGARIAN TITLE. Egyedi fák detektálása és térbeli eloszlásának elemzése referenciaadatok nélkül.
KULCSSZAVAK. LiDAR, fadetektálás, G-statisztika, K-statisztika, Monte Carlo-módszer, kvadrátpróba.
KEYWORDS. LiDAR, tree detection, G statistic, K statistic, Monte Carlo method, quadrat test.
* Corresponding author.

local maxima method [21] for tree identification is significantly influenced by the quality and preprocessing of data, as well as the type and structure of the forest [4]. The problem can be relatively simply stated: even the best and most widely adopted methods currently available are not sufficiently reliable, and often, results considered acceptable are not truly satisfactory [5]. The variable effectiveness of ITD (individual tree detection) methods also impacts the estimation of other metrics [9].

Several attempts have been made to address these issues. Rasterizing the point-based method can reduce the number of errors under certain conditions [17]. It can be satisfactory within a narrow scope but are limited in their applicability for surveying "unknown" forest areas, which is essentially the ultimate goal. Machine learning methods [12], Monte Carlo methods [1], principal component analysis [8], object-based labelling [15], and optimization methods [18] can predict and sometimes reduce the error magnitude for specific tree and forest types, but this does not necessarily imply that these methods can be generalized to most forest types. Resizing the TWS (tree window size) used in the analysis has led to more usable results in some forest types [13], but this also does not seem to be a generalizable method. Using the L function for refining detections appears to be a more general method [11], but the applied procedure is still not entirely independent of the expected and known results from field reference data. Of course, improving technical conditions, such as using high-density, close-range, multispectral LiDAR recordings [6], can also contribute to increasing the efficiency of the procedure.

In this article, an attempt is made to provide an example of how to outline a procedure based on the principles of nearest-neighbour distances without relying on precise reference data. Rather than focusing on the exact identification of individual trees, this method assesses the overall characteristics of the forest. This approach enables the detection and monitoring of specific "anomalies" within the forest using almost exclusively LiDAR imagery, thus facilitating rapid and efficient surveys.

2 Materials and methods

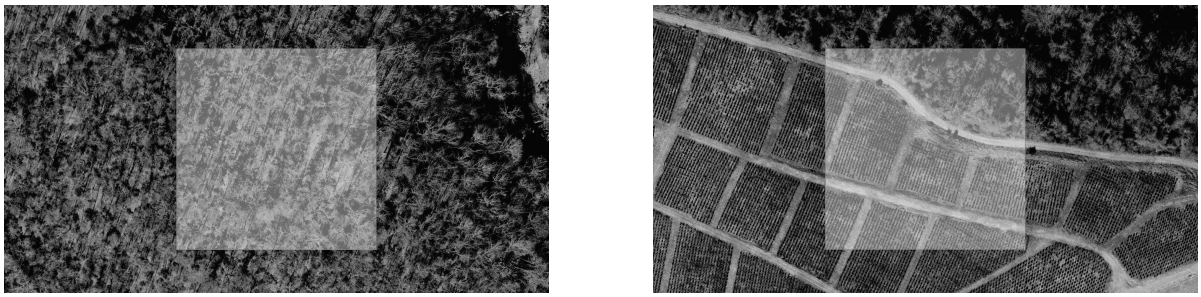
2.1 Data

The data described in Table 1. serve as the source for analysis. This LiDAR dataset was gathered by NCALM for Paula Figueiredo at North Carolina State University [14].

Features	Forest	Mixed vegetation
Horizontal Coordinates	WGS84 / UTM Zone 17N Metres [EPSG: 32617]	
Vertical Coordinates	Ellipsoid	
Number of Points	800743	528590
X_{min}, Y_{min}	490915, 4038800	491825, 4038310
X_{max}, Y_{max}	491015, 4038900	491925, 4038410

Table 1. Features of LiDAR datasets

The first LiDAR point cloud represents a purely forested area, while the second one depicts a forest patch and a plantation separated by a road. (Figure 1.) Both areas are 100 by 100 metres in size. Similar to common field conditions, we have a general overview of the area's characteristics, but we lack precisely surveyed reference data, such as the exact locations of the vegetation (trees) [20].



