GEOSPATIAL DATA COLLECTION AND INTEGRATION FOR FOREST ECOSYSTEM CHARACTERIZATION IN SUPPORT OF WILDFIRE ANALYSIS

Davis Dinkov, Atanas Kitev, Desislava Hristova, Lyubka Pashova

Assoc. Prof. Davis Dinkov, Ph.D.;
National Institute of Geophysics, Geodesy and Geography -Bulgarian Academy of Sciences;
Acad. Georgi Bonchev Str., Bl. 3, 1113 Sofia, Bulgaria;
+359888312877, davis.dinkov@gmail.com

Assoc. Prof. Atanas Kitev, Ph.D.;
National Institute of Geophysics, Geodesy and Geography - Bulgarian Academy of Sciences;
Acad. Georgi Bonchev Str., Bl. 3, 1113 Sofia, Bulgaria;
+359887787916, atanaskitev@abv.bg

Mgr. Desislava Hristova;
National Institute of Geophysics, Geodesy and Geography - Bulgarian Academy of Sciences;
Acad. Georgi Bonchev Str., Bl. 3, 1113 Sofia, Bulgaria;
+359888583814, dessisslava.hristova@abv.bg

Assoc. Prof. Lyubka Pashova, Ph.D.;
National Institute of Geophysics, Geodesy and Geography - Bulgarian Academy of Sciences;
Acad. Georgi Bonchev Str., Bl. 3, 1113 Sofia, Bulgaria;
+359887538880, lpashova@geophys.bas.bg

Abstract

Forest fires pose an increasing threat to forest ecosystems under the combined influence of climate change, land-use dynamics, and extreme weather events. Reliable forest ecosystem characterization and wildfire analysis require high-quality, harmonized geospatial datasets that accurately reflect the environmental conditions governing fire ignition, spread, and impact. The GeoFireData project aims to systematically collect, integrate, and process geospatial data to support advanced wildfire analysis and future modeling efforts. The first phase of the project focuses on integrating heterogeneous geospatial sources – UAV-derived DEM/DSM, national forest inventory data, Ministry of Agriculture and Food (MoAF) orthophotos and DSM, and specialized layers for agricultural areas from the Land Parcel Identification System (LPIS) – into a unified analytical layer.

The primary objective is to derive a statistically representative sample from the forest area for general ecosystem description and analysis, using GIS tools and statistical methods. Forest stand data layers were processed, including attributes such as forest type, age structure, mean tree height, mean diameter, and canopy height (CH). Terrain characteristics – elevation, slope, and aspect – were also incorporated. Using five stratification criteria (forest type, age class, elevation, slope, and canopy height) and GIS tools, the spatial distribution and characteristics of a representative sample were established for the "Vratsa Balkan" study area (2,132.4 ha). The sample was derived using a stratified random sampling method with proportional area allocation, in which the number of sample plots in each stratum is proportional to the stratum's accessible area within the study zone. The stratification yielded 50 actual combinations, out of a theoretical maximum of 162. A two-stage area-based generalization was applied: strata below 10 ha were merged with the nearest significant stratum while preserving forest type and age class; strata below 20 ha were subsequently merged following the same hierarchical principle, resulting in 12 final strata. Proportional allocation determined 27 sample plots (0.5 ha each). Sample representativeness was verified using a two-sample Kolmogorov-Smirnov test for elevation, slope, and CH distributions, a χ^2 test for the proportional representation of forest types and stand age classes, and a spatial coverage analysis – all results confirmed the representativeness of the sample across all tested indicators.

The GeoFireData approach underlines the critical role of data-driven geospatial infrastructures in wildfire research and management. By focusing on reliable data acquisition and processing, the project contributes to more reliable forest ecosystem assessments and provides a robust foundation for future wildfire modeling and simulation.

Keywords: *Wildfires, Forest fuel, Forest Ecosystems, Forest Inventory, Sampling Design, GIS, Geospatial Analysis, UAVs*

INTRODUCTION

Wildfires pose an increasing threat to forest ecosystems worldwide, particularly due to climate change, land-use dynamics, and the growing frequency of extreme weather events. Effective wildfire risk management – from prevention and early detection to monitoring and restoration – depends critically on the availability of high-quality, harmonized geospatial datasets. This data must accurately reflect the complex environmental conditions that influence a fire's ignition, spread, and overall impact.

This study presents a methodological approach for advanced methods for acquisition, processing and verification of data, as an important stage of pre-processing and quality control of different datasets integrated into a common layer in a GIS environment for subsequent 3D modeling, analysis and evaluation for the purposes of the GeoFireData project.

The goal is to establish a statistically representative model that can be applied to the study, modeling, and monitoring of wildfires.

Early traditional models of wildfire spread (Rothermel 1972), based on empirical relationships, laid the foundations, but their accuracy depends heavily on the input data. With advances in GIS, it has become possible to integrate topographic data, vegetation information, and meteorological conditions into more complex simulation models (Chuvieco et al. 1989). Sentinel-2 has established itself as the preferred tool for mapping burned areas (Roteta et al. 2019; Liu et al. 2023). The combined use of Landsat-8 and Sentinel-2 data further improves spatiotemporal coverage and detection accuracy, particularly for small and fragmented fires (Roy et al. 2019). The spectral indices derived from these sensors – NDVI, dNBR, and RBR – are widely used to assess fire severity and monitor post-fire vegetation recovery (Howe et al. 2022). Meteorological reanalysis products (ERA5, MERRA-2) complement the geospatial framework. The integration of these diverse layers into coherent geospatial databases is an essential prerequisite for reliable wildfire modeling and simulation (Chuvieco et al. 2019; Sakellariou et al. 2022).

