horticulturae

Article

Mapping Frost Damage in Coffee Plantations Using RPA (Remotely Piloted Aircraft)-Mounted LiDAR (Light Detection and Ranging) Structural Metrics

Amara Lana Valim Abreu [1], Gabriel Araújo e Silva Ferraz [1,*], Laila Luana Campos [1], Rosalra Maria Alves de Morais [1], Rafael de Oliveira Faria [1], Giuseppe Rossi [2] and Matteo Barbari [2]

[1] Department of Agricultural Engineering, School of Engineering, Federal University of Lavras (UFLA), s/n, P.O. Box 3037, Lavras 37200-900, Brazil; amaralana.1000@gmail.com (A.L.V.A.); lailaluanacampos@gmail.com (L.L.C.); rosalra.morais2@estudante.ufla.br (R.M.A.d.M.); rafael.faria@ufla.br (R.d.O.F.)

[2] Department of Agriculture, Food, Environment and Forestry (DAGRI), University of Florence, Via San Bonaventura 13, 50145 Florence, Italy; giuseppe.rossi@unifi.it (G.R.); matteo.barbari@unifi.it (M.B.)

* Correspondence: gabriel.ferraz@ufla.br

Abstract

Precision agriculture (PA) relies on data collection and analysis to understand spatial variability in agricultural crops, using technologies such as sensors, GNSS, remotely piloted aircraft (RPAs), and management software. This study aimed to evaluate the potential of a LiDAR sensor coupled to an RPA to map damage in coffee plants after frosts and fill scientific gaps. To this end, the experiment was conducted in Santo Antônio do Amparo, Brazil, in May 2023, obtaining data on height and canopy diameter of three coffee plantations, aged 3, 4, and 10 years, affected by frosts in different years (2019 and 2021). Multivariate statistical analyses were performed using non-hierarchical clustering tools to classify the coffee plants into two groups according to frost incidence. Two new classifications were proposed for the two groups, based on k-means non-hierarchical cluster analysis. Thus, the study showed significant differences in the variables, with a p-value of $<2.2 \times 10^{-16}$ for the proposed groupings 1 and 2. Therefore, the groups are statistically different.

Keywords: digital and precision agriculture; remote sensing; automated aircraft; coffee growing

Check for updates

Academic Editor: Georgios Koubouris

Received: 15 June 2026
Revised: 14 July 2026
Accepted: 16 July 2026
Published: 23 July 2026

1. Introduction

Coffee cultivation plays a prominent role in the Brazilian economy, with Brazil being the world's largest coffee producer and accounting for a substantial share of global coffee production [1,2]. With a cultivated area exceeding 2.2 million hectares and an estimated production of 54.94 million 60-kg bags, coffee remains one of the country's most important agricultural commodities. Despite the technological advances achieved over recent decades, the sector continues to face challenges related to operational costs and the need to optimize crop monitoring and crop recovery, highlighting the importance of developing more efficient methodologies and technologies for coffee crop management [3–5]. Monitoring the structural development of coffee plants is an essential component of efficient coffee crop management, as changes in canopy architecture directly reflect vegetative growth, plant vigor, and responses to environmental conditions. Furthermore, canopy structural characteristics are closely associated with solar radiation interception, photosynthetic efficiency, biomass accumulation, and yield potential, making their quantification an important step for agricultural monitoring and decision-making in precision agriculture [6–10]. Among the main agronomic parameters used to evaluate plant development, plant height, canopy diameter, canopy volume, and leaf area index (LAI) stand out as important indicators of vegetative growth, plant vigor, and yield potential [11–14]. These variables reflect changes in canopy architecture and plant structural development and are frequently used to evaluate vegetation responses to different environmental conditions and stress factors [15–17]. In addition, they directly influence agronomic management practices, including irrigation, fertilization, crop health monitoring, and yield estimation [12–14]. Traditionally, these parameters are obtained through manual measurements, which present limitations related to high labor demand, time-consuming data acquisition, limited spatial representativeness, and susceptibility to operational errors, particularly in large agricultural areas [1,14]. In this context, remote sensing technologies and remotely piloted aircraft systems (RPAS) have emerged as efficient alternatives for acquiring these variables in a rapid, accurate, and non-destructive manner [15,16]. These technologies enable large-scale data acquisition with high spatial resolution, allowing detailed characterization of the structural variability of agricultural crops [17,18]. In addition to acquiring structural parameters, these technologies enable continuous crop monitoring throughout the growing season, facilitating the identification of spatial and temporal variations associated with plant growth, water, nutritional, and thermal stress, while supporting decision-making in precision agriculture [19–21]. Furthermore, remote sensing has proven effective for estimating biophysical parameters, such as plant height, canopy diameter, and leaf area index (LAI), which are directly associated with photosynthetic capacity and yield potential in coffee crops [22–24]. Among emerging technologies, Light Detection and Ranging (LiDAR) has stood out due to its ability to acquire highly accurate three-dimensional data through the generation of dense and detailed point clouds. These data enable precise structural characterization of plants, even under low-light conditions or in areas with dense vegetation. LiDAR sensors mounted on remotely piloted aircraft systems (RPAS) allow the generation of detailed three-dimensional crop models, reducing the need for field measurements and improving the efficiency of agricultural monitoring [15,25,26]. Recent studies have demonstrated that this technology provides high accuracy in estimating structural parameters in perennial crops, including coffee [15]. Another critical factor affecting coffee production is frost occurrence, a climatic phenomenon capable of causing severe damage to plants and resulting in significant economic losses [27–29]. Exposure to temperatures below −2 °C may induce tissue necrosis, loss of apical dominance, and the death of productive branches, ultimately impairing plant development and productivity [30,31]. Frost damage is directly influenced by microclimatic and topographic factors. In areas with steeper slopes, cold air, due to its higher density, tends to flow downhill and accumulate in lower regions, such as valleys and depressions, where temperatures may reach critical levels [16,27,32,33]. Under these conditions, the freezing of water within plant tissues causes structural damage to cells, compromising their integrity and physiological function [34]. This process results in reduced photosynthetic activity, leaf necrosis, structural damage to branches, and impaired vegetative and reproductive development, directly affecting plant formation, quality, and productivity [23,35,36]. Several methodologies have been employed to map frost risk and frost-induced damage in agricultural systems [32,34]. Traditional approaches are generally based on topographic analyses, historical meteorological records, field observations, and sample data, enabling the identification of areas that are more susceptible to cold air accumulation and the occurrence of thermal inversions [23,27]. Although these methods provide valuable information regarding frost occurrence, they present limitations related to low spatial resolution, limited population representativeness, and the inability to adequately capture microclimatic variability within cultivated areas. More recently, remote sensing techniques have been employed to improve the monitoring and assessment of frost-induced damage, combined with field sample data [37]. Studies using multispectral imagery and vegetation indices have demonstrated the potential to detect vegetation stress resulting from frost events [23,27]. However, these approaches are predominantly based on spectral responses and may not fully represent the structural changes in plant architecture caused by this type of stress. In this context, LiDAR technology has emerged as a promising alternative because it enables the acquisition of high-resolution three-dimensional structural information, allowing a more detailed representation of canopy damage and spatial variability within crop fields [36]. In this context, precision agriculture has emerged as an important tool for monitoring and managing coffee plantations. The use of RPAS equipped with remote sensing sensors enables detailed mapping of spatial variability, contributing to optimized input use, reduced operational costs, and improved production efficiency [15,36]. Despite the considerable potential of LiDAR technology for the structural characterization of perennial crops, there is still a limited number of studies exploring its application for detecting and spatially characterizing frost-induced damage in coffee plantations. Furthermore, few studies have investigated the integration of LiDAR-derived structural parameters with spatial clustering techniques to identify damage patterns associated with climatic stress events. In addition, recent studies have shown limited exploration of non-hierarchical clustering techniques for the analysis of LiDAR-derived structural data, particularly for identifying spatial patterns and delineating homogeneous zones in areas affected by climatic stress. The application of these techniques can significantly contribute to the identification of regions exhibiting different levels of damage, enabling a more detailed analysis of spatial variability and supporting site-specific crop management decisions. Although unsupervised clustering approaches, such as K-means, have been widely applied in agricultural remote sensing, their use has been predominantly focused on general vegetation segmentation and crop damage identification using multispectral or time-series remote sensing data [38]. For example, K-means-based machine learning approaches have been applied to identify agricultural areas affected by hail damage using remote sensing time-series data [38]. Similarly, studies have demonstrated the applicability of K-means clustering directly to LiDAR point clouds for canopy structural segmentation [39], while other authors have employed LiDAR-derived three-dimensional descriptors and intensity metrics for vegetation detection and classification [40]. However, despite these advances, studies integrating structural metrics derived from RPAS-mounted LiDAR with unsupervised clustering techniques to assess frost damage in coffee plantations remain extremely limited. In this context, the present study provides a methodological contribution by combining high-resolution structural metrics obtained from RPAS-mounted LiDAR with K-means clustering ($k = 2$) to discriminate frost-affected and non-affected coffee plants based on canopy structural responses. This approach represents an innovative application of unsupervised machine learning in precision coffee farming, particularly for the rapid mapping of post-frost damage using three-dimensional structural information derived from LiDAR data. Therefore, the proposed methodology expands the applications of LiDAR-based agricultural monitoring by integrating vegetation structural characterization, high-resolution data acquisition using a remotely piloted aircraft system (RPAS), and non-hierarchical clustering techniques for the assessment of frost damage in coffee plantations. The hypothesis of this study is that coffee plants of different ages exposed to frost exhibit measurable structural differences that can be quantified using LiDAR-derived metrics, including plant height, canopy diameter, canopy volume, leaf area index (LAI), and elevation. These metrics enable the discrimination of areas exhibiting different levels of frost damage, previously identified through field observations following frost events. In this context, the objective of this study was to evaluate the ability of these structural variables, derived from LiDAR data acquired by an RPAS and combined with non-hierarchical clustering techniques, to identify spatial patterns associated with frost damage severity in coffee plantations. By integrating high-resolution three-dimensional information with clustering-based spatial analysis methods, this approach provides an effective alternative for identifying and characterizing frost-affected zones, thereby supporting agricultural monitoring and the development of site-specific management strategies in precision agriculture.

2. Materials and Methods

The present experiment was conducted in the rural area of the city of Santo Antônio do Amparo, at the Bom Jardim farm, located at the confluence of the southern and western regions of Minas Gerais, Brazil, using three distinct areas, described in Figure 1 and Table 1.

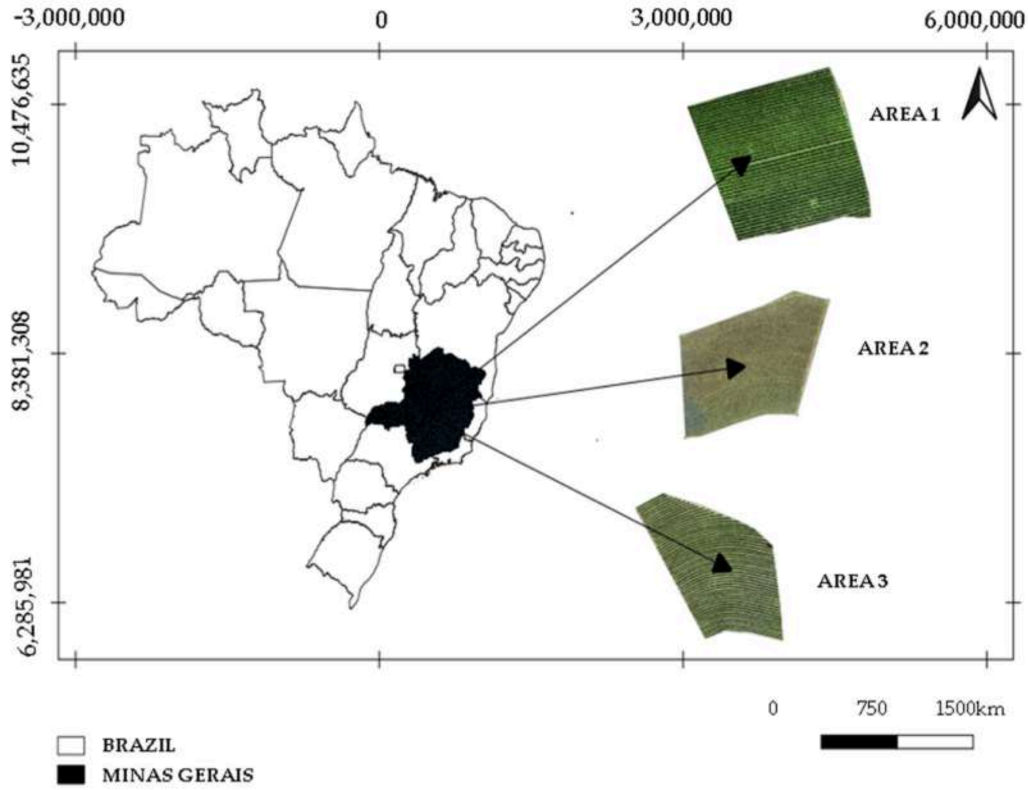

Figure 1. Location of the cultivated areas on the Bom Jardim Farm, where the data for the present study were collected.

Table 1. Summary of the main characteristics of the areas under study.

Region	Variety	Coordinates (UTM, SIRGAS 2000)	Average Altitude (m)	Area (ha)	Year of Planting	Number of Frost Occurrences
Area 1	Red Catucaí IAC 144	7,677,527.305288 N e 603,932.761361 W	926	3.56 ha	2013	2
Area 2	Red Catucaí IAC 144	508,019.6 N e 7,676,692.0 W	932	3.52 ha	2019	1
Area 3	Catucaí Red IAC 144 and Arara	508,093.6 N e 7,676,753.1 W	935	2.10 ha	2020	1

In Area 1, planting was carried out on 21 December 2013, with coffee plants spaced 0.70 m between plants and 3.10 m between rows. In Areas 2 and 3, the adopted spacing was 0.5 m between plants and 3.5 m between rows. Plant age was determined based on the time elapsed between the planting date and the data collection performed in this experiment, corresponding to 3, 4, and 10 years. The soil in these areas was characterized as dystrophic Red-Yellow Latosol. Area 1 was affected by frost events in two distinct periods, in 2019 and 2021, whereas Areas 2 and 3 were affected only by the frost events of 2021. Therefore, the plant populations in these areas were impacted by this climatic event and consequently experienced damage, which had been previously verified and classified [23,27]. The damages are presented in Figure 2.