(a) Forest area (b) Mixed vegetation area

Figure 1. Vegetation areas

2.2 Canopy and tree detection

For ground classification, a Progressive Morphological Filter (PMF) was utilized. The original raster-based method [22] was modified by the developers of the R `lidR` package [19] used for data processing transforming it into a point-based approach [16]. During this process, the default values (window size = 3, threshold = 5) were used.

After ground classification, the digital terrain model (DTM) was created using the kriging method, which is more accurate but more resource-intensive than the commonly used triangulation method. Kriging is the most sophisticated approach, employing advanced geostatistical interpolation techniques that consider the spatial relationships and distances between the returns. Height normalization was ensured using point cloud-based normalization method. This model is superior in terms of computational accuracy by normalizing with a continuous terrain instead of a discretized terrain.

In the next step, the canopy height model was created using a point-to-raster method.

From this model, the coordinates of individual trees, along with their corresponding height values, were computed using a Local Maximum Filter (LMF) with window size = 5. During each step, various other methods were considered, but upon testing, these methods yielded essentially identical results [20].

2.3 Spatial statistical analysis

The density-intensity of tree locations was visually analysed. Subsequently, the nearest-neighbour distances and, more generally, the pairwise distances within each sample plot were analysed as follows [2].

The cumulative distribution function of the nearest-neighbour distances for a regular point in the point (tree) pattern in a stationary point process $\mathbb{X}$ is

$$G(r) = \mathbb{P}\left(d(u, \mathbb{X} \setminus \{u\}) \leq r \,|\, u \in \mathbb{X}\right),$$

where u is a random location, and $d(u, \mathbb{X} \setminus \{u\})$ is the shortest distance between u and the points of the $\mathbb{X}$ pattern, excluding u.

The observed distribution function for the distances to the measured nearest neighbours is

$$G^*(r) = \frac{1}{n(\mathbf{x})} \sum_i \mathbf{1}\{t_i \leq r\}.$$

Edge corrections based on the empirical cumulative distribution function is

$$\hat{G}(r) = \sum_i e(x_i, r)\mathbf{1}\{t_i \leq r\}. \tag{1}$$

The $e(x_i, r)$ edge correction weight in the above equation ensures the approximate unbiasedness of $\hat{G}(r)$.

The distribution function of the nearest-neighbour distances in a homogeneous Poisson point process with intensity λ is

$$G_{\text{pois}}(r) = 1 - \exp(-\lambda\pi r^2). \tag{2}$$

In the case where $\hat{G}(r) > G_{\text{pois}}(r)$, the pattern is considered clustered, while when $\hat{G}(r) < G_{\text{pois}}(r)$, the pattern is classified as regular. In addition to the theoretical Poisson distribution, the Hanisch estimate [7], the border-corrected estimate, and the Kaplan-Meier estimate [10] for $\hat{G}(r)$ are used.

The biased patterns of $s_{ij} = \|x_i - x_j\|$ pairwise distances, namely the overrepresentation of smaller distances, can be attributed to reasons similar to those mentioned before. The expected number of other points of the process within a distance r from a typical point of the process is denoted as

$$K(r) = \frac{1}{\lambda}\mathbb{E}\left[n\left(\mathbb{X} \cap b(u, r) \setminus \{u\}\right) \mid u \in \mathbb{X}\right].$$

The anticipated number of points within the region $b(u, r)$ is $\lambda\pi r^2$. For a homogeneous Poisson process, this is independent of the intensity.

$$K_{\text{pois}}(r) = \pi r^2. \tag{3}$$

Estimators for this value are adjusted and normalized as empirical distribution functions of the pairwise distances, which is

$$\hat{K}(r) = \frac{1}{\hat{\lambda}^2 \,\text{area}(W)} \sum_i \sum_{j \neq i} \mathbf{1}\{\|x_i - x_j\| \leq r\}\, e(x_i, x_j; r), \tag{4}$$

where $e(u, v, r)$ is the edge adjustment factor. If $\hat{K}(r) > K_{\text{pois}}(r) = \pi r^2$, clustering is observed, while if $\hat{K}(r) < K_{\text{pois}}(r) = \pi r^2$, a regular pattern is indicated.