Despite advances in satellite-based methods, the detailed three-dimensional characterization of forest fuel remains a challenge. UAV-LiDAR systems demonstrate high potential for precise mapping of forest types. Hoffrén et al. (2024) reported an average point density of 452 points/m² across 73 forest sample plots in Spain, and the integration of HMLS/SLAM improved the identification of fuel in the lower structural layers. Aerial LiDAR scanning allows for a direct assessment of fuel bulk density (Martin-Ducup et al. 2025). FastFuels demonstrates how integrating LiDAR data can provide input for QUIC-Fire simulators (Marcozzi et al. 2025). Ground-based SLAM systems complement aerial measurements, providing detailed information on ground-level fuel (Holvoet et al. 2025). Multispectral UAV cameras expand analytical capabilities through NDVI and the assessment of fuel moisture (Santos et al. 2024).

This study is dedicated to a current scientific problem that is poorly studied in a national context. As an initial stage in the GeoFireData project implementation, it offers a modern approach to harmonizing and integrating diverse geospatial datasets and information. While stratified sampling and GIS-based forest inventory methods are well established in the literature, this study introduces several specific contributions.

First, a five-criteria stratification scheme is proposed – combining forest type, age class, elevation, terrain slope and canopy height – specifically designed to capture the multidimensional variability of mixed forest ecosystems relevant to wildfire behavior.

Second, a two-stage area-based generalization (≥10 ha → ≥20 ha) was implemented through an explicit hierarchical merging procedure that preserves ecologically meaningful stratum attributes – forest type and age class as priority criteria – while ensuring statistical stability of within-stratum estimates.

Third, four heterogeneous geospatial sources – high-resolution UAV-derived DEM and DSM, national forest management inventory data, Ministry of Agriculture and Food (MoAF) orthophotos and DSM, and specialized layers for agricultural areas from the Land Parcel Identification System (LPIS) – were integrated into a unified analytical layer through a reproducible QGIS-based workflow.

Fourth, an improved forest type classification algorithm was applied: instead of relying solely on the dominant tree species (DV1), the classification is based on the relative proportion of all tree species in the main canopy layer (DV1-DV5) with a threshold of ≥70%, reflecting the actual species composition of the stand, improving the delineation of mixed forest stands, and correcting their systematic underrepresentation.

Fifth, sample representativeness was verified using three complementary methods: a two-sample Kolmogorov-Smirnov test comparing the distributions of quantitative indicators (elevation, terrain slope, canopy height); a χ^2 test assessing the proportional representation of categorical indicators (forest type, age class); and a spatial coverage analysis (NearestHub) evaluating the uniformity of sample plot distribution across the study area.

The need for statistically representative samples and methodological frameworks for their extraction using GIS remains a current research issue. The GeoFireData project addresses this need by offering a systematic approach to acquiring, integrating and validating geospatial data to support advanced wildfire analysis.

METHODOLOGY: THE GEOFIREDATA APPROACH

The methodology developed as part of the GeoFireData project is structured around the acquisition, processing, and verification of geospatial data to derive a statistically representative model for wildfires. Figure 1 illustrates the complete workflow used to conduct the studies, perform statistical and GIS analyses, and present the results.