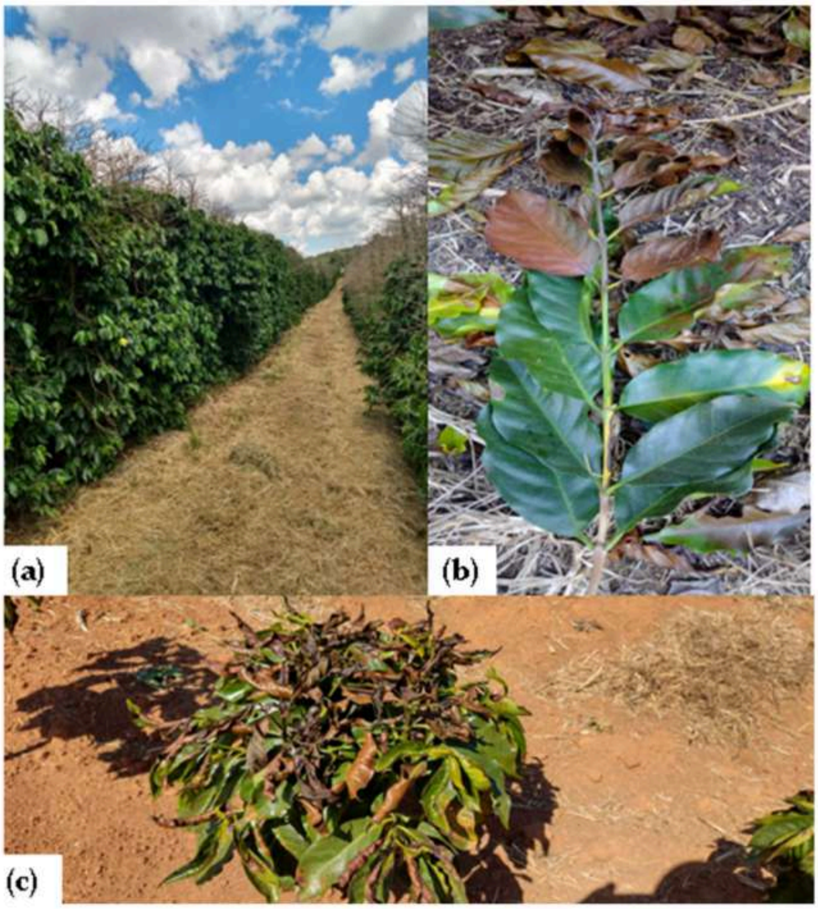

Figure 2. Records of crops affected by frost in 2021: area 1 (**a**), area 2 (**b**), and area 3 (**c**).

Although the three experimental areas differed in planting year, plant age, row spacing, frost history, and management practices, these factors were not analyzed jointly. Each experimental area was treated as an independent dataset during the clustering and statistical analyses, thereby minimizing the influence of potential confounding factors. Consequently, the clustering procedure was performed separately for each area, allowing the structural variability associated with frost damage to be evaluated under relatively homogeneous agronomic conditions rather than through direct comparisons among fields.

In addition to the structural variables obtained through remote sensing, the clustering results were subsequently compared with field records previously collected regarding the levels of frost damage in the studied areas, as described in [23,27]. This comparison allowed an independent validation of the groups identified by the clustering algorithm, verifying whether the detected spatial patterns were consistent with the actual conditions observed in the field.

In sloped terrains, when temperatures fall below the freezing point of water, the cold air that accumulates near the surface tends to flow downslope toward the bottom of the slopes and accumulate in depressions or valleys [32,33]. Exposure of plants to low temperatures can lead to the freezing of water within cellular tissues, rupturing cell walls and causing damage, which gives the plant a burned-like appearance [34].

Ambient air temperature data were collected by an automatic meteorological station located on the farm. The first frost event in 2019 occurred on 8 and 9 July of that year, and the minimum temperatures observed on those days were 1.8 °C and 0.3 °C. In 2021, the first frost occurred on 20 July, with a minimum observed temperature of −1.7 °C. The climatic classification of the region is Cwa, humid subtropical, according to the Köppen–Geiger classification, characterized by hot and humid summers and cold and dry winters, with mean, minimum, and maximum temperatures of 20 °C, 14 °C, and 26 °C, respectively. The average total precipitation is 1400 mm and the mean annual total precipitation is 1670 mm.

Considering that frost-induced damage may be influenced both by structural characteristics of the plants and by topographic factors of the terrain, different clustering strategies were evaluated in this study in order to investigate how these variables contribute to the identification of spatial damage patterns in coffee fields.

For this study, considering the analyzed areas, two distinct clustering methods were statistically defined. The first clustering (Clustering 1) includes analyses based on plant height and canopy diameter. The second clustering (Clustering 2) comprises analyses that consider plant height, canopy diameter, and terrain elevation. The use of these two clustering strategies made it possible to evaluate the effect of including topographic variables in the identification of plant structural patterns associated with frost damage. Thus, it was possible to compare the performance of clustering approaches based exclusively on plant structural characteristics with those incorporating terrain information.

For each proposed clustering approach, the classification of coffee plants was performed according to the degree of frost damage, thereby generating two groups, referred to as group 1 and group 2, corresponding to the most and least affected plants, respectively.

The number of groups defined for the non-hierarchical clustering in the present study was based on recent studies conducted in the same area [23,27], which evaluated the occurrence and validation of frost damage in coffee plants. The selection of two groups is also associated with the agronomic hypothesis that plants may present two predominant response levels after frost events: severely affected plants and plants with a lower level of damage. This assumption is consistent with field observations reported in previous studies conducted in the same experimental area.

On 21 May 2023, an overflight was conducted using an unmanned aerial vehicle (UAV) equipped with a LiDAR sensor over the study areas to obtain data on coffee plants in each of the three areas analyzed. For this purpose, a quadcopter-type UAV, model DJI Matrice 300, equipped with a real-time kinematic (RTK) global positioning system, was used (Figure 3a). This UAV has a flight autonomy of up to 55 min, a six-direction sensing and positioning system, an enhanced transmission system, and real-time automatic frequency switching between the 2.4 and 5.8 GHz bands, resulting in greater flight stability in areas with high levels of interference, such as power transmission lines.

A Zenmuse L1 sensor (Figure 3b) was mounted on this UAV, consisting of a Livox LiDAR module, a high-precision inertial measurement unit (IMU), and a 20 MP RGB camera with a 1-inch CMOS sensor, 8.8 mm focal length, and a mechanical shutter, all mounted on a three-axis stabilized gimbal.

To ensure greater accuracy in flight execution and in the georeferencing of the information obtained by the LiDAR sensor, a GNSS antenna system was used, consisting of an RTK L1 and L2 base integrated with the UAV and another base fixed on the ground for correction using the DJI D-RTK 2 Mobile Station.

https://doi.org/10.3390/horticulturae12080904

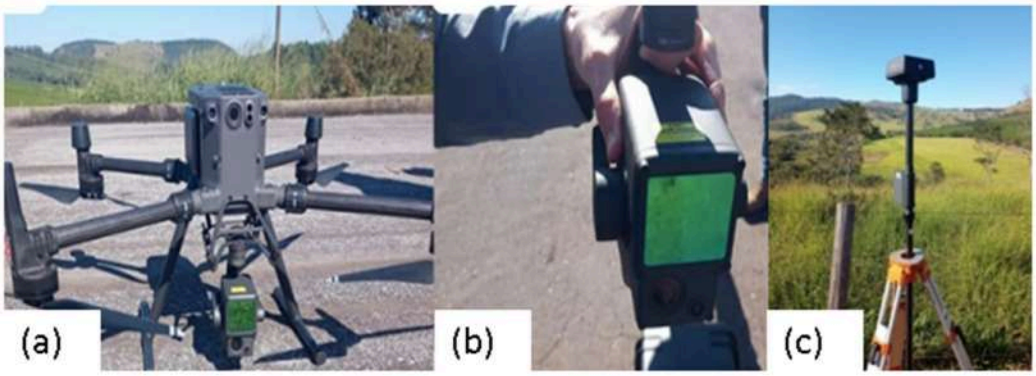

Figure 3. RPA Matrice 300 RTK (**a**) and Zenmuse L1 sensor (**b**). Fixed base on the ground for GNSS signal correction, DJI DRTK-2 model (**c**).

Before the flight began, the parameters were defined, such as a flight altitude of 75 m relative to the ground, coverage area, and a speed of 4.1 m/s. The image survey was carried out using the DJI Pilot 2 app (Figure 4), employing a point cloud density of 685 points/m^2; a GSD of 0.96 cm/pixel; a lateral overlap (LiDAR) of 50%; a forward LiDAR (visible) overlap of 70%; a lateral LiDAR (visible) overlap of 61%; and a 10-m margin around the study area for flight planning, ensuring that the area of interest was well covered. Additionally, a sampling rate of 160 kHz was chosen to ensure that the images and collected data could provide high quality and meet the expected objectives.

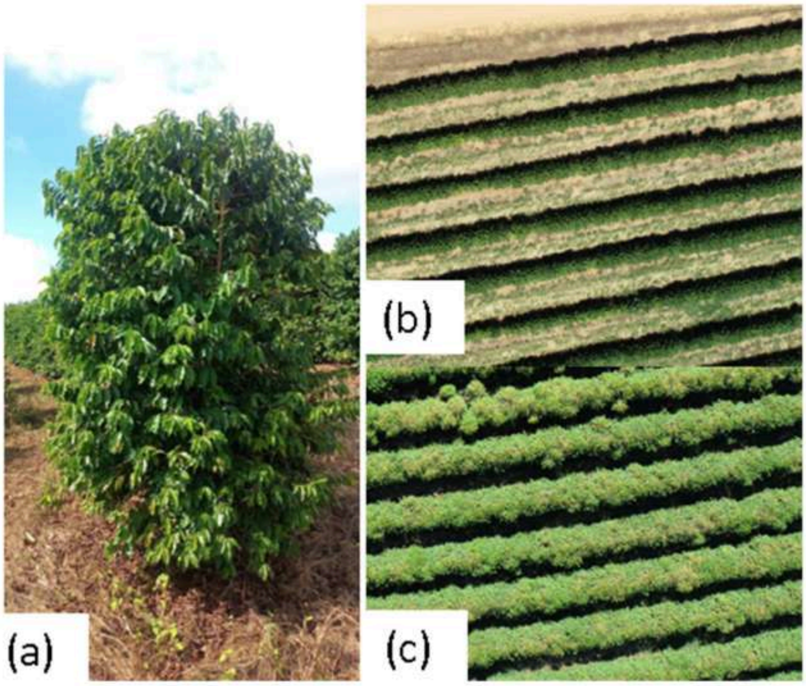

Figure 4. Coffee plant specimen found in the field, area 1 (**a**); top view of plants from area 2 (**b**); top view of plants from area 3 (**c**).

2.1. Processing of Data Obtained from the LiDAR Sensor

During the alignment of the images, a sparse point cloud was generated through homologous points automatically identified in the process. Subsequently, the point cloud generated in the previous process was densified, increasing the number of points in the cloud and reducing empty spaces. Finally, the orthomosaic was created; in this phase, the images are orthorectified, where they are reprojected orthogonally and at a constant scale, in order to eliminate or minimize distortions caused by the sensor system and the surface.

With the images properly corrected, the software performs mosaicking and creates a single product. The survey generated georeferenced LiDAR point clouds in (.las) format, whose data were processed in DJI Terra software. The point clouds were georeferenced using the information obtained from the sensor and the RTK accuracy of the UAV and the stationary base. After collecting the data from the LiDAR sensor, the geographic spatial reference was corrected to orthometric (Sirgas 2000 UTM; zone 23S). The adopted image correction methodology was triangulation; the point cloud was classified automatically and then manually to define representative ground points. The orthomosaic was georeferenced based on RTK, ensuring an accuracy of 0.02 mm. From these, the digital surface model (DSM) and the digital terrain model (DTM) were generated. The next steps involve conducting analyses in the QGIS software to obtain information on each georeferenced plant, using data from the LiDAR sensor and the RTK database. Through geoprocessing in QGIS, the crown diameters and plant heights were determined. Plant heights were obtained using raster tools, considering the differentiation between Digital Surface Models (DSMs) and Digital Terrain Models (DTMs), generated from processing the LiDAR sensor data. The crown diameter measurements were obtained using the vector measurement tools in QGIS. The canopy diameter of the plants was determined directly on the orthomosaic, measured along the largest central axis horizontally, using the ruler tool to measure distances in metric units. For the plant volume variable, a calculation was proposed based on the canopy diameter and coffee plant height information, using the cylinder equation (Equation (1)). In the present study, this variable was not intended to represent the actual canopy volume or to perform a three-dimensional reconstruction of the plant. Instead, canopy volume was included solely as a standardized structural indicator to evaluate its behavior in conjunction with the other biometric variables analyzed, considering that the plants exhibited the corresponding canopy structure under field conditions. This approach was adopted due to the similarity of the plants to a cylindrical structure, as observed in Figure 4

$$V = 3.1415 \cdot R^2 \cdot H \quad (1)$$

where:

R = radius;
H = plant height.

The structure of the plant following a cylindrical shape can be explained through mechanized harvesting, as this can cause deformations in the plant's branches, thus modifying the structural shape of the coffee plant [35]. The leaf area index (LAI) was calculated according to Equation (2), developed in [1] and described in [25,37].

$$LAI = 0.0134 + 0.7276 \cdot D^2 \cdot H \quad (2)$$

where:

D = Plant canopy diameter;
H = Plant height.