In addition to the theoretical Poisson model $K(r)$, boundary-corrected estimation, translation-corrected estimation, and isotropic correction estimation are computed.

Numerous other metrics and procedures are associated with the statistics of spatial points, but these appear to be the most important. The previous findings must also be tested using inferential statistical methods.

Given the significance of the K statistic even within this narrow scope, the results of the K statistic were tested using a Monte Carlo method suitable for spatial data [3].

A key question in the K statistic is whether there is a difference between $\hat{K}$ and K_{pois}.

The initial hypothesis is

$$H_0 : \text{The observed point set is a representation of a random spatial process.} \tag{5}$$

The reference curve for the procedure was the K function under complete spatial randomness (CSR). M independent simulations (with $M = 39$) for a two-sided test at a 5% significance level, as $\alpha = \frac{2}{M+1}$ were run for the study regions W. The estimated K functions $\hat{K}^{(j)}(r)$ for

$j = 1, \ldots, M$ were calculated for each realization. The lower (L) and upper (U) pointwise envelopes for these simulated curves are

$$L(r) = \min_j \hat{K}^{(j)}(r)$$

and

$$U(r) = \max_j \hat{K}^{(j)}(r).$$

For a fixed r, the probability of $K_b(r)$ exceeding the envelope $[L(r), U(r)]$ for simulated curves indicates rejection of the null hypothesis of a uniform Poisson process, with a significance level of $\alpha = \frac{2}{M+1}$. Alternatively, using pointwise order statistics provides a test with exact size $\alpha = \frac{2k}{M+1}$ for the k-th largest and k-th smallest values.

In addition to the Monte Carlo test, a quadrat test was conducted. This test does not rely on the L-K statistic, making it suitable for verifying our previous results. The study regions are divided into equal-sized quadrats (2x2), and the number of points within each quadrat is counted. Under the H_0, the points (trees) follow a homogeneous Poisson process, meaning they are randomly distributed, similarly to the previous null hypothesis. The observed frequency distribution of points per quadrat is compared to the expected Poisson distribution. The chi-square test statistic is used to quantify the difference is

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i},$$

where O_i is the observed frequency and E_i is the expected frequency.

3 Results

During the process, 257 trees were identified in the forested area and 47 trees in the mixed vegetation area (Figure 2).

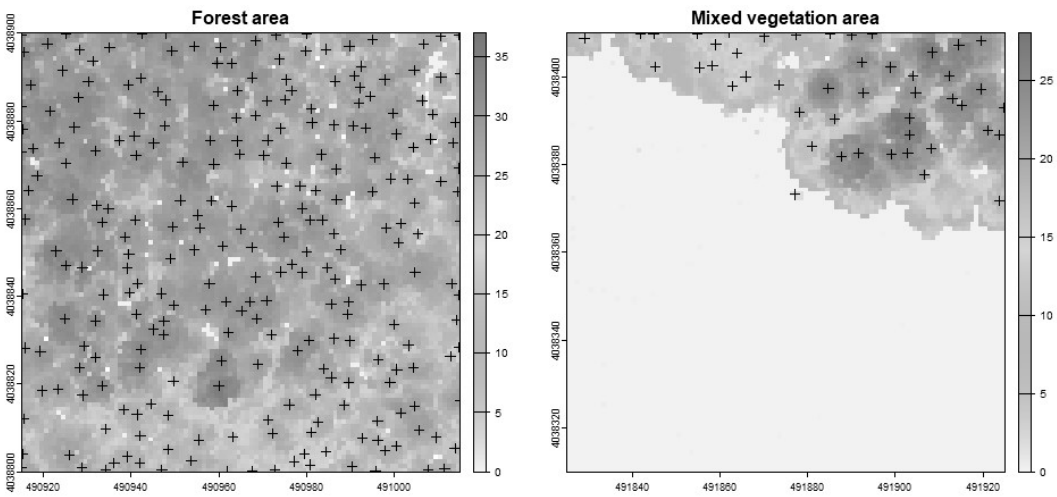


Figure 2. Canopy and trees

The estimation of $G(r)$ from (1) suggests that the pattern of trees in the forest area is regular. Specifically, $G(r) = 0$ for $r \leq 2.5$ metres, indicating that there are no nearest-neighbour distances less than 2.5 metres. The difference between the two plant covers is reflected in the