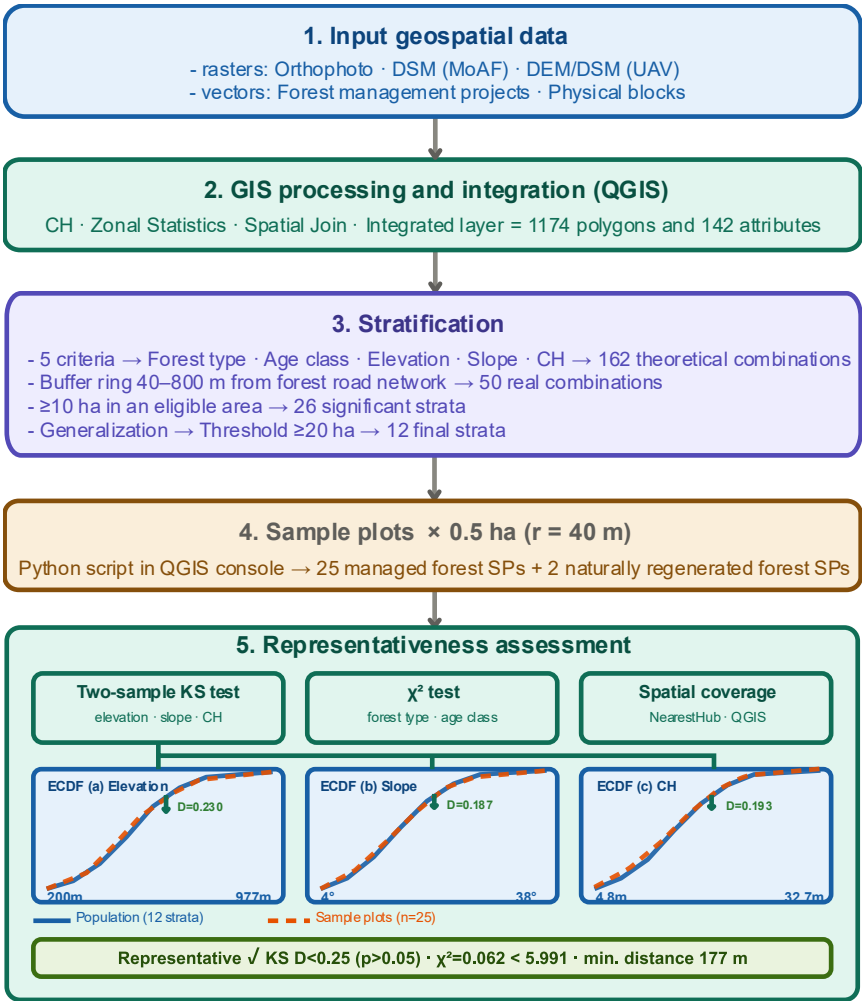

Figure 1. GeoFireData workflow – from input data (1) to representativeness assessment (5)

Study area

The study area is located administratively in the Montana Province, Varshets Municipality, within the territorial boundaries of the villages of Dolno Ozirovo and Gorno Ozirovo, and falls within the boundaries of the "Vratsa Balkan"

Nature Park (Figure 2). The study area covers 2,132.37 hectares, with a length of approximately 7 km and a width of approximately 3 km. The area is located in the Stresherski division of the Vratsa Mountain, at an elevation of 270 to 1,000 m, with an elevation difference of 800 m.

The geological foundation consists of sandstones, conglomerates, siltstones, and argillites, and in the eastern parts, of limestones and dolomitic limestones (Haydutov et al. 1995). The soils include gray forest soils (Gleyic-Eutric Planosols, Lithi-Haplic Luvisols), brown forest soils with rendzinas (Lithi-Eutric Cambisols, Rendzic Leptosols), and deluvial-meadow soils (Koinov et al. 1968). The climate is temperate continental. The vegetation is predominantly broadleaf – mixed stands of Quercus cerris, Quercus frainetto and Quercus pubescens (Bohn et al. 2000; Agrolesproekt 2019). According to data from the "Vratsa Balkan" Nature Park, one incident /fire/ was reported on February 16, 2024, in the vicinity of the village of Dolno Ozirovo.

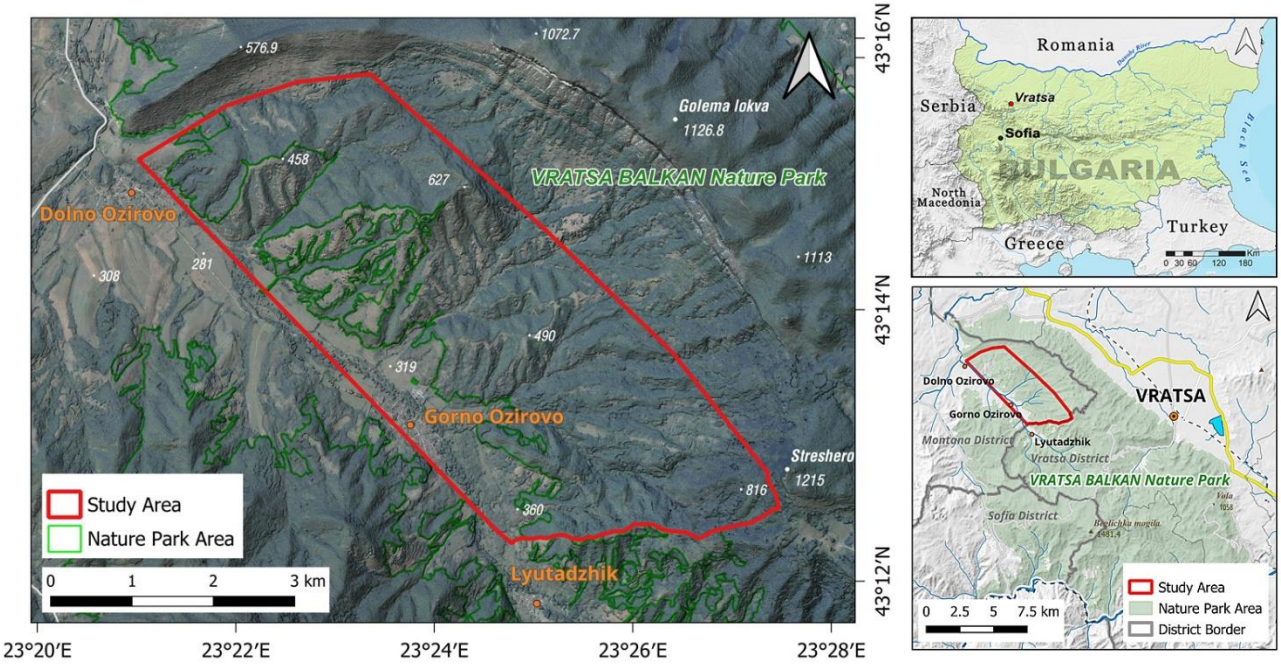

Figure 2. Map of the study area – "Vratsa Balkan" Nature Park, Varshets Municipality, Montana Province, Bulgaria; Area: 2,132.4 ha; elevation range: 270-1,000 m a.s.l.; centre: 43°13'38"N, 23°24'9"E

Available geospatial data

The following data is available for the study area: a digital orthophoto map from the MoAF (0.40 m/pix, RGB, 2020) and DSM (4 m/pix); UAV/ drone survey from 2022 with an orthophoto map (0.15 m/pix), DSM (0.5 m/pix) and DEM (0.5 m/pix) with the following limitation: 55% of the area was captured with trees still without leaves; ESRI shapefiles of the physical blocks (PB) from the LPIS of the MoAF; and a forest management shapefile organized by subdivisions of the Regional Forest Directorate Berkovitsa, including complete inventory attributes (up to five tree species per subdivision, along with species distribution, origin, age, height, and diameter).