The altimetric information of the terrain was obtained through raster tools, by extracting information from digital terrain models, derived from LiDAR data processing and geoprocessing using QGIS software, as previously described.

The use of LiDAR as a reference is supported by its extensive validation and well-established reliability in the scientific literature for accurately characterizing vegetation structure, particularly in estimating plant metrics. Thus, the methodological approach focused on a non-hierarchical clustering analysis of LiDAR-derived data in areas affected after frost occurrence, emphasizing the consistency and relative performance of the estimates generated by the clusters in comparison with previous methodologies, both in frost-affected areas and in the literature.

2.2. Statistical Analysis

In the present study, a non-hierarchical clustering analysis was performed. In hierarchical clustering, the objective is to identify possible groupings based on the inherent structure of the data. In contrast, in non-hierarchical clustering, as applied in this study, the objective is to optimize a clustering criterion based on a predefined number of groups specified by the researcher, and this choice must be properly justified, as proposed by [41–44].

The classification of frost-risk areas was conducted and justified based on the methodology described in [23,27], in which each area was subdivided in order to classify the degree of risk of damage to plants. The authors considered visual criteria that evaluate topographic variation in the same areas proposed in the study, taking into account terrain slope and field data sampled within the areas. In contrast, the present study addresses the classification through non-hierarchical clustering based on population data of affected plants, incorporating plant structural characteristics, elevation, and statistical validation of the defined groups.

Independent evaluations were performed regardless of the clustering structure, and additional derived variables not directly used during the clustering procedure were assessed, including canopy area, estimated canopy volume, and leaf area index. Significant differences between the clusters for these variables reinforced the biological and structural consistency of the identified groups.

The k-means method, proposed by [43], is a non-hierarchical clustering algorithm widely used in exploratory data analysis and pattern recognition. It is an iterative method whose objective is to partition a dataset containing n observations into k distinct groups, such that each observation belongs to the group whose centroid (cluster center) is closest, based on a distance measure, typically the Euclidean distance. The algorithm aims to minimize the sum of the squared distances between each observation and the centroid of its respective group, known as the within-cluster sum of squares (WCSS).

Non-hierarchical cluster analysis aims to classify observations in a dataset in such a way that similar observations are allocated to the same group; therefore, observations belonging to different groups are considered dissimilar.

$$kj = 1 \sum ni \;||\; xij - \mu i \;\;||\; 2$$

Among the main advantages of the k-means algorithm are its simplicity, computational efficiency, and ease of interpretation, especially when applied to large datasets. However, the method also presents some limitations, such as the need to define the number of groups (k) in advance, which is a fundamental characteristic of non-hierarchical clustering methods. In the context of this study, the k-means method was used to classify plants into two distinct groups based on structural variables obtained through remote sensing. This approach enabled the objective identification of plants that were more and less affected by frost based on statistical and geometric criteria. This approach was based on the prior assumption that two groups of plants existed in the studied areas [3,16], hereafter referred to as Group 1 and Group 2, as defined by the clustering procedure.

Thus, for the development of Cluster 1 through clustering analysis, the variables plant height and canopy diameter from a population of frost-affected coffee plants were used, obtained by the LiDAR sensor. For Cluster 2, the variables considered were plant height, canopy diameter, and terrain elevation in the study area, also obtained by the LiDAR sensor. Because the experimental areas differed in agronomic characteristics, including planting year, plant age, row spacing, cultivar composition, and frost history, the clustering procedure was performed independently for each area. This strategy ensured that the identified clusters represented structural variability within each field, thereby avoiding potential bias associated with direct comparisons among heterogeneous production systems. A significance level of 5% was adopted for all statistical analyses.

Data analyses were performed using the software R Core Team (2023). Initially, to compare the mean values of the variables between Cluster 1 and Cluster 2 and to identify plants more and less affected by frost in the three analyzed areas, an analysis of variance (ANOVA) was performed. This approach allowed the evaluation of differences in plant height, canopy diameter, plant volume, and leaf area index (LAI) between Cluster 1 and Cluster 2 defined by the non-hierarchical k-means clustering method.

The variables plant volume and leaf area index (LAI), which were not directly used in the clustering process, were subsequently analyzed to determine whether they presented significant differences between the formed clusters. This strategy allowed an additional validation of the clusters by verifying whether the identified groups exhibited consistent structural differences beyond the variables used in the clustering process.

Additionally, an ANOVA was conducted to compare the mean values between the treatments (Group 1 and Group 2) for each studied variable. In this test, the null hypothesis (H_0) states that the group means are equal, whereas the alternative hypothesis (H_1) indicates that at least one mean differs between treatments. Since the study involved only two groups, the application of Tukey's multiple-comparison test was not necessary, as the ANOVA F-test is sufficient to determine the existence of significant differences between two means.

Analysis of variance assumes that the residuals follow a normal distribution, present homogeneity of variances, and are independent. In this study, the independence of errors was ensured through the random sampling process. The normality of residuals was evaluated using the Shapiro–Wilk test, whose null hypothesis states that the data follow a normal distribution, whereas homogeneity of variances was verified to ensure the validity of the ANOVA results.

To validate the optimal number of clusters, three complementary criteria were used: the Elbow Method, the Silhouette Index, and the Gap Statistic. The Elbow Method was used to evaluate the reduction in the Within-Cluster Sum of Squares (WSS) as the number of clusters increased. The Silhouette Index was employed to simultaneously measure the internal cohesion and separation between groups. Additionally, the Gap Statistic was used to compare the observed clustering structure with a random reference distribution, allowing verification of whether the identified clusters represented a real structure in the data rather than a pattern generated by chance. The optimal number of clusters was defined based on the agreement among the evaluated methods.

Finally, thematic maps were developed in a Geographic Information System (GIS) environment to spatially represent the identified clusters. These maps allowed the comparison of the spatial patterns of damage obtained by the proposed method with the traditional approach described by [27], which classifies frost damage risk primarily based on visual evaluation and topographic characteristics. This comparison made it possible to evaluate the potential of the approach based on LiDAR data and statistical clustering techniques for identifying spatial patterns of damage in coffee plantations.

3. Results and Discussion

This section presents the results of the non-hierarchical clustering analyses performed for Areas 1, 2, and 3. Descriptive analyses of the plant height and coffee canopy diameter variables were used as the basis for the non-hierarchical clustering. The original dataset comprised 3981 coffee plants in Area 1, 4690 coffee plants in Area 2, and 4586 coffee plants in Area 3.

https://doi.org/10.3390/horticulturae12080904

The analyses were performed using two groupings. Grouping 1 was based on two variables, whereas Grouping 2 was based on three variables, as described in Section 2. Furthermore, each grouping was divided into two groups, representing coffee plants that were more severely affected and less severely affected by frost, respectively. A margin of error of 5% and a confidence level of 95% were adopted for the statistical analyses.

The observations for area 1 of the diameter of group 1 of the coffee plant population showed average values of 1.79 m, and for group 2, average values of 1.84 m, when observing Cluster 1. For the plant height variable, the observed average values were 2.63 m for group 1 and 2.78 m for group 2. For the observations of the plant volume variable in area 1, the average values found for group 1 were 7.49 m^3 and for group 2 were 8.24 m^3. For the observations of the leaf area index (LAI), the average values found for group 1 were 0.930 and for group 2 were 0.1023.

For area 2, groupings 1 and 2, as observed in Table 2, did not show differences in the descriptive analyses. However, when comparing groups 1 and 2, differences were observed between them, with group 1 presenting higher averages for all the variables analyzed in the two proposed groupings. This can be justified by the region being less affected in 2021, with only one frost occurring in the studied area, and by the recovery of the plants over time, according to [23]. The study analyzed two experimental sets, each composed of two groups of plants. In both groupings, a decrease in the canopy diameter was observed, a pattern also identified for the parameters of plant height and calculated volume. The consistency of these metrics was maintained between the groups, regardless of the evaluated grouping. Regarding leaf area, the data showed constant values for group 1 across all samples, with equivalent stability in group 2. The homogeneity observed in the subgroups suggests the possible influence of common environmental variables, which may have uniformly modulated the measured characteristics throughout the analyses.

Table 2. Descriptive statistics of the groupings with 2 and 3 variables for areas 1, 2, and 3.

	Group	D	A	V	LAI
			Area 1		
AGR 1	**1**	1.79	2.63	7.49	0.930
	2	1.84	2.78	8.24	0.1023
AGR 2	**1**	1.79	2.63	7.49	0.930
	2	1.84	2.78	8.24	0.1023
			Area 2		
AGR 1	**1**	1.41	0.72	1.23	0.01533
	2	1.21	0.68	0.86	0.01070
AGR 2	**1**	1.41	0.72	1.23	0.01533
	2	1.21	0.68	0.86	0.01071
			Area 3		
AGR 1	**1**	2.42	0.96	4.46	0.055
	2	1.86	0.81	2.27	0.028
AGR 2	**1**	2.23	0.90	3.74	0.046
	2	1.99	0.84	2.74	0.034

D: crown diameter; V: crown volume; A: plant height; AGR 1: grouping 1; AGR 2: grouping 2.

The data analyzed in Area 3, as presented in the Table 2, indicate that in Cluster 1, the observations of crown diameter showed average values of 2.42 m for Group 1 and 1.86 m for Group 2. For the plant height variable, the average values were 0.96 m in Group 1 and 0.81 m in Group 2. Regarding plant volume, Group 1 had an average of 4.46 m^3, while

Group 2 had an average of 2.27 m^3. Finally, for the leaf area index (LAI), the observed average values were 0.055 in Group 1 and 0.028 in Group 2. For Cluster 2, the observations of crown diameter of the coffee plants showed average values of 2.23 m for Group 1 and 1.99 m for Group 2. For the plant height variable, the values also differed between the groups. Regarding plant volume, the average values were 3.74 m^3 for group 1 and 2.74 m^3 for group 2. In relation to the leaf area index (LAI), the average value observed was 0.046 for group 1 and 0.034 for group 2. For both groupings under study, different values can be observed for all the variables analyzed.

The analysis of data from Area 3 revealed consistent differences between the clusters across all evaluated biometric parameters, demonstrating the existence of structural variability within the coffee plant population. In cluster 1, higher mean values were observed compared to cluster 2 for all variables, including crown diameter, plant height, plant volume, and Leaf Area Index (LAI). This pattern suggests that plants belonging to cluster 1 exhibit greater vegetative development and, consequently, higher physiological vigor compared to those in cluster 2.

The larger crown diameter and greater plant volume indicate a more robust architecture, while the higher LAI values reflect a greater capacity for solar radiation interception and photosynthetic potential, factors directly associated with plant growth and productivity.

The difference observed in plant height, although numerically smaller compared to the other variables, also reinforces this distinction between the clusters. Plant height is an important indicator of longitudinal growth and may reflect cumulative differences over time in response to environmental conditions and stress history. In areas subject to frost occurrence, such as in the present study, plants with lower height and reduced volume may be associated with locations where the impact of thermal stress was more severe, thereby compromising vegetative growth. Conversely, plants with greater structural development may indicate areas with lower damage severity or more favorable microenvironmental conditions for recovery.

In Clustering 2, the same pattern was observed, with cluster 1 again presenting higher mean values than cluster 2 for all analyzed variables. This consistency between clusters reinforces the reliability of the classification and suggests that the identified clusters reflect real differences in plant structure and development. The repetition of this pattern across different clustering analyses indicates the influence of environmental and physiological factors that affect plant growth differently across space.

Although the data generally showed low absolute variation, the relative differences between clusters are agronomically relevant. Small variations in parameters such as LAI and plant volume may result in significant differences in photosynthetic capacity, biomass accumulation, and, consequently, plant productivity potential. Furthermore, these structural differences may reflect microenvironmental variations associated with topography, altitude, and cold air dynamics, factors that directly influence the intensity and spatial distribution of frost damage.

In this context, analyses based solely on data range, considering mean or extreme values in isolation, may not be sufficient to identify spatial and functional patterns within the study area. The use of non-hierarchical clustering makes it possible to overcome this limitation, as it allows the identification of homogeneous clusters with similar characteristics by simultaneously considering multiple variables. This approach is particularly relevant in agricultural systems, where variability is influenced by several interrelated factors.

Moreover, non-hierarchical clustering enables the identification of zones with different levels of vegetative development, which may be associated with varying levels of frost exposure and susceptibility. This information is essential for site-specific management, allowing the identification of more vulnerable areas and supporting the implementation of targeted monitoring and mitigation strategies. Therefore, the adopted methodology not only improves the understanding of plant structural variability but also contributes to the generation of more precise and applicable information within the context of precision agriculture.

The results obtained in Area 3 demonstrate that, even under conditions of apparently low overall variability, important structural differences exist between plant clusters, which can be more efficiently identified using non-hierarchical clustering techniques. This approach enables a more detailed characterization of spatial variability, contributing to a better understanding of plant responses to environmental conditions and providing support for the development of more efficient and targeted management strategies.

3.1. Comparison Test

For each variable studied in the coffee plant areas, an analysis of variance was performed to compare the means of the groups for each variable under study, where the same results were observed for groupings 1 and 2, respectively with 2 and 3 variables, considering a 5% significance level. The null hypothesis (H_0) to be tested is that the means are equal, and the alternative hypothesis (H_1) is that at least one mean differs among the treatments.

Comparison Test of the Variables of Groups 1 and 2 in Areas 1, 2, and 3

With the observed *p*-values for area 1, it can be stated that the only variable that showed significant differences was the height variable with a *p*-value of $<2.2 \times 10^{-16}$ for the original grouping and the proposed groupings 1 and 2. The aim was to assess the statistical difference between (group 1) and (group 2) across the three areas. The *p*-value $p < 2.2 \times 10^{-16}$ indicates evidence to reject the hypothesis of equality between the group means at a 5% significance level (Table 3). Therefore, the groups are statistically different, allowing for a profile to be suggested for these groups.