G statistic (1) and (2). In the second area, there are no trees within a distance of 2.5 metres as well, but for $r \geq 3.5$, the forest exhibits strong clustering characteristics (Figure 3).

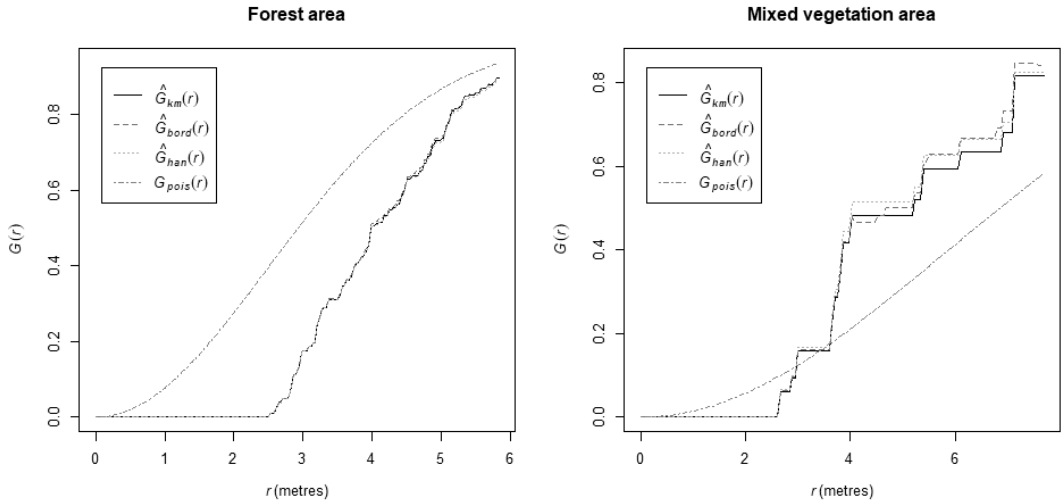


Figure 3. G statistic

Unlike what was observed with the G statistic, the K statistic (3) and (4) indicates that in the fully forested area, a pattern similar to the Poisson model can be seen. However, in the mixed vegetation area, clustering is observed, similar to previous findings (Figure 4).