Creation of an integrated geospatial layer

All geoinformation layers are projected into the BGS2005 / UTM Zone 34N coordinate system. The slope and aspect were calculated from a UAV DEM (0.5 m/pixel) using the algorithm described by Horn (1981), implemented in QGIS (QGIS Development Team 2024). The canopy height (CH) is calculated as CH = DSM MoAF (4 m) − DEM UAV (resampled to 4 m using average resampling). Zonal Statistics (Zonal Statistics in QGIS) extracts mean elevation, mean slope, and mean canopy height (CH) for each forest management subdivision. A spatial join (Spatial Join in QGIS) identifies forest polygons within the area's boundaries.

The classification of forest type is based on the relative proportion of coniferous and broadleaf species in the first floor (dominant) (attributes for the tree species included in the subsection – 5 tree species are considered): ≥70% coniferous → coniferous; ≥70% broadleaf → broadleaf; otherwise → mixed. This approach takes into account the actual species composition in the subdivision, rather than just the dominant species.

Stratification Scheme

Stratification divides the studied forest area into **homogeneous spatial units** – **strata**, based on five criteria: (1) forest type – broadleaf/coniferous/mixed (based on ≥70% coverage in the first floor (dominant)) (2) stand stage – young (≤40 years) / mature (41-80 years) / overmature (>80 years) (based on stand age); (3) elevation – low (200-350 m) / medium (350-500 m) / high (>500 m); (4) slope – gentle (<15°) / moderate (15-30°) / steep (>30°) (Lundbäck et al. 2021), with slope calculated from the DEM using the algorithm by Horn (1981); (5) Canopy height (CH) – low/medium (≤10 m) / high (>10 m). The theoretical maximum is 3×3×3×3×2 = 162 combinations. An additional criterion is the accessibility of forest areas; for this reason, the existing forest roads in the study area were included, and buffer zones (40-800 m) were generated based on the forest road network. A filter is also applied to ensure that participating subdivisions have a minimum area (≥ 5000 m²). Strata with an area below the threshold value within the eligible zone are merged with the nearest significant strata according to a hierarchical principle, while maintaining forest type and stage as priority criteria. Verification is performed using SQL queries in the QGIS DB Manager to assess the completeness and accuracy of the attributes according to the five stratification criteria for each polygon, as well as the total area by strata relative to the specified threshold values. Forest polygons with identical attributes across the five stratification criteria are grouped into strata (multipolygons) through geospatial vector processing in QGIS.

Stratified Random Sampling

Stratified random sampling is formed by sample plots (SPs) randomly located within each stratum, subject to a minimum distance from the stratum boundary and from adjacent sample plots. Each sample plot is circular in shape with an area of 0.5 ha (r ≈ 40 m). The eligible area is defined as the intersection of forest polygons with a buffer ring from the forest road network. The lower boundary of the buffer corresponds to the radius of the sample area (r = 40 m), which ensures that the circle fits completely within the polygon. The upper limit is determined by the operational scope of the field team. The centroids of the sample plots are randomly distributed within the permitted area, subject to a minimum distance from the polygon boundary and a minimum distance between adjacent sample plots, implemented via a Python script in the QGIS console. Naturally regenerated forest (NRF) areas are treated separately with a fixed number of sampling plots for pilot characterization.

The target number of sample plots **n** was determined based on the operational capacity of the field team for single-season multisensory surveying. The distribution among strata is carried out using proportional allocation by area (Cochran 1977):

$$n_h = \max(1, \text{round}(A_h / \Sigma A_h \times n))$$

where n_h is the specific number of sampling plots obtained for stratum h, A_h is the area of stratum h within the accessible forest area in the study area, and ΣA_h is the total area of all strata in the same area. The function max(1, ...) guarantees a minimum of 1 SP for each stratum regardless of its area, ensuring coverage of all final strata. Larger strata receive proportionally more SPs – the specific values of n_h for each stratum are presented in the Results section.

Assessment of the representativeness of the sample

The representativeness of the stratified random sampling is assessed using two complementary approaches.

For the quantitative indicators (elevation, slope, CH), the two-sample Kolmogorov-Smirnov test (KS test) is applied, a nonparametric method suitable when the distributions are not known in advance (Massey 1951; Cochran 1977). The test statistic **D** is the maximum vertical distance between the empirical cumulative distributions of the sample and the reference population:

$$D = \max |F_n(x) - F_N(x)|$$

where $\mathbf{F_n(x)}$ is the cumulative distribution in the n sample plots, and **F_N(x)** is the cumulative distribution of the reference population. The ECDF of the sample is compared against a stratified reference population, defined as the 12 final strata, each represented by its stratum mean value weighted by area in the accessible zone. By plotting the cumulative proportion of observations below each value of the indicator, a small maximum vertical distance **D** between the two curves indicates that the sample closely follows the distributional shape of the reference population. This approach is consistent with the stratified random sampling design, which aims to represent strata-level distributions rather than individual subdivision-level variability. If $p > 0.05$ ($\alpha = 0.05$), the null hypothesis of homogeneity is not rejected – the two distributions are considered statistically indistinguishable and the sample is deemed representative with respect to that indicator.