Table 3. At a 5% significance level; D: crown diameter; V: crown volume; A: plant height; LAI: leaf area index.

Variable	Shapiro—Wilk	Bartlett	Mann–Whitney
	Area 1		
D	1.47×10^{-19}	0.036	0.96
A	3.86×10^{-35}	1.84×10^{-12}	$<2.2 \times 10^{-16}$
V	3.19×10^{-43}	0.314	0.095
LAI	3.19×10^{-43}	0.338	0.095
	Area 2		
D	2.94×10^{-22}	0	$<2.2 \times 10^{-16}$
A	$<7.82 \times 10^{-14}$	$<1.51 \times 10^{-9}$	$<8.21 \times 10^{-14}$
V	$<2.2 \times 10^{-16}$	0	$<2.2 \times 10^{-16}$
LAI	$<1.29 \times 10^{-19}$	0	$<2.2 \times 10^{-16}$
	Area 3		
D	2.94×10^{-22}	0	$<2.2 \times 10^{-16}$
A	$<7.82 \times 10^{-14}$	$<1.51 \times 10^{-9}$	$<8.21 \times 10^{-14}$
V	$<2.2 \times 10^{-16}$	0	$<2.2 \times 10^{-16}$
LAI	$<1.29 \times 10^{-19}$	0	$<2.2 \times 10^{-16}$

Thus, through the analysis, it was observed that the variables studied in the areas of 3- and 4-year-old plants, plant height, plant volume, and LAI showed different medians for groups 1 and 2. The crown diameter variable did not show different medians between groups 1 and 2. For areas 2 and 3, the non-hierarchical cluster analysis considering cluster

1 and cluster 2 showed similar results. In both cases, differences were detected among all variables of the 2 groups and 2 areas, in accordance with the principle of parsimony. The study conducted with the original clustering as proposed by [16] did not detect any differences.

- *Grouping 1*

1. *Area 1*

As previously presented, the grouping of the plants into two groups was considered, and there was a clear division between them, as observed in Figure 5, in which group 1 is concentrated in the upper part and group 2 in the lower part. It is possible to observe six numerical values distributed into two clusters organized along a single axis. The calculated centroids reflect this delimitation strategy, based on the height and diameter of the crowns, prioritizing the internal homogeneity of each cluster. Thus, it can be stated that there was a very clear classification of group 1 in red and group 2 in blue.

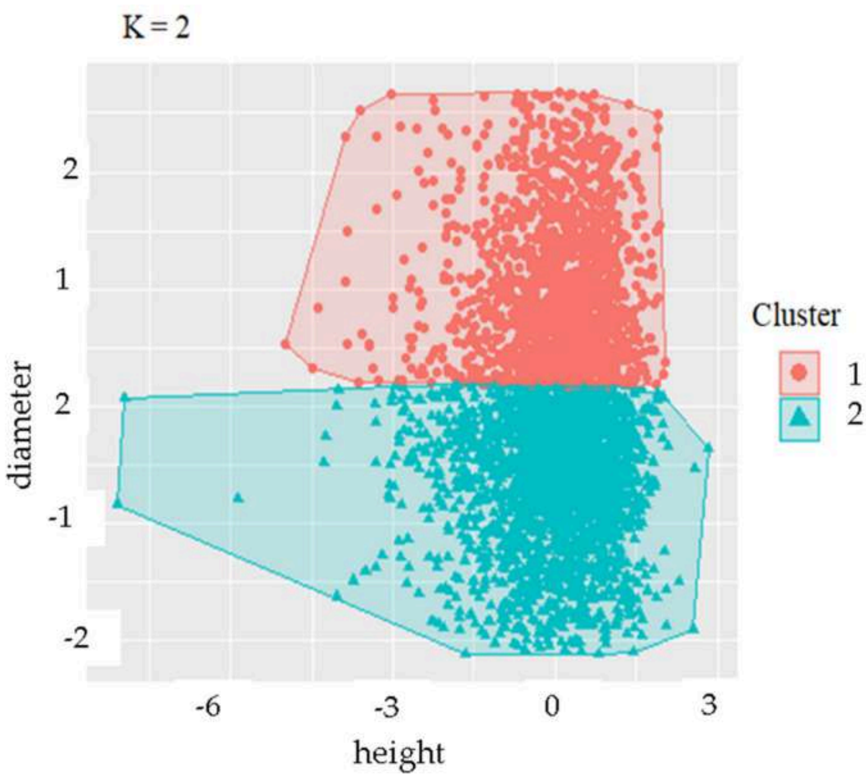

Figure 5. Scatter plot with clustering for Area 1.

Group 1 consisted of 1508 coffee plants, corresponding to 39.60% of the coffee plants in the field, and Group 2 was composed of 2300 plants, corresponding to 60.39% of the total coffee plants in the field. As shown in Figure 6, Group 1 has higher crown diameter values than Group 2. However, the plant height values are similar among the observations comprising Group 2.

2. *Area 2*

As mentioned earlier, the number of clusters was set to 2, and there was a clear division between the groups, group 1 in red and group 2 in blue. Extracting the information on plant height and canopy diameter made it possible to identify in area 2 important characteristics for the clear separation of coffee plants, corroborating the data obtained in Figure 7, where it is possible to see that group 1 is more concentrated in the right quadrant and group 2 in the left quadrant.

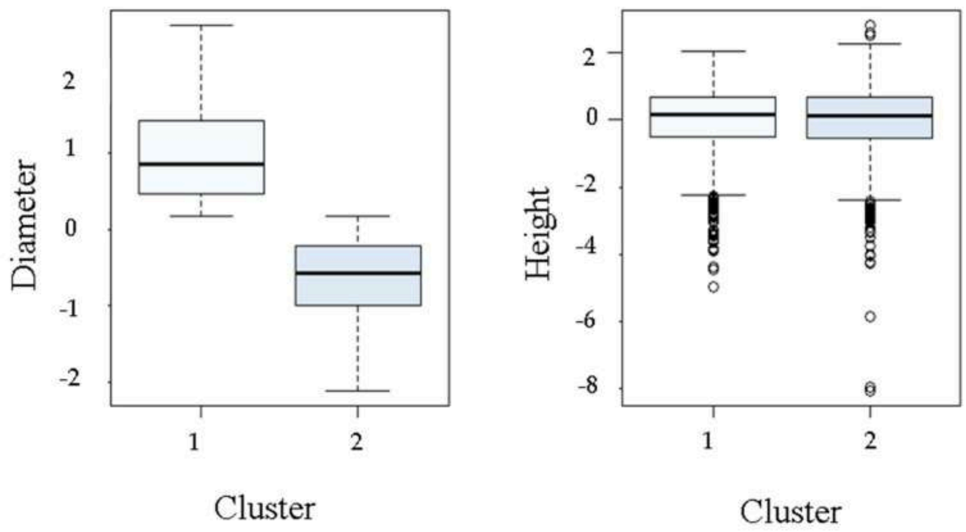

Figure 6. Boxplot of the canopy diameter and plant height variables in the Area 1 clusters.

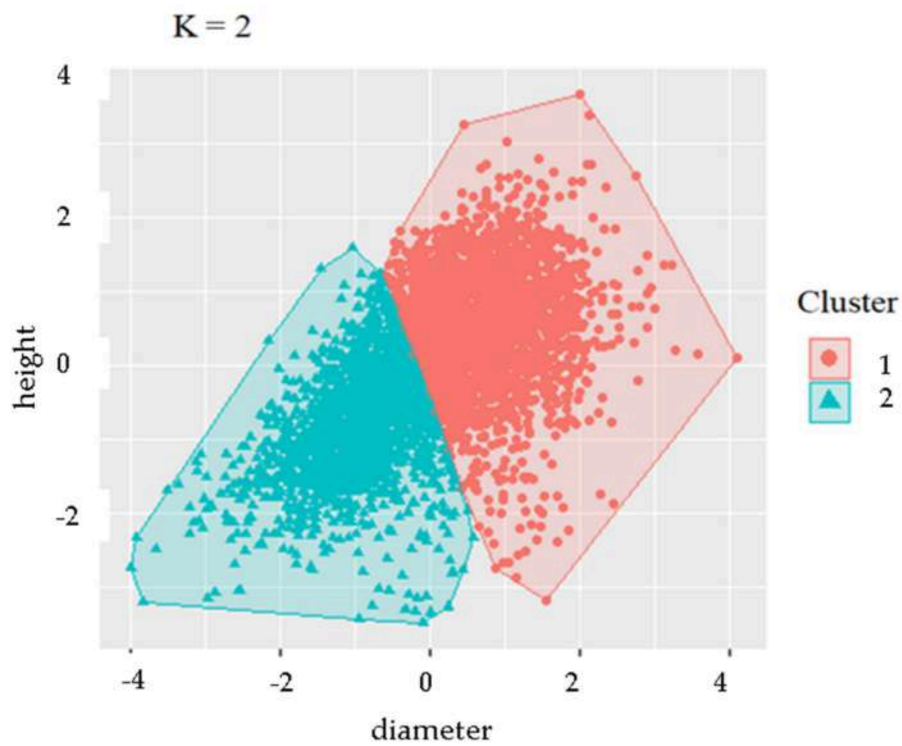

Figure 7. Scatter plot with clustering related to Area 2.

Group 1 was made up of 2307 observations and group 2 of 2385. In Figure 8, it can be seen that group 1 has higher values for both canopy diameter and plant height.

3. *Area 3*

From the analysis of Figure 9, it can be observed that there was a very clear division between the groups, with group 1 concentrated in the right corner and group 2 in the left corner. Group 1 consists of 2027 observations and group 2 of 2560. In this grouping, it can be observed that group 1 had a smaller portion of individuals than group 2.

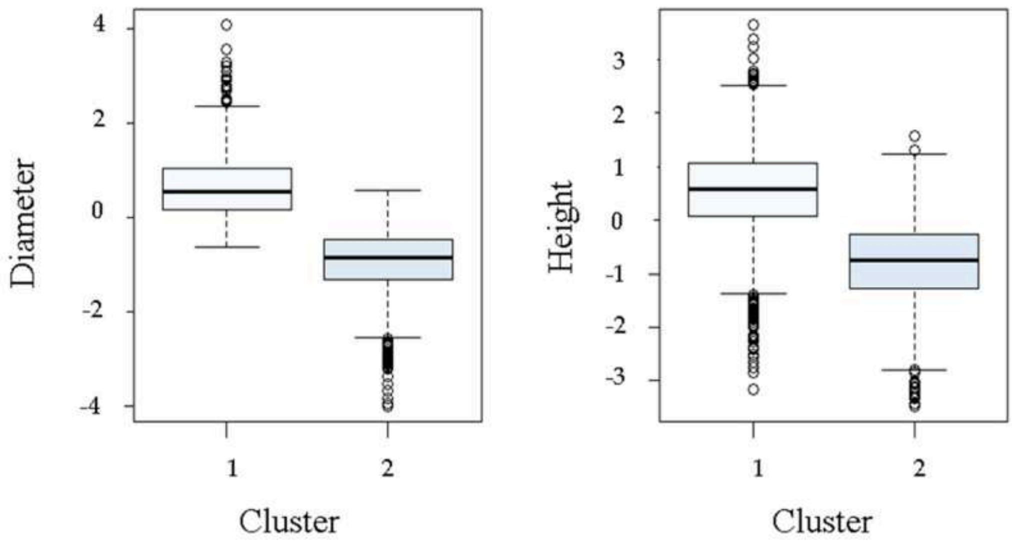

Figure 8. Boxplot of crown diameter and plant height variables in the clusters of Area 2.

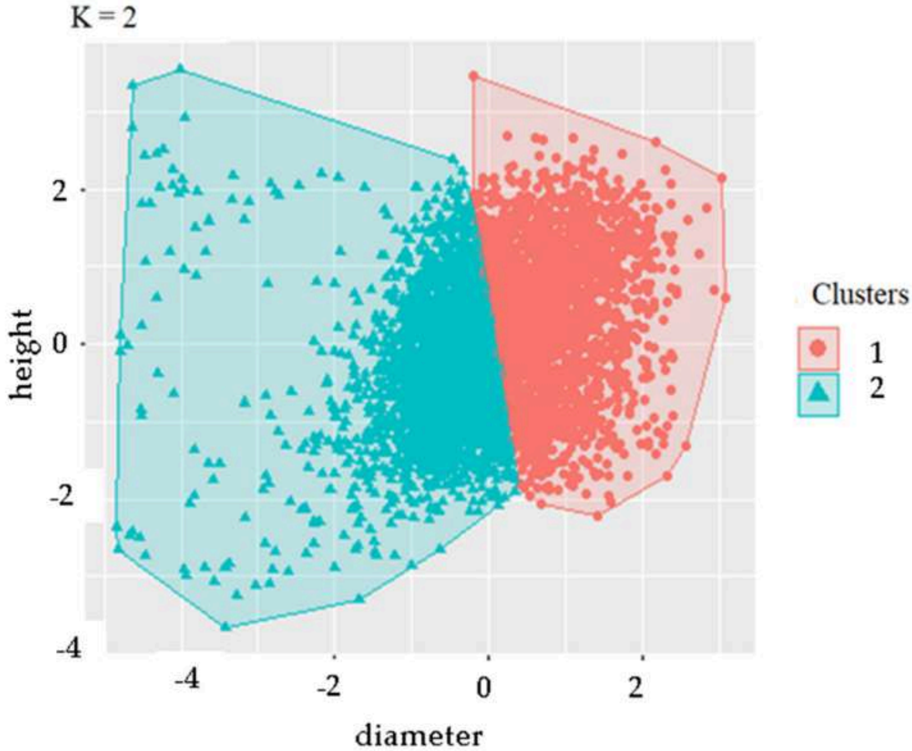

Figure 9. Scatter plot with clustering related to Area 3.