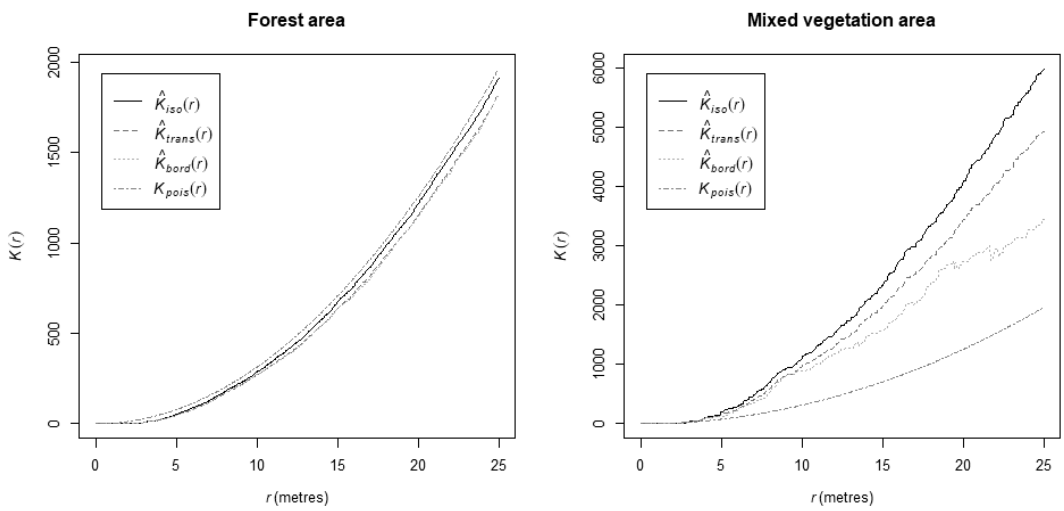


Figure 4. K statistic

In the first case, H_0 from (5) of randomness is not definitively rejected, whereas in the second case, it is unequivocally rejected. This is consistent with our prior findings (Figure 5).

The obtained result was confirmed by the χ^2 test (Table 2 and Figure 6).

Area	χ^2	df	p-value
Forest Area (2x2)	2.2529	3	0.9568
Mixed Vegetation Area (2x2)	62.362	3	3.678×10^{-13}

Table 2. Results of quadrat tests

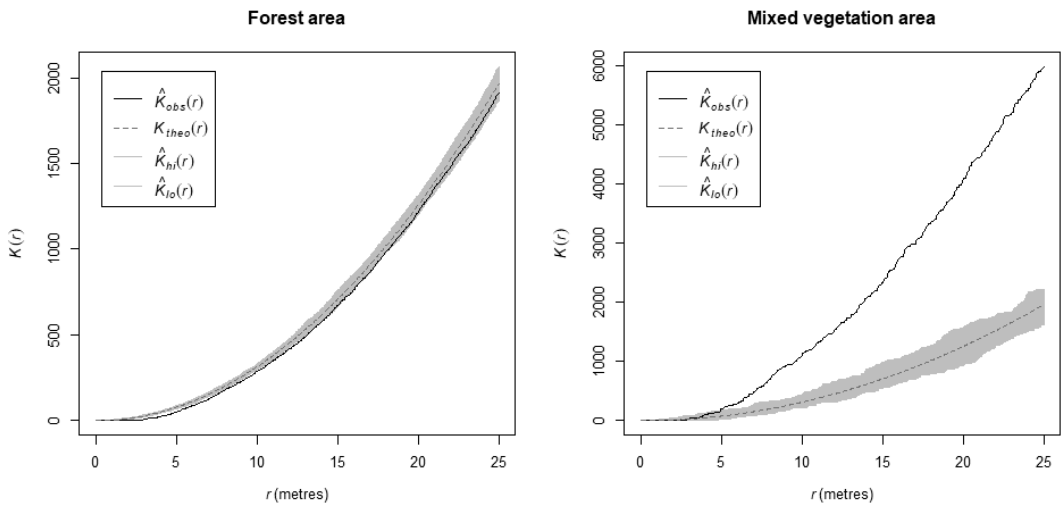


Figure 5. Pointwise envelopes

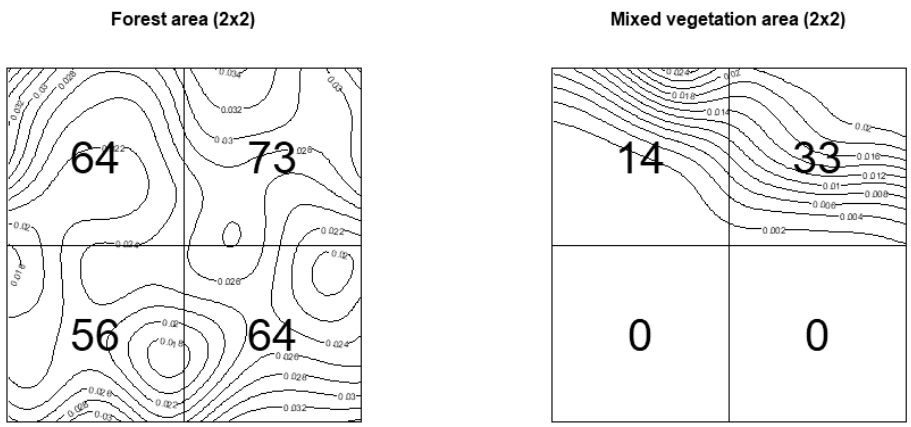


Figure 6. Quadrats with number of trees and contour lines

4 Conclusions

While the current models for individual tree identification using LiDAR data still require refinement, the incorporation of known field data and advanced statistical methods like G and K statistics holds promise for improving survey efficiency and accuracy. The continued development and validation of these models are essential for their widespread adoption in forestry applications, particularly in the context of sustainable forest management and conservation.