For categorical indicators (forest type, age class), the χ^2 test is applied:

$$\chi^2 = \Sigma\ (O_i - E_i)^2 / E_i$$

where O_i is the observed frequency of category ***i*** in the sample, and E_i is the expected frequency under proportional representation. Spatial coverage is assessed by the average distance from the centroid of each forest subdivision to the nearest SPs (Average Nearest Hub Distance in QGIS).

RESULTS

Structure of the Forest Fund

The integrated geospatial layer includes 1,174 polygons. The forests cover 770 subdivisions (1,234.6 ha, 57.9% of the study area), unclassified forests cover 335 polygons (312.2 ha, 14.6%), and non-forested areas – 69 polygons (21.8 ha). The remaining 563.8 ha include forest roads, clearings, and forest meadows.

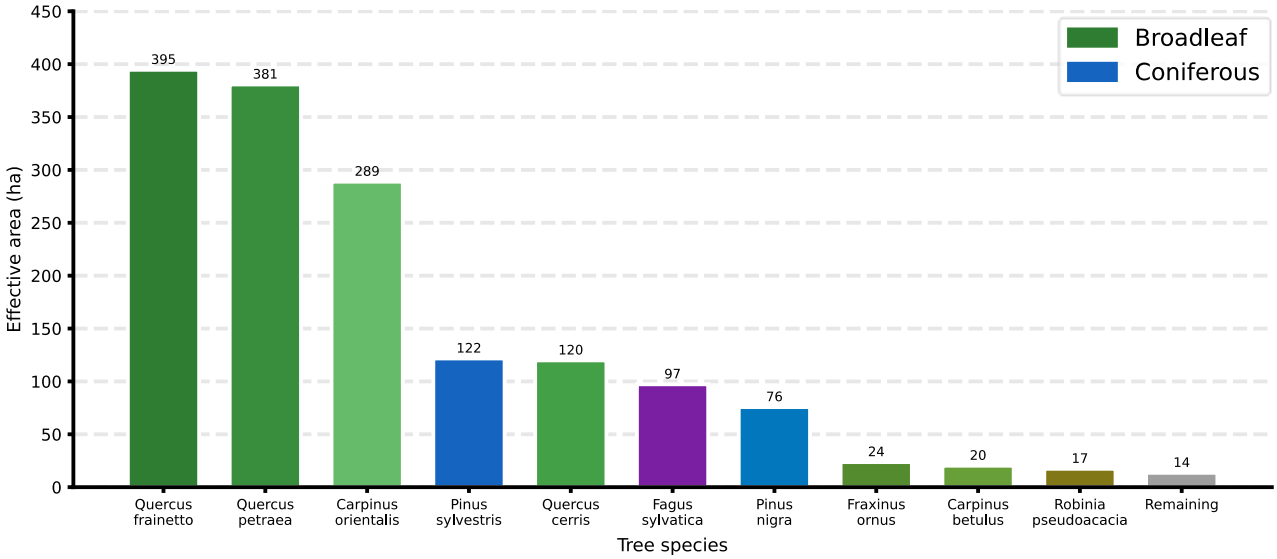

Figure 3. Tree species composition of the forest fund by effective area (from tree species 1 (DV1) to tree species 5 (DV5); forest management shapefile, Regional Forest Directorate Berkovitsa)

The tree species composition (Figure 3, Figure 4a, Table 1) is dominated by Quercus frainetto (394.7 ha, 25.4%) and Quercus petraea (380.8 ha, 24.5%), followed by Carpinus orientalis (288.9 ha, 18.6%). Broadleaf species occupy 82.1% of the area, conifers (Pinus sylvestris and Pinus nigra) occupy 13.5%, and mixed stands occupy 4.4%. In terms of stage (Figure 4b), mature stands predominate (57.2%), followed by young stands (39.6%) and overmature stands (3.1%). The classification was based on the relative proportion of species in the main canopy layer with a threshold of 70%.

Table 1. Tree species composition of the forest fund (effective area = area × proportion/10; DV1–DV5)

Tree species	Effective area (ha)	% of total	Average height H (m)	Average age (yrs)	Average diameter D (CMNE)	Forest type
Quercus frainetto	394.7	25.4	9.9	42	10.9	Broadleaf
Quercus petraea	380.8	24.5	10.9	44	12.3	Broadleaf
Carpinus orientalis	288.9	18.6	4.9	45	3.7	Broadleaf
Pinus sylvestris	121.7	7.8	13.9	38	16.8	Coniferous
Quercus cerris	119.9	7.7	11.4	47	16.5	Broadleaf
Fagus sylvatica	97.3	6.3	15.5	67	19.4	Broadleaf

Pinus nigra	75.7	4.9	13.4	38	18.2	Coniferous
Fraxinus ornus	23.7	1.5	8.1	39	8.6	Broadleaf
Carpinus betulus	20.3	1.3	12.6	51	12.9	Broadleaf
Robinia pseudoacacia	17.4	1.1	7.0	12	5.8	Broadleaf
Remaining	13.6	0.9	—	—	—	—
Total	**1554.0**	**100.0**	—	—	—	—