It is observed that both plant height and crown diameter were greater in group 1, as shown in Figure 10. Some studies suggest that, in contrast, during periodic frosts, plants may "prepare" for their annual occurrence through a gradual acclimation process of their physiological mechanisms [2]. These authors also state that this phenomenon promotes plant adaptation to adverse conditions, allowing for a more efficient recovery after exposure to thermal stress caused by frost.

4. *Zoning of Grouping 1*

In Figure 11a,c,e, it can be observed that it is possible to identify more accurately the areas that were affected by frost and this allows the adoption of strategies for the recovery of these areas with greater precision, considering that the identification was carried out using a high-capacity information-gathering technology (LiDAR Sensor) and statistical analyses.

The study's methodology made it possible to extract information about the characteristics of the plants, namely, plant height and crown diameter, enabling the producer to strategically plan agricultural practices based on plant characteristics, height, and diameter. Among the practices that can benefit from this methodology are fertilization, spraying, and those necessary for the recovery of the planted area, or even planning for the recovery of the coffee plantation. In Figure 11b,d,f, the mapping based on the methodology proposed by [27] in areas 1, 2, and 3 can be observed. When looking at Figure 11 and comparing the two methodologies under study, it is noticeable that there is a difference between the two maps, as seen in the green and yellow regions. It is observed that the map proposed by the methodology of this work (Figure 11a,c,e) appears more heterogeneous, being more precise in the characteristics of the plants, since it is based only on plant height and coffee plant canopy diameter, thus being more susceptible to the variability of the plant's morphological state and allowing for a less limited study of the effects of frost.

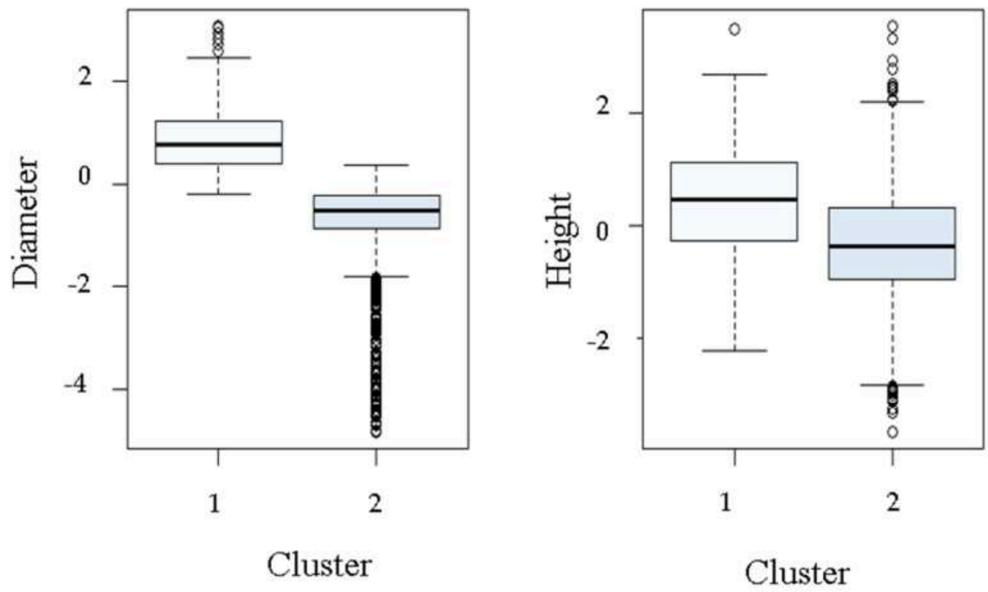

Figure 10. Boxplot of crown diameter and plant height variables in the clusters of Area 3.

- *Grouping 2*

1. *Area 1*

In Figure 12, it can be observed that the same method was applied to separate the groups but now using the elevation variable in grouping 2 as the third variable. The new multidimensional space generated caused the separation of the clusters to change, while maintaining the main division structure. The addition of the elevation variable provides greater refinement in the group data, as it allows for a more precise definition of the convex regions. Despite this, the overlap of the clusters suggests a complex, unexplained relationship. Thus, the addition of elevation did not prove to be a determining factor, but it reduced the internal variability of the proposed groups.

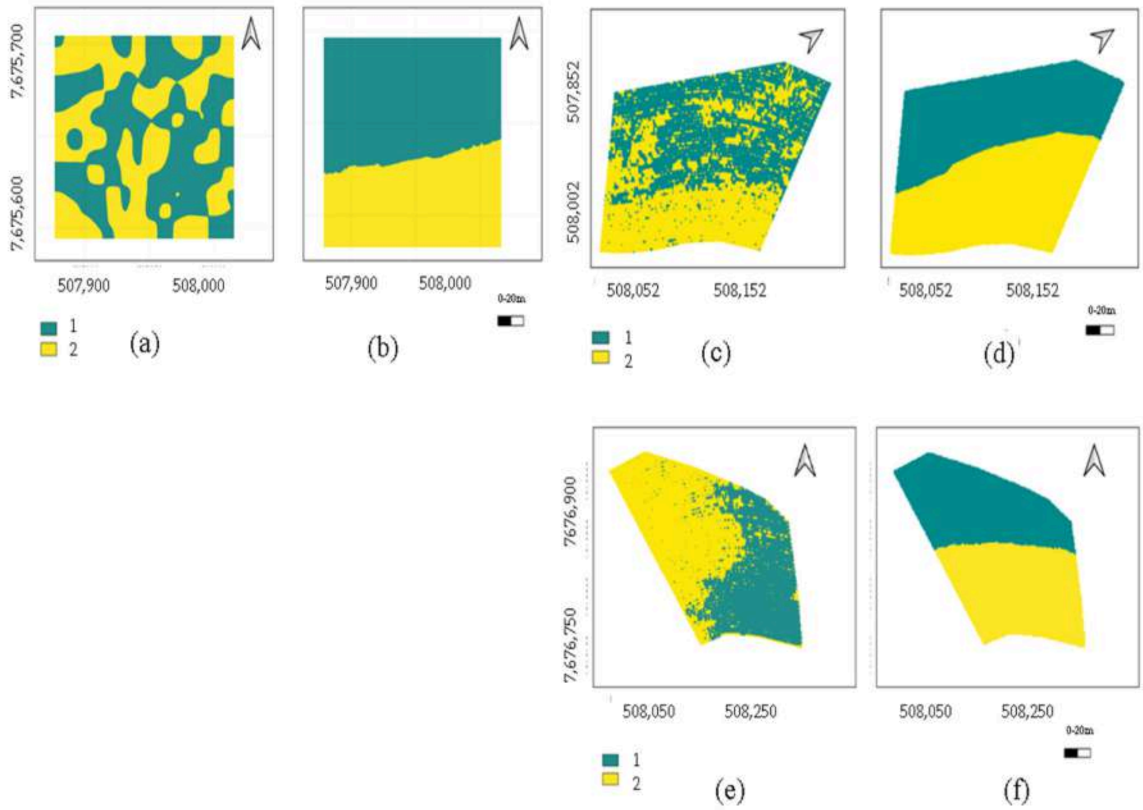

Figure 11. Maps developed using the methodology of this work according to plant height and crown diameter (**a**,**c**,**e**) and map developed based on the methodology proposed by [27] (**b**,**d**,**f**).

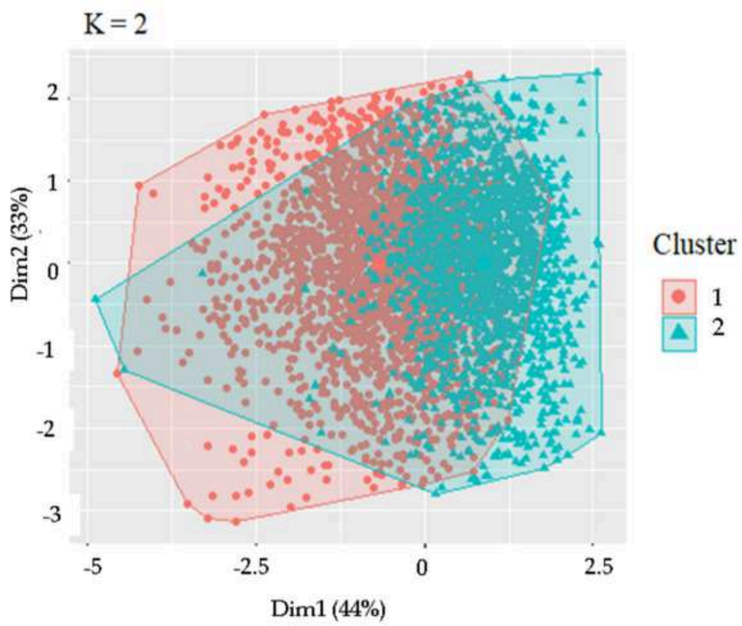

Figure 12. Scatter plot with clustering performed in Area 1.

The division between the groups was not very clear, as can be seen in Figure 12, which can be explained by the addition of the elevation variable. Group 1 consists of 2146 observations, corresponding to 33% of the data, and group 2 consists of 1662 observations, corresponding to 44% of the data. In Figure 10, the results of the cluster analysis (K = 2) can be observed, and the gap between −2.5 and 0.0 indicates a clear separation between the groups, suggesting one cluster in the negative values (e.g., −5.0; −2.5) and another in the positive values (0.0; 2.5) [45].

Groups 1 and 2 have approximate median values for crown diameter and plant height, with group 1 showing lower values compared to group 2. For 10-year-old plants, it was observed that the most affected plants were located in areas characteristic of lowland valleys. According to [46], the critical influence of topographic characteristics in defining areas susceptible to frost damage is highlighted. This susceptibility is intrinsically linked to specific local variables, such as minimum temperature limits, plant phenology, and elevation. The number of frost occurrences also becomes a determining factor in the recovery of coffee plants. According to [47], thermal availability has a direct influence on plant phenology. Higher temperatures accelerate plant development, while lower temperatures prolong the cycle. If the annual temperature fluctuation is pronounced, with a harsh winter, many perennial species enter a resting period (dormancy), returning to the annual vegetative cycle as soon as thermal conditions become suitable, according to Figure 13.

2. *Area 2*

The original data considered in the study included three variables: plant height, crown diameter, and altitude. The dataset contains 4692 observations in the population. In Figure 14, it can be observed that Group 1 is more concentrated in the lower corner, while Group 2 is located in the upper corner, highlighting the separation between the groups based on the variables analyzed. It is also noted that the same method was applied to separate the groups of 3-year-old coffee plants but now using the altitude variable in grouping 2 as a third variable. The new multidimensional space generated caused the separation of the clusters to change, while maintaining the main division structure. The addition of the altitude variable provides greater refinement in the data for the groups, as the convex regions are better defined. Despite this, the overlap of the clusters suggests a complex, unexplained relationship. Thus, the addition of altitude did not prove to be a determining factor, but it did reduce the internal variability of the groups proposed in the clustering.

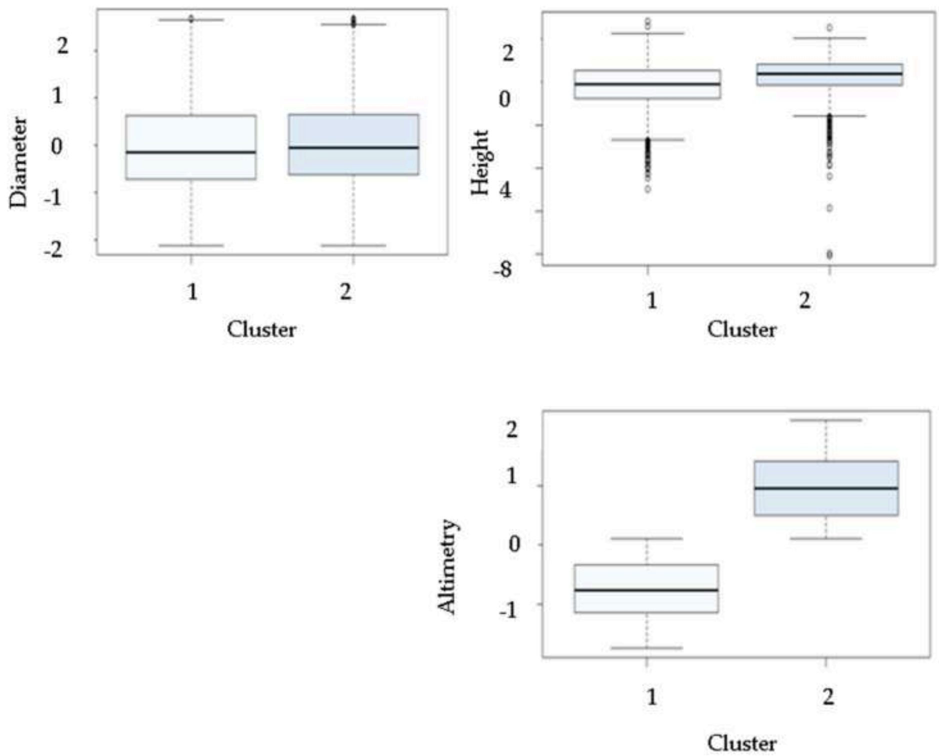

Figure 13. Boxplot of the variable height, crown diameter, and altimetry in the clusters of Area 1.