The comparison of the two areas using applied identification and statistical methods has demonstrated that, essentially, valid conclusions can be drawn about the trees covering the area even without field surveys, given a cost-effective technological background. We believe that the refinement of models beyond the results of field surveys and the more precise identification of trees is often impractical, as such methods are generally not generalizable. Reverting to random distributions and filtering significant anomalies alongside the methods we use offers a limited but important means of drawing conclusions in forestry and the timber industry. The applied method provides an opportunity to identify anomalies that have occurred in the forest area (tree mortality, tree cutting, natural damage) based on previous recordings.

Acknowledgements

This article was made in frame of the project TKP2021-NVA-13 which has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021-NVA funding scheme.

Bibliography

[1] **Apostol, B., Petrila, M., Lorenţ, A., Ciceu, A., Gancz, V., and Badea, O.**: *Species discrimination and individual tree detection for predicting main dendrometric characteristics in mixed temperate forests by use of airborne laser scanning and ultra-high-resolution imagery*, Science of The Total Environment, **698** (2020), 134074. doi: 10.1016/j.scitotenv.2019.134074.

[2] **Baddeley, A.**: *Analysing spatial point patterns in R*, (2010).

[3] **Besag, J. and Clifford, P.**: *Generalized Monte Carlo Significance Tests*, Biometrika, **76** (1989), No. 4, 633–642, publisher: [Oxford University Press, Biometrika Trust]. doi: 10.2307/2336623.

[4] **Douss, R. and Farah, I. R.**: *Extraction of individual trees based on Canopy Height Model to monitor the state of the forest*, Trees, Forests and People, **8** (2022), 100257. doi: 10.1016/j.tfp.2022.100257.

[5] **Duncanson, L. I., Cook, B. D., Hurtt, G. C., and Dubayah, R. O.**: *An efficient, multi-layered crown delineation algorithm for mapping individual tree structure across multiple ecosystems*, Remote Sensing of Environment, **154** (2014), 378–386. doi: 10.1016/j.rse.2013.07.044.

[6] **Hakula, A., Ruoppa, L., Lehtomäki, M., Yu, X., Kukko, A., Kaartinen, H., Taher, J., Matikainen, L., Hyyppä, E., Luoma, V., Holopainen, M., Kankare, V., and Hyyppä, J.**: *Individual tree segmentation and species classification using high-density close-range multispectral laser scanning data*, ISPRS Open Journal of Photogrammetry and Remote Sensing, **9** (2023), 100039. doi: 10.1016/j.ophoto.2023.100039.

[7] **Hanisch, K.-H.**: *Scattering Analysis of Point Processes and Random Measures*, Mathematische Nachrichten, **117** (1984), No. 1, 235–245, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/mana.3211170119. doi: 10.1002/mana.3211170119.

[8] **Jaskierniak, D., Lucieer, A., Kuczera, G., Turner, D., Lane, P. N. J., Benyon, R. G., and Haydon, S.**: *Individual tree detection and crown delineation from Unmanned Aircraft System (UAS) LiDAR in structurally complex mixed species eucalypt forests*, ISPRS Journal of Photogrammetry and Remote Sensing, **171** (2021), 171–187. doi: 10.1016/j.isprsjprs.2020.10.016.

[9] **Kansanen, K., Vauhkonen, J., Lähivaara, T., Seppänen, A., Maltamo, M., and Mehtätalo, L.**: *Estimating forest stand density and structure using Bayesian individual tree detection, stochastic geometry, and distribution matching*, ISPRS Journal of Photogrammetry and Remote Sensing, **152** (2019), 66–78. doi: 10.1016/j.isprsjprs.2019.04.007.