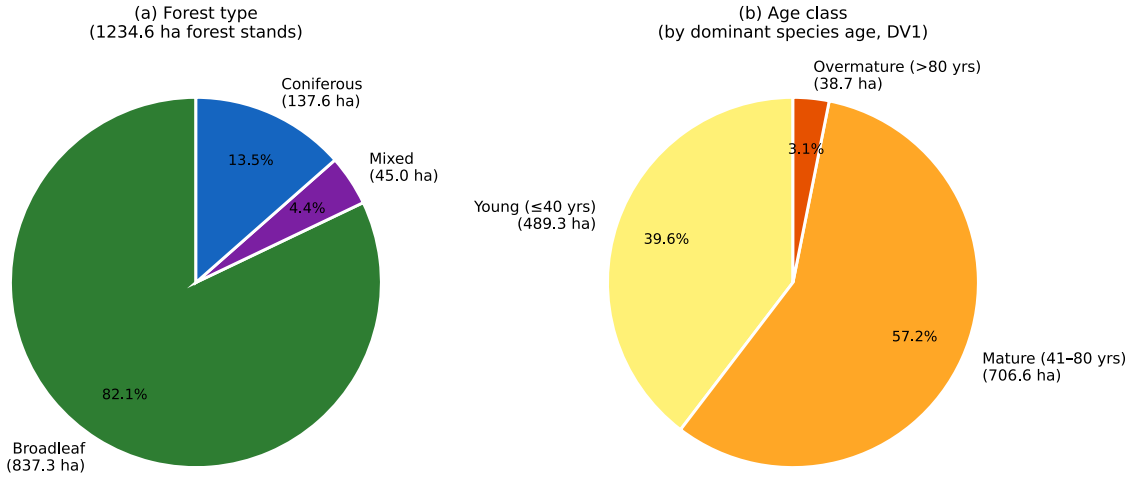

Figure 4. Structure of the forest fund by forest type (a) and stand stage (b)

Topographic and structural characteristics

Analysis of the DEM (0.5 m/pixel; Horn 1981) shows a weighted average elevation of 455 m (range 288-977 m). The dominant area is “Medium” (350-500 m), covering 827.4 ha (67.0%). The average slope is 19.6°, with moderate slopes (15-30°) occupying 81.4%, gentle slopes (<15°) – 15.1%, and steep slopes (>30°) – 3.4%. The height of the canopy ranges from 4.8 to 32.7 m (average 12.8 m); trees with high canopy (>10 m) account for 69.0% (Figure 5).

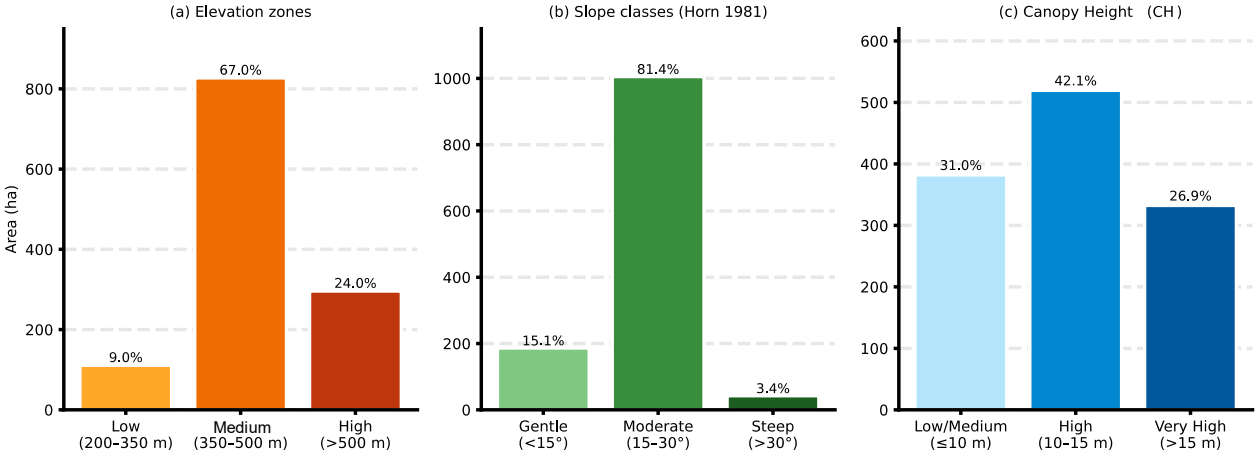

Figure 5. Topographic and structural characteristics: (a) elevation zones, (b) slope (Horn 1981), (c) CH (canopy height) classes

Stratification scheme and layout of the sample plots

From the theoretical maximum of 162 combinations, 50 combinations exist in the real data (after a buffer of 40-800 m and a filter area ≥ 5000 m²). Of these, 26 have an area ≥10 ha in the accessible forest area in the study area (1019.8 ha) and are preserved as a separate vector layer for comparative visualization. Subsequently, a generalization with a threshold of ≥20 ha was applied (Cochran 1977; Westfall et al. 2011), which reduced the scheme to 12 final strata. The merging follows a hierarchy of the following principle steps (PS): PS1 – preserving forest type, stage and CH, elevation is classified to the "Medium" class; PS2 – preserving forest type and stage, transferring elevation and slope to the dominant combination, with the CH class being preserved (High or Low) if there is a suitable target stratum; PS3 – the "Mixed Forest" class with a small area is merged with "Broadleaf" while preserving the stage. Figure 6 and Table 2 show the final distribution of the 12 generalized strata along with their corresponding sample plots.