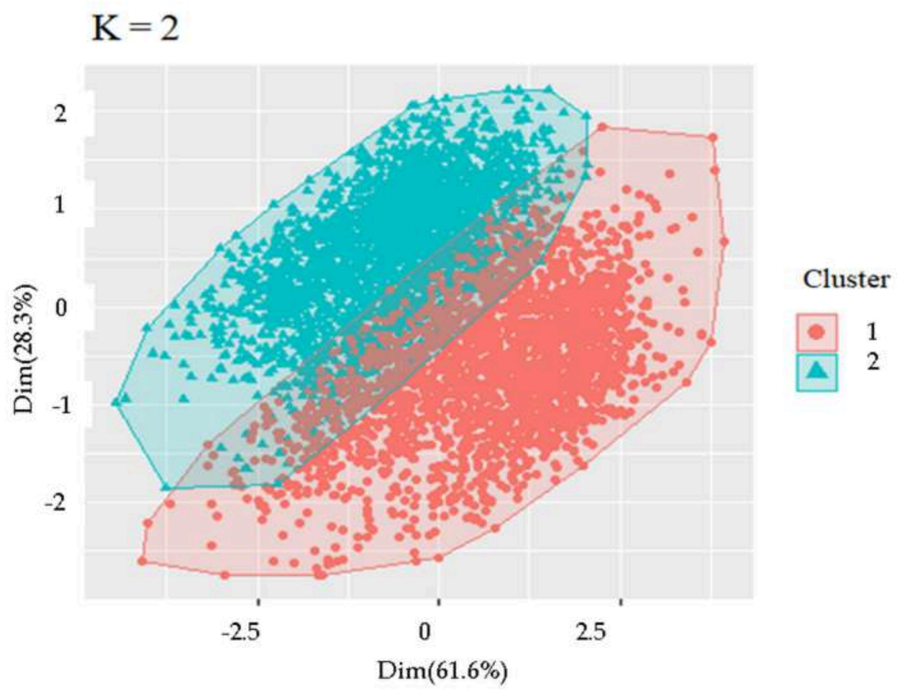

Figure 14. Scatter plot with clustering performed in Area 2.

Group 1 is composed of 2307 observations and group 2 of 2385. For the analysis of variance to compare means, a margin of error of 5% and a 95% confidence level were considered. As observed in Figure 15, group 1 obtained higher values for crown diameter, plant height, and altitude, respectively, when compared to group 2.

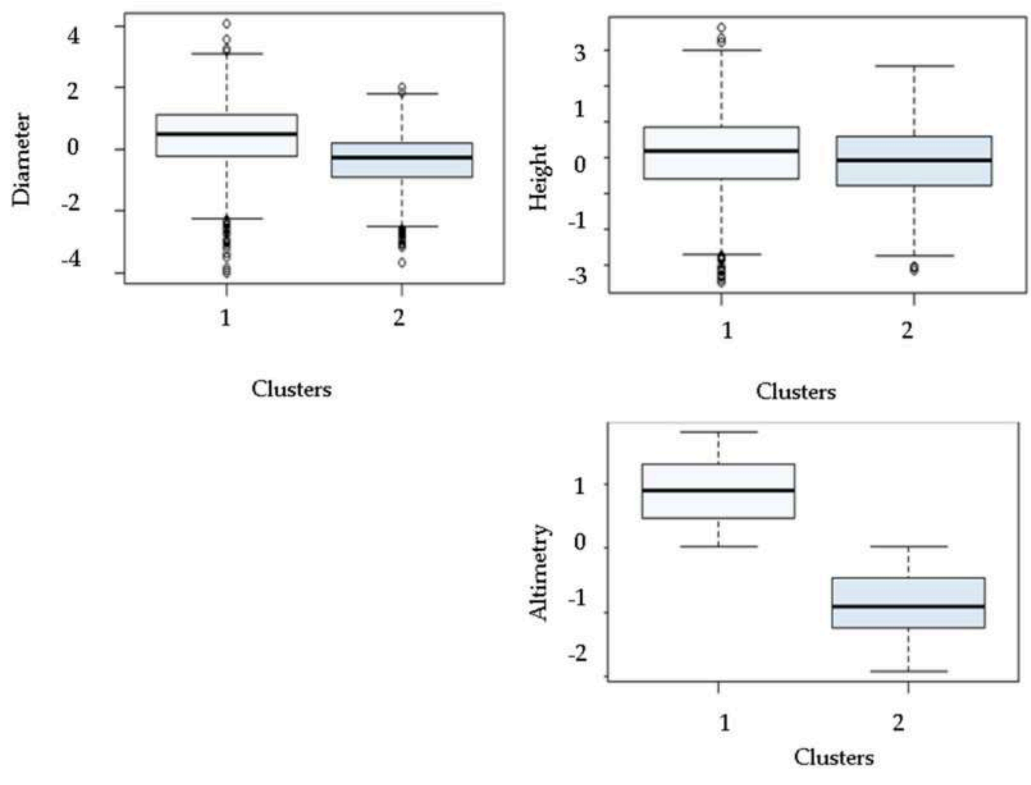

Figure 15. Boxplot of the variables height, crown diameter, and elevation in the clusters of Area 2.

3. *Area 3*

The data considered in the study, therefore, contained 4586 observations, equally distributed between groups 1 and 2. The division between the groups was not very clear, as shown in Figure 16. It is observed that the same method was applied to separate the groups, but now using the altitude variable in grouping 2, for 4-year-old plants. The new multidimensional space generated caused the cluster separation to change, while maintaining the main division structure. The addition of the altitude variable provides greater refinement in the group data, as the convex regions are better defined. Despite this, the overlap of the clusters suggests a complex, unexplained relationship. Thus, the addition of altitude did not prove to be a determining factor, but it did reduce the internal variability of the proposed groups.

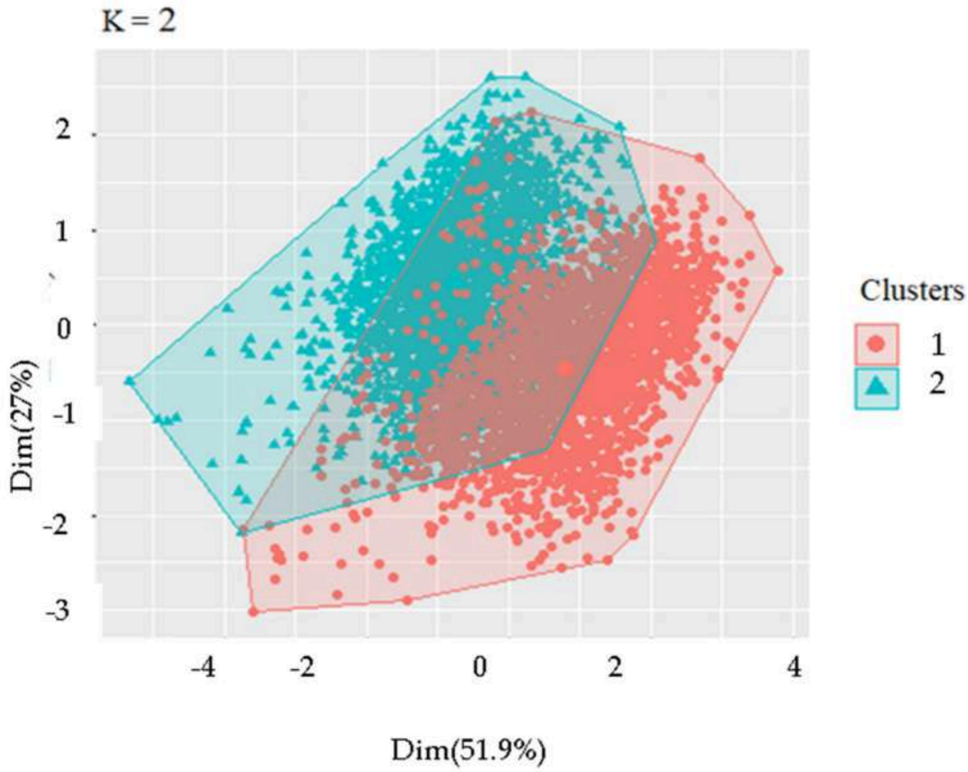

Figure 16. Scatter plot with clustering performed in Area 3.

Group 1 is composed of 2280 observations and group 2 of 2307. As observed in Figure 17, group 1 had higher values for crown diameter, plant height, and altitude compared to group 2. In this grouping methodology, it can be seen that group 2 had more individuals than group 1.

4. *Zoning of Grouping 2*

In the methodology proposed by the study, based on the analysis of the maps of the groups generated by the non-hierarchical clustering of coffee plants in the field in Figure 18a, it was observed that the characteristics of plant height and population diameter showed greater similarity when the variable elevation was added to the map representation, compared to the classification proposed by [27] (Figure 18b). In the classification proposed by this study, a difference is observed at the boundary separating group 1 and group 2, smoothing the borders when compared to the classification proposed by [27], while maintaining an approximate pattern between them. Based on the analysis of the maps of the groups generated by the non-hierarchical clustering of the coffee plants in the field in Figure 18c, it was observed that the characteristics of plant height and population diameter appeared more homogeneous compared to the classification proposed by [27], as shown in Figure 18d. In the classification proposed in this work, a difference can be seen in the boundary separating Groups 1 and 2, smoothing the edges while still following an approximate pattern of the classification proposed by [27]. This approach reinforces the methodology of the aforementioned author, maintaining the logic of dividing the groups while improving the definition of the regions, providing a more precise analysis of the areas affected by frost. The northern regions in green were more affected by frost incidence, while the southern regions in yellow were less affected For four-year-old coffee plants, based on the analysis of the group maps generated by the non-hierarchical clustering of the plants in the field, according to Figure 18e, it was observed that the characteristics of plant height and canopy diameter were more similar compared to the classification proposed by [27] (Figure 18f). The methodology proposed in this study promoted a classification that demonstrated a difference in the location separating groups 1 and 2, softening this boundary line, but following an approximate pattern in relation to the methodology proposed by [27].

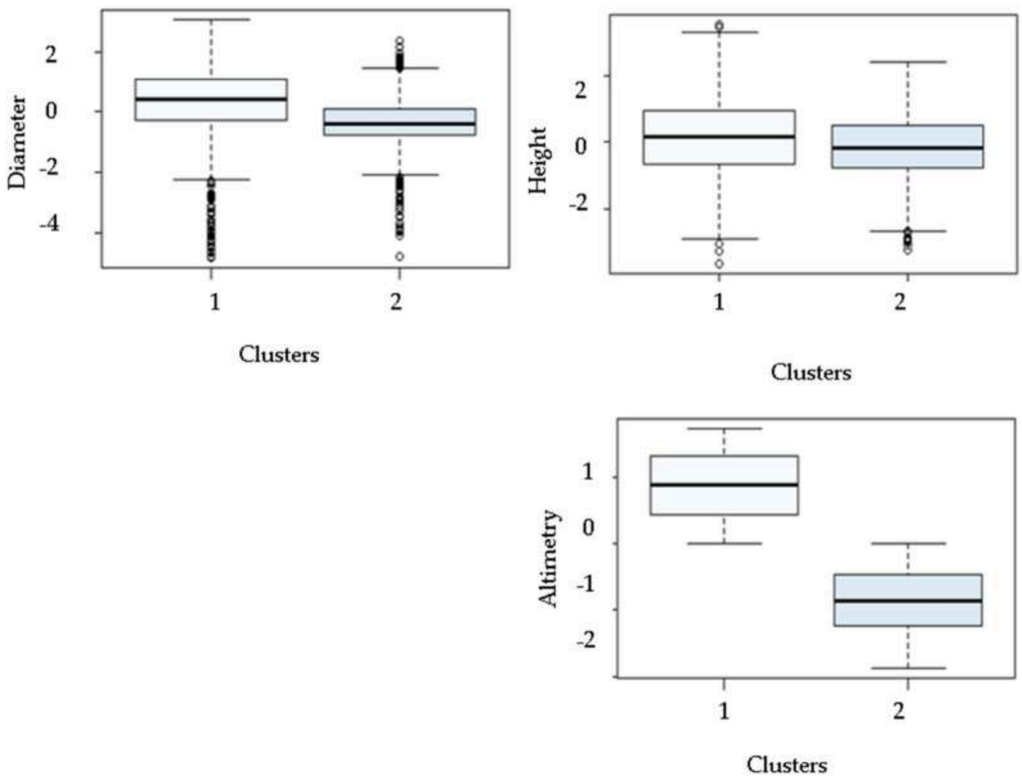

Figure 17. Boxplot of the variables height, crown diameter, and altimetry in the clusters of Area 3.

Plant height is considered an interesting and useful tool for different studies, as it allows the determination of various models, including those for predicting coffee yield and the leaf area index of the population. The same applies to leaf area size, since frost has the ability to reduce the plant's leaf area through necrosis, thus causing significant reductions in its development [48].

Thus, it can be observed that the classification carried out with the data obtained through the LiDAR sensor, together with the non-hierarchical clustering, identified the two regions, with plants showing higher and lower incidence, similar to the classification proposed by [27]. The northern regions in green were affected by frost incidence, and the southern regions in yellow were less affected. Therefore, the most suitable regions to start planning management and recovery would be the green zones, located in the north, in both classifications. The variations observed in the areas analyzed in this study indicate a direct influence of climatic and topographic conditions on the development of coffee plants. The occurrence of frosts between 2019 and 2021 affected plant growth and reduced leaf area, as already reported in other studies [27,49]. Plants in lower regions showed greater damage due to the accumulation of cold air, corroborating findings from published literature [22,50]. According to the literature, gradual acclimatization allows plants exposed to periodic frosts to develop physiological mechanisms of adaptation and tolerance, enabling more efficient recovery after exposure to thermal stress caused by frost [2]. This recovery capacity is more evident in young plants, as they have a more efficient enzymatic metabolism for this purpose [17]. The use of LiDAR data combined with non-hierarchical clustering proved to be effective in identifying areas affected by frost with high spatial accuracy. The proposed methodology allowed for distinguishing zones with different levels of impact and can guide the management and recovery of coffee plantations. Adding the altitude variable refined the delineation of the groups, although it was not decisive for the main separation. The maps resulting from this work indicate that the most affected regions are located to the north of the studied areas, suggesting that these zones should be prioritized for management and replanting. Thus, the method reduces the need for field inspections and provides technical support for agricultural planning, aligning with precision agriculture practices and the efficient use of geospatial data [1,48]. Thus, the methodology enables more effective data collection, reducing the need for multiple field visits as currently carried out by the agricultural producer.