[10] **Kaplan, E. L. and Meier, P.**: *Nonparametric Estimation from Incomplete Observations*, Journal of the American Statistical Association, **53** (1958), No. 282, 457–481, publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1958.10501452. doi: 10.1080/01621459.1958.10501452.

[11] **Kostensalo, J., Mehtätalo, L., Tuominen, S., Packalen, P., and Myllymäki, M.**: *Recreating structurally realistic tree maps with airborne laser scanning and ground measurements*, Remote Sensing of Environment, **298** (2023), 113782. doi: 10.1016/j.rse.2023.113782.

[12] **Lisiewicz, M., Kamińska, A., and Stereńczak, K.**: *Recognition of specified errors of Individual Tree Detection methods based on Canopy Height Model*, Remote Sensing Applications: Society and Environment, **25** (2022), 100690. doi: 10.1016/j.rsase.2021.100690.

[13] **Mohan, M., Mendonça, B. A. F. d., Silva, C. A., Klauberg, C., de Saboya Ribeiro, A. S., Araújo, E. J. G. d., Monte, M. A., and Cardil, A.**: *Optimizing individual tree detection accuracy*

and measuring forest uniformity in coconut (Cocos nucifera L.) plantations using airborne laser scanning, Ecological Modelling, **409** (2019), 108736. doi: 10.1016/j.ecolmodel.2019.108736.

[14] **OpenTopography**: *Lidar Survey of Sparta Earthquake Rupture, NC 2020*, 2022. URL https://opentopography.org/meta/OT.122022.32617.1

[15] **Ramiya, A. M., Nidamanuri, R. R., and Krishnan, R.**: *Individual tree detection from airborne laser scanning data based on supervoxels and local convexity*, Remote Sensing Applications: Society and Environment, **15** (2019), 100242. doi: 10.1016/j.rsase.2019.100242.

[16] **Roussel, J.-R., Caspersen, J., Béland, M., Thomas, S., and Achim, A.**: *Removing bias from LiDAR-based estimates of canopy height: Accounting for the effects of pulse density and footprint size*, Remote Sensing of Environment, **198** (2017), 1–16. doi: 10.1016/j.rse.2017.05.032.

[17] **Schaller, C., Ginzler, C., van Loon, E., Moos, C., Seijmonsbergen, A. C., and Dorren, L.**: *Improving country-wide individual tree detection using local maxima methods based on statistically modeled forest structure information*, International Journal of Applied Earth Observation and Geoinformation, **123** (2023), 103480. doi: 10.1016/j.jag.2023.103480.

[18] **Sun, Y., Jin, X., Pukkala, T., and Li, F.**: *Two-level optimization approach to tree-level forest planning*, Forest Ecosystems, **9** (2022), 100001. doi: 10.1016/j.fecs.2022.100001.

[19] **Tompalski, P. J.-R. R., Tristan R. H. Goodbody**: *The lidR package*. URL https://r-lidar.github.io/lidRbook/

[20] **Tóth, Z.**: *Data and R code for processing and analyzing LiDAR data of the Sparta area*, (2024), publisher: Zenodo. doi: 10.5281/ZENODO.14006303.

[21] **Valbuena, R., Vauhkonen, J., Packalen, P., Pitkänen, J., and Maltamo, M.**: *Comparison of airborne laser scanning methods for estimating forest structure indicators based on Lorenz curves*, ISPRS Journal of Photogrammetry and Remote Sensing, **95** (2014), 23–33. doi: 10.1016/j.isprsjprs.2014.06.002.

[22] **Zhang, J., Sohn, G., and Brédif, M.**: *A hybrid framework for single tree detection from airborne laser scanning data: A case study in temperate mature coniferous forests in Ontario, Canada*, ISPRS Journal of Photogrammetry and Remote Sensing, **98** (2014), 44–57. doi: 10.1016/j.isprsjprs.2014.08.007.