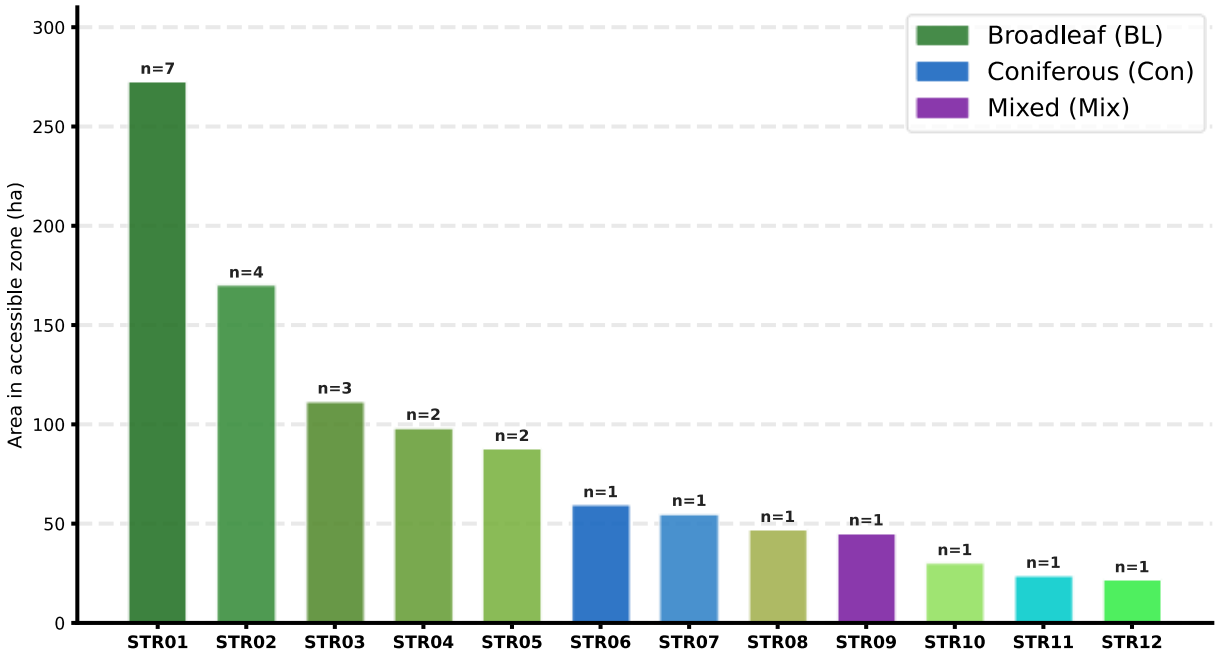

Figure 6. Final 12 strata and proportional allocation of sample plots (n - number of sample plots; buffer 40-800 m; threshold ≥20 ha, Cochran 1977)

Table 2. Final stratification scheme and distribution of sample plots (Cochran 1977; Westfall et al., 2011)

ID*	Stratum**	Area (ha)	% of total	n SP	ha/SP	Forest type
STR01	BL_Mat_Med_Mod_HCH	272.4	26.7	7	38.9	Broadleaf
STR02	BL_You_Med_Mod_HCH	169.8	16.6	4	42.5	Broadleaf
STR03	BL_You_Med_Mod_LCH	111.4	10.9	3	37.1	Broadleaf
STR04	BL_Mat_Med_Mod_LCH	98.1	9.6	2	49.1	Broadleaf
STR05	BL_Mat_High_Mod_HCH	87.5	8.6	2	43.8	Broadleaf
STR06	Con_Mat_Med_Mod_HCH	59.2	5.8	1	59.2	Coniferous
STR07	Con_You_Med_Mod_HCH	54.8	5.4	1	54.8	Coniferous
STR08	BL_Mat_High_Mod_LCH	46.7	4.6	1	46.7	Broadleaf
STR09	Mix_You_Med_Mod_HCH	45.0	4.4	1	45.0	Mixed

STR10	BL_Mat_Med_Gen_HCH	30.0	2.9	1	30.0	Broadleaf
STR11	Con_Mat_High_Mod_HCH	23.6	2.3	1	23.6	Coniferous
STR12	BL_You_Med_Gen_LCH	21.4	2.1	1	21.4	Broadleaf
NRF	Naturally regenerated forest***	312.2	—	2	—	—
	Total	**1019.8 + 312.2**	**100.0**	**27**	—	—

** ID: unique stratum identifier used in Figure 6; ** Strata abbreviations: BL = Broadleaf · Con = Coniferous · Mix = Mixed ; Mat = Mature (41-80 yrs) · You = Young (≤40 yrs) ; Med = Medium elevation (350-500 m) ; High = High elevation (>500 m) ; Mod = Moderate slope (15-30°) ; Gen = Gentle slope (<15°) ; HCH = High Canopy Height (>10 m) ; LCH = Low/Medium Canopy Height (≤10 m).*
**** Naturally regenerated forest (312.2 ha): formerly agricultural land, now naturally regenerated with coppice forest species; characterized using DEM, CH, and RGB orthophotos.*

A total of 27 sample plots were established: 25 in the managed forest strata and two in the naturally regenerated forest. The sampling intensity is 1.32% of the eligible area (1,019.8 ha) and 0.93% of the total surveyed forest area (1,458.3 ha), corresponding to 1 sampling plot per 40.8 ha and 1 sampling plot per 54.0 ha. All SPs are located within 40-800 meters of the forest road network, with a minimum distance of 40 meters from the boundary of the polygon and 120 meters between neighboring SPs (Figure 7).