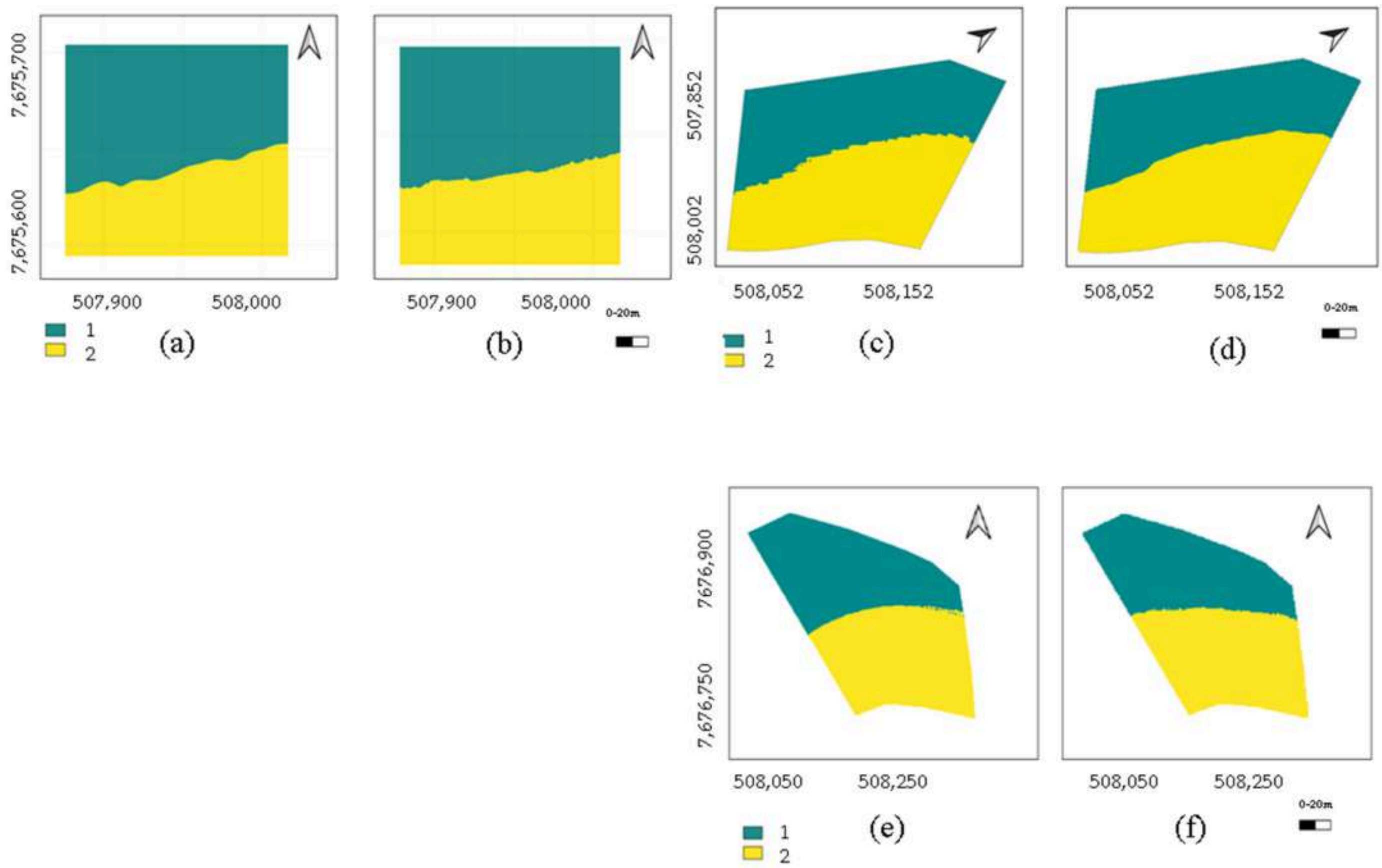

Figure 18. Maps developed using the methodology of this work according to plant height (**a**), crown diameter (**c**), and elevation (**e**), and corresponding maps developed based on the methodology proposed by [27] (**b**,**d**,**f**).

The classification of frost risk areas is essential for agricultural planning and the mitigation of crop damage, especially in production systems sensitive to microclimatic variations. Traditionally, as described in previous studies [23,27], this classification has been based on visual criteria associated with topography, considering factors such as terrain slope, landscape position, and field observations. These criteria are relevant because topography directly influences cold air dynamics, favoring its accumulation in lower areas and increasing the risk of frost occurrence. However, this approach presents limitations related to the subjectivity of visual assessment and the dependence on the evaluator's experience, which may introduce variability and reduce the reproducibility of the results.

In this context, the present study proposes an alternative approach based on non-hierarchical clustering, using population data from affected plants, biometric characteristics, and environmental variables such as altitude. This methodology represents a significant advancement, as it enables a more objective classification based on quantitative data. The use of clustering techniques allows the identification of natural patterns in the data, grouping areas with similar characteristics in terms of plant response to frost stress, without relying exclusively on visual interpretation. Furthermore, the incorporation of statistical validation of cluster 1 and cluster 2 strengthens the robustness of the results, ensuring that the classification reflects real and significant differences between the evaluated areas.

Another relevant aspect of this approach is the integration of plant physiological variables and environmental factors, enabling a more comprehensive understanding of crop responses to frost. Unlike purely topographic analyses, which primarily consider geomorphological factors, the use of population-based data allows the direct capture of the climatic event's impact on plants, making the classification more representative of actual field conditions. This is particularly important in precision agriculture scenarios, where decision-making depends on detailed and spatially explicit information.

Additionally, the use of statistical methods contributes to reducing subjectivity and increasing the reliability of the classification, allowing its application across different areas and conditions. This approach also facilitates integration with remote sensing technologies and Geographic Information Systems (GIS), enabling the generation of more accurate frost risk maps and supporting agricultural management, as proposed in this study. As a practical implication, the results may support mitigation strategies, such as planting planning, the selection of less susceptible areas, and the adoption of protective measures.

Finally, although the clustering-based approach provides greater robustness and objectivity, its effectiveness depends on the quality and representativeness of the collected data. Therefore, future studies may explore the integration of additional climatic variables, such as minimum temperature, humidity, and cold air dynamics, in order to further improve classification accuracy. In this way, the combination of statistical methods with agronomic and environmental data represents a promising tool for frost risk zoning and for strengthening the resilience of agricultural systems.

The statistical analysis revealed that, in Area 1, plant height was the only variable that exhibited a significant difference between the evaluated clusters, with a p-value below the 5% significance level ($p < 0.05$ and $p < 2.2 \times 10^{-16}$). This result provides strong evidence to reject the null hypothesis of equal group means, confirming that plant height was a key discriminating factor in the differentiation of the identified clusters. Plant height is a critical indicator of vegetative vigor and reflects the physiological response of plants to stress conditions, including frost events. More developed plants may be either more exposed to atmospheric conditions or, alternatively, exhibit greater recovery capacity, depending on the environmental context. Therefore, the identification of this variable as statistically significant indicates that plant height plays a central role in characterizing structural variability among groups.

https://doi.org/10.3390/horticulturae12080904

Furthermore, the analysis of areas composed of 3- and 4-year-old plants showed that plant height, plant volume, and leaf area index (LAI) presented distinct median values between Groups 1 and 2, reinforcing the presence of structural and physiological differences between clusters. These variables are directly associated with plant architecture and photosynthetic capacity, reflecting the level of development and overall physiological condition. LAI, in particular, is widely used as an indicator of radiation interception capacity and productive potential and is also sensitive to abiotic stress. Consequently, significant differences in these variables indicate that the groups represent populations with distinct physiological characteristics, potentially associated with different levels of exposure and susceptibility to frost effects.

In contrast, canopy diameter did not show significant differences between groups in Area 1, indicating that this parameter alone may not be sufficiently sensitive to discriminate stress effects or structural differences among plants. This finding suggests that, although canopy diameter is an important component of plant architecture, it may exhibit lower relative variability or reduced sensitivity compared to variables such as height and volume, particularly at specific developmental stages or under certain environmental conditions.

In Areas 2 and 3, the non-hierarchical clustering analysis considering Clusters 1 and 2 produced consistent and even more pronounced results, with significant differences detected in all analyzed variables between groups. This pattern reinforces the robustness of the clustering methodology, demonstrating its ability to identify distinct structural patterns within the plant population. The consistency of results across different areas indicates that the method is sufficiently sensitive to capture real structural variation, reflecting differences in plant development and potentially in response to environmental stress. The application of the parsimony principle further supports the adequacy of the adopted model, as it effectively explains the observed variability without introducing unnecessary complexity.

By contrast, the original clustering proposed by [17] did not detect significant differences between groups, highlighting limitations in its ability to capture the variability present in the dataset. This result suggests that the original method may not have been sufficiently sensitive or may have relied on criteria that do not directly reflect the physiological and structural differences among plants. The absence of significant differences in that model reinforces the superiority of the non-hierarchical clustering approach employed in this study, which demonstrated greater discriminative power and stronger alignment with patterns observed in the biometric variables.

The optimal number of clusters was determined using complementary cluster validation methods, including the Elbow Method, the Silhouette Index, and the Gap Statistic, which are widely recommended in the clustering literature to evaluate cluster compactness, separation, and structural robustness [44,51–53]. The three clustering validation methods consistently indicated the two-cluster solution (k = 2) as the most appropriate (Figure 19). The Elbow Method showed a sharp reduction in within-cluster variability between (k = 1) and (k = 2), followed by stabilization for higher values of (k). Similarly, the Silhouette Index presented its highest average value for (k = 2), indicating better cohesion and separation between groups. The Gap Statistic also showed its maximum value at (k = 2), demonstrating that the two-cluster structure exhibited the greatest difference compared to a random reference distribution. Furthermore, the two-group solution explained approximately 53.2% of the total data variability (BSS/TSS), reinforcing the adequacy of the adopted partition.

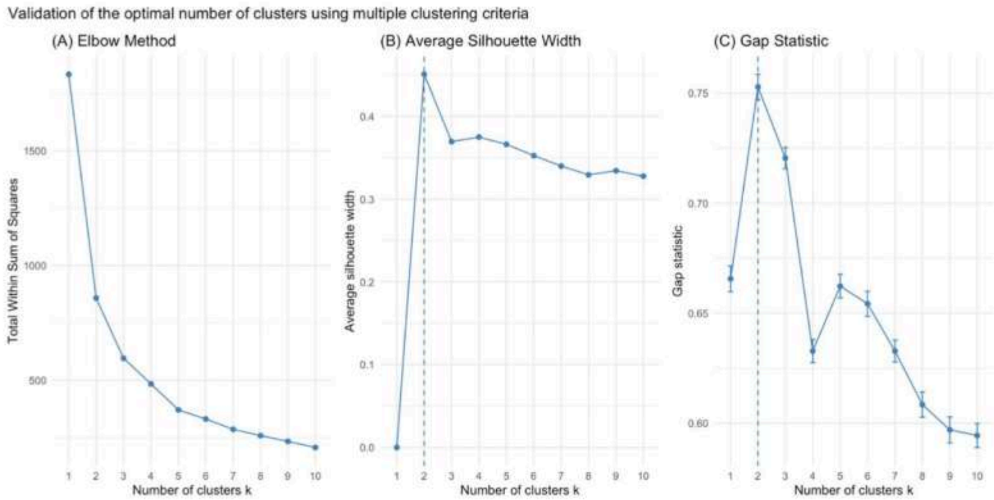

Figure 19. Methods used for clustering validation and determination of the number of clusters.

Overall, the results confirm that the proposed methodology was effective in identifying statistically distinct groups, enabling the definition of differentiated structural profiles within the evaluated areas. The ability to detect differences in variables directly related to plant vigor and development highlights the potential of this approach for agricultural monitoring and for identifying areas with varying levels of susceptibility to environmental stress, such as frost.

The proposed study did not include an individual classification of frost damage severity for each plant, which precluded the construction of a confusion matrix and the calculation of performance metrics such as accuracy, F1-score, and Cohen's kappa coefficient. Therefore, cluster validation was based on the consistency between the structural patterns detected by LiDAR across the population of frost-affected plants within the experimental areas, the agronomic field observations, and the internal clustering validation indices (Elbow Method, Silhouette Index, and Gap Statistic).

These findings contribute to advancing data-driven classification techniques, promoting a more objective and precise assessment of spatial variability in agricultural systems.

4. Conclusions

This study made it possible to develop and test an innovative methodology for mapping frost occurrence zones, using data obtained from LiDAR sensors mounted on UAVs. This methodology was based on data clustering techniques and was able to efficiently identify and classify areas with different levels of frost impact, meeting the first objective of the research. The results showed that groups of coffee plants affected by frost exhibited statistically significant differences in the analyzed variables: plant height, canopy diameter, plant volume, and Leaf Area Index (LAI). These data highlight the potential of the proposed approach to detect morphological changes associated with thermal stress caused by frost. Finally, by using data generated by LiDAR sensors mounted on RPAs in the clustering process, the applicability of this technology in post-frost monitoring was demonstrated, allowing for rapid, large-scale analyses with a high level of detail. The developed methodology proved to be superior to approaches found in the literature, representing a significant advancement for precision and digital coffee cultivation and contributing to the development of more effective strategies for mitigating and managing the impacts of climate on coffee crops.

Author Contributions: Conceptualization, G.A.e.S.F.; Data Curation, A.L.V.A. and L.L.C.; Formal analysis A.L.V.A. and L.L.C.; Funding Acquisition, G.A.e.S.F. and R.d.O.F.; Investigation A.L.V.A., L.L.C. and R.M.A.d.M.; Methodology, G.A.e.S.F., R.d.O.F., G.R. and M.B.; Project Administration, G.A.e.S.F.; Validation, A.L.V.A. and R.M.A.d.M.; writing—Original Draft Preparation, A.L.V.A. and G.A.e.S.F.; Writing—Review and Editing, A.L.V.A., G.A.e.S.F., R.M.A.d.M., G.R. and M.B.; Visualization, A.L.V.A. and G.A.e.S.F.; Supervision, G.A.e.S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by EMBRAPA Café—Coffee Research Consortium, Project No. 10.18.20.041.00.00; by CNPq (National Council for Scientific and Technological Development), Projects No. 310186/2023-4 and 404462/2025-1; by the Minas Gerais Research Support Foundation (FAPEMIG) Project APQ-00661-22; and by CAPES (Coordination for the Improvement of Higher Education Personnel—Finance Code 001), through the provision of scholarships.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors upon request.

Acknowledgments: The authors acknowledge EMBRAPA Café and CNPq (National Council for Scientific and Technological Development) for funding the project; CAPES (Coordination for the Improvement of Higher Education Personnel) for the provision of scholarships; the Graduate Program in Agricultural Engineering (PPGEA) at the Federal University of Lavras (UFLA); the Nucleos of Agricultural Systems Studies (NESA); and the Federal University of Lavras and University of Florence (UniFI), for institutional and technical support in the development of this work.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Favarin, J.L.; Dourado Neto, D.; García y García, A.; Villa Nova, N.A.; Favarin, M.D.G.G.V. Equações para a estimativa do índice de área foliar do cafeeiro. *Pesqui. Agropecuária Bras.* **2002**, *37*, 769–773. [CrossRef]
2. Sakai, A.; Larcher, W. *Frost Survival of Plants: Responses and Adaptation to Freezing Stress*; Springer: Berlin, Germany, 1987.
3. Marin, D.B.; Alves, M.D.C.; Pozza, E.A.; Gandia, R.M.; Cortez, M.L.J.; Mattioli, M.C. Sensoriamento remoto espectral na identificação e mapeamento de variáveis bióticas e abióticas do cafeeiro. *Rev. Ceres* **2019**, *66*, 142–153. [CrossRef]
4. Lopes, P.R. A Biodiversidade Como Fator Preponderante para a Produção Agrícola em Agroecossistemas Cafeeiros Sombreados no Pontal do Paranapanema. Ph.D. Thesis, Universidade de São Paulo, São Paulo, Brazil, 2014.
5. Zhang, N.; Wang, M.; Wang, N. Precision agriculture—A worldwide overview. *Comput. Electron. Agric.* **2002**, *36*, 113–132. [CrossRef]
6. Rakočević, M. Coffee Plant Architecture. In *Advances in Botanical Research*; Elsevier: Amsterdam, The Netherlands, 2025; Volume 114, pp. 29–61. [CrossRef]
7. Murchie, E.H.; Burgess, A.J. Casting Light on the Architecture of Crop Yield. *Crop Environ.* **2022**, *1*, 74–85. [CrossRef]
8. Omasa, K.; Hosoi, F.; Konishi, A. 3D LiDAR Imaging for Detecting and Understanding Plant Responses and Canopy Structure. *J. Exp. Bot.* **2007**, *58*, 881–898. [CrossRef] [PubMed]
9. Hosoi, F.; Omasa, K. Factors Contributing to Accuracy in the Estimation of the Woody Canopy Leaf Area Density Profile Using 3D Portable LiDAR Imaging. *J. Exp. Bot.* **2007**, *58*, 3463–3473. [CrossRef] [PubMed]
10. Rakočević, M.; Ribeiro, R.V. Soil Ca_2SiO_4 Supplying Increases Drought Tolerance of Young Arabica Coffee Plants. *Plants* **2025**, *14*, 3666. [CrossRef] [PubMed]
11. Rosísca, J.R.; Nagashima, G.T.; de Oliveira, C.M.G.; Morais, H.; Sera, G.H.; Caramori, P.H.; de Aguiar, M.A. Parâmetros associados à resistência de genótipos de café a temperaturas negativas. *Semin. Ciênc. Agrár.* **2022**, *43*, 2293–2308. [CrossRef]
12. Kriegel, H.-P.; Kröger, P.; Zimek, A. Clustering high-dimensional data. *ACM Trans. Knowl. Discov. Data* **2009**, *3*, 1. [CrossRef]
13. Sasaki, T.; Imanishi, J.; Ioki, K.; Morimoto, Y.; Kitada, K. Estimation of leaf area index and canopy openness in broadleaved forest using an airborne laser scanner in comparison with high-resolution near-infrared digital photography. *Landsc. Ecol. Eng.* **2008**, *4*, 47–55. [CrossRef]
14. Souza, V.Q.D.; Caron, B.O.; Schmidt, D.; Behling, A.; Bamberg, R.; Vian, A.L. Resistência de espécies arbóreas submetidas a extremos climáticos de geada em diferentes sistemas agroflorestais. *Ciênc. Rural* **2011**, *41*, 972–977. [CrossRef]
15. Omia, E.; Bae, H.; Park, E.; Kim, M.S.; Baek, I.; Kabenge, I.; Cho, B.K. Remote sensing for monitoring field crops: A comprehensive review of sensor systems, data analysis, and recent advances. *Remote Sens.* **2023**, *15*, 354. [CrossRef]

16. Damatta, F.M.; Ramalho, J.D.C. Impactos da seca e do estresse térmico na fisiologia e produção do café: Uma revisão. *Braz. J. Plant Physiol.* **2006**, *18*, 55–81.
17. Costa, A.C.d. Efeito de Irradiâncias Contrastantes nas Características Fotossintéticas e na Atividade de Enzimas Antioxidantes em Plantas Jovens de Chichá. Master's Thesis, Instituto Federal Goiano, Rio Verde, Brazil, 2013.
18. Panagiotidis, D.; Abdollahnejad, A.; Surový, P.; Chiteculo, V. Determining tree height and crown diameter from high-resolution UAV imagery. *Int. J. Remote Sens.* **2016**, *3*, 2392–2410. [CrossRef]
19. Maes, W.H.; Steppe, K. Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. *Trends Plant Sci.* **2019**, *24*, 152–164. [CrossRef] [PubMed]
20. Hunt, E.R., Jr.; Daughtry, C.S.T. What Good Are Unmanned Aircraft Systems for Agricultural Remote Sensing and Precision Agriculture? *Int. J. Remote Sens.* **2018**, *39*, 5345–5376. [CrossRef]
21. Wang, J.; Zhang, S.; Lizaga, I.; Zhang, Y.; Ge, X.; Zhang, Z.; Zhang, W.; Huang, Q.; Hu, Z. UAS-Based Remote Sensing for Agricultural Monitoring: Current Status and Perspectives. *Comput. Electron. Agric.* **2024**, *227*, 109501. [CrossRef]
22. Camargo, M.B.P.; Santos, M.A.; Pedro, M.J., Jr.; Fahl, J.I. Modelo agrometeorológico para monitoramento e predição da produtividade do café (*Coffea arabica* L.) no estado de São Paulo, Brasil. In *Proceedings of the Anais Da 21st International Conference on Coffee Science, Montpellier, França, 10–15 September 2006*; ASIC: Paris, France, 2006; pp. 1125–1131.
23. Valente, G.F.; Ferraz, G.A.E.S.; Schwerz, F.; Faria, R.O.; Fernandes, F.A.; Marin, D.B. Remotely Piloted Aircraft for Evaluating the Impact of Frost in Coffee Plants: Interactions between Plant Age and Topography. *Remote Sens.* **2024**, *16*, 3467. [CrossRef]
24. Holman, F.H.; Riche, A.B.; Michalski, A.; Ostler, G.; de Sa, V.G.O.; Brown, J.; Mellers, G.; Smith, M. High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sens.* **2016**, *8*, 1031. [CrossRef]
25. Santos, L.M.D.; Ferraz, G.A.E.S.; Carvalho, M.A.F.; Teodoro, S.A.; Campos, A.A.V.; Menicucci Neto, P. Utilização de imagens RPA no mapeamento do índice de clorofila de café. *Sustainability* **2022**, *14*, 13118.
26. Wang, A.; Zhang, W.; Wei, X. A review on weed detection using ground-based machine vision and image processing techniques. *Comput. Electron. Agric.* **2019**, *158*, 226–240. [CrossRef]
27. Marin, D.B.; Faria, R.O.; Zezzi, G.C.P.; Colaço, A.F.; Santana, L.S.; Ferraz, G.A.E.S.; Molin, J.P. Unmanned Aerial Vehicle to Evaluate Frost Damage in Coffee Plants. *Precis. Agric.* **2021**, *22*, 1845–1860. [CrossRef]
28. Partelli, F.L.; Vieira, H.D.; Viana, A.P.; Batista-Santos, P.; Rodrigues, A.P.; Leitão, A.E.; Ramalho, J.C. Low temperature impact on photosynthetic parameters of coffee genotypes. *Pesqui. Agropecuária Bras.* **2009**, *44*, 1404–1415. [CrossRef]
29. Soffiati, A. *Holoceno: Época Geológica e Antrópica*; Editora Autografia: Rio de Janeiro, Brazil, 2022.
30. Lunz, A.M.P. Crescimento e Produtividade do Cafeeiro Sombreado e a Pleno Sol. Ph.D. Thesis, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, Brazil, 2006.
31. Taiz, L.; Zeiger, E.; Møller, I.M.; Murphy, A. *Fisiologia e Desenvolvimento Vegetal*, 6th ed.; Artmed: Porto Alegre, Brazil, 2017.
32. Kumhálová, J.; Moudrý, V. Topographical characteristics for precision agriculture in the conditions of the Czech Republic. *Geogr. Ann.* **2014**, *50*, 90–98. [CrossRef]
33. Braga, G.B.; Imbuzeiro, H.M.A.; Pires, G.F.; de Oliveira, L.R.; Barbosa, R.A.; Vilela, K.D.F. Frost risk and rural insurance in Brazil. *Rev. Bras. Meteorol.* **2021**, *36*, 703–711. [CrossRef]
34. Al-Issawi, M.; Rihan, H.Z.; El-Sarkassy, N.; Fuller, M.P. Frost Hardiness Expression and Characterisation in Wheat at Ear Emergence. *J. Agron. Crop Sci.* **2013**, *199*, 66–74. [CrossRef]
35. Shammi, S.; Sohel, F.; Diepeveen, D.; Zander, S.; Jones, M.G. A survey of image-based machine learning techniques for frost detection in plants. *Inf. Process. Agric.* **2023**, *10*, 164–191. [CrossRef]
36. Bento, N.L.; Ferraz, G.A.E.S.; Santana, L.S.; Faria, R.O.; Rossi, G.; Bambi, G. Plant height and soil compaction in coffee crops based on LiDAR and RGB sensors carried by remotely piloted aircraft. *Remote Sens.* **2025**, *17*, 1445. [CrossRef]
37. UFV. Avaliação de índice de vegetação na avaliação de danos por geada em cafeeiros de um ano. In *Anais Do 32º Simpósio Mineiro de Pós-Graduação UFV*; UFV: Lavras, Brazil, 2024; pp. 1000–1257. Available online: https://simppea.ufv.br/ (accessed on 15 July 2026).
38. Sosa, L.; Justel, A.; Molina, Í. Detection of crop hail damage with a machine learning algorithm using time series of remote sensing data. *Agronomy* **2021**, *11*, 2078. [CrossRef]
39. Mu, Y.; Zhou, G.; Wang, H. Canopy lidar point cloud data k-means clustering watershed segmentation method. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *6*, 67–73. [CrossRef]
40. Alencar, C.J.; Galo, M.; Dos Santos, R.C. Uso de descritores 3D e intensidade na detecção de árvores em ambiente urbano por meio de dados LiDAR. *Rev. Bras. Cartogr.* **2023**, *75*, 1. [CrossRef]
41. Box, G.E.P.; Cox, D.R. An analysis of transformations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1964**, *26*, 211–252. [CrossRef]
42. Casella, G.; Berger, R.L. *Statistical Inference*, 2nd ed.; Cengage Learning: São Paulo, Brazil, 2014.
43. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [CrossRef]

44. Osborne, J.; Waters, E. Four assumptions of multiple regression that researchers should always test. *Pract. Assess. Res. Eval.* **2002**, *8*, 2.
45. Watson, D.J. The physiological basis of variation in yield. *Adv. Agron.* **1952**, *4*, 101–144. [CrossRef]
46. Oliveira, E.; Silva, F.M.D.; Souza, Z.M.D.; Figueiredo, C.A.P.D. Influência da colheita mecanizada na produção cafeeira. *Ciênc. Rural* **2007**, *37*, 1466–1470. [CrossRef]
47. De Souza, C.M.A.; Ferraz, G.A.E.S.; Marcatti, G.E. Comparação entre derriça manual e mecânica de frutos de cafeeiro. *Ceres* **2006**, *53*, 305.
48. Matiello, J.B.; Almeida, S.R. Deu Geada No Cafezal, Não Façam Podas Já. 2016. Available online: http://www.cccrj.com.br/revista/839/24.pdf (accessed on 10 October 2025).
49. Allevato, E.; Chirici, G.; Vacchiano, G.; Foderi, C.; Lombardi, F.; Rinaldi, M.; D'Ottavio, D.; Cutini, A.; La Marca, O. Canopy damage by spring frost in European beech along the Apennines: Effect of latitude, altitude and aspect. *Remote Sens. Environ.* **2019**, *225*, 431–440. [CrossRef]
50. Barros, R.S.; Magalhães, A.C.; Rodrigues, P.A. Determinação de área de folhas do café (*Coffea arabica* L. cv. 'Bourbon Amarelo'). *Rev. Ceres* **1973**, *20*, 44–52.
51. Sagala, N.T.; Gunawan, A.A.S. Discovering the optimal number of crime cluster using elbow, silhouette, gap statistics, and nbclust methods. *ComTech Comput. Math. Eng. Appl.* **2022**, *13*, 1–10. [CrossRef]
52. Corrar, J.L.; Paulo, E.; Dias Filho, J.M. *Análise Multivariada: Para Cursos de Administração, Ciências Contábeis e Economia*; Atlas: São Paulo, Brazil, 2014.
53. Yuan, C.; Yang, H. Research on K-value selection method of K-means clustering algorithm. *J* **2019**, *2*, 226–235. [CrossRef]

https://doi.org/10.3390/horticulturae12080904