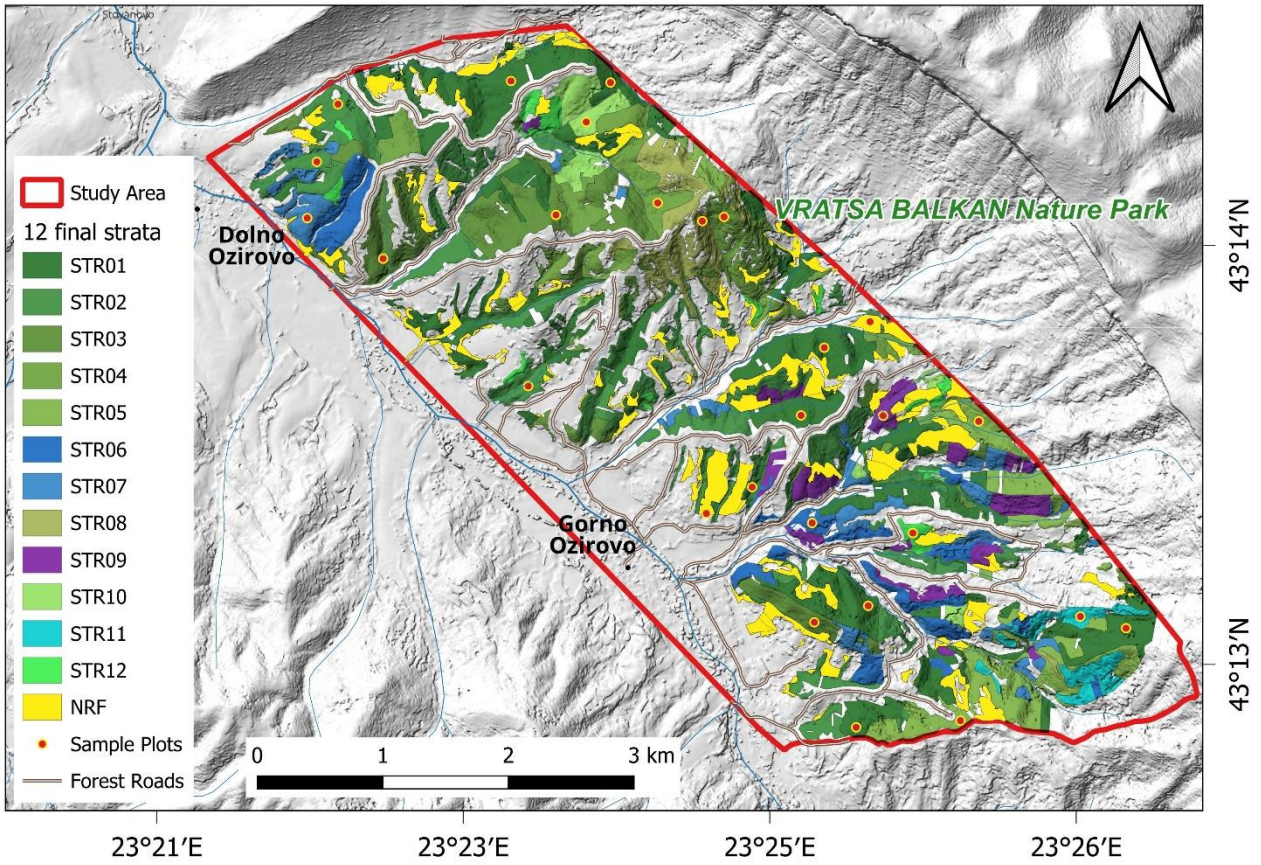

Figure 7. Map of the spatial distribution of the 12 final strata and 27 sample plots (SPs) within the accessible forest area of the study zone (1,019.8 ha; buffer 40-800 m from the forest road network). Strata IDs (STR01-STR12) correspond to Table 2. Yellow polygons – naturally regenerated forest areas (NRF, 312.2 ha). Red dots with yellow outline – sample plots (n = 27; 25 managed forest SPs + 2 NRF SPs).

Assessment of representativeness

The results of the representativeness assessment are summarized in Table 3. The two-sample KS test shows $D < 0.25$ for all three quantitative indicators at $p > 0.05$ (elevation: $D = 0.230$, $p = 0.684$; slope: $D = 0.187$, $p = 0.876$; CH: $D = 0.193$, $p = 0.847$), which means that the sample does not differ statistically from the stratified reference population on any of these indicators.

Figure 8 shows the ECDF curves for elevation, slope, and CH. In all three panels, the sample plot curve (orange dashed line) closely follows the stratified population curve (blue line) across the full range of values. The largest deviation for elevation occurs in the medium elevation range (350-500 m), reflecting the dominance of the Medium elevation class in both the stratified population (67.0%) and the sample. For slope, the curves are nearly identical across the moderate range (15-30°), which accounts for 81.4% of the accessible forest area. The CH curves show the closest agreement, confirming that the two CH classes (≤10 m and >10 m) are proportionally represented in the sample.

The χ^2 test for forest type ($\chi^2 = 0.062$, $\chi^2_crit = 5.991$, $df = 2$) and for age class ($\chi^2 = 0.041$) confirms the proportional representation of all categories. The minimum distance between SPs is 177 m (exceeding the threshold of 120 m), and the spatial coverage of 6.74 × 4.98 km encompasses the entire test area.

Table 3. Results of the assessment of sample representativeness (α = 0.05)

Indicator / test	Method	Value	Threshold (α=0.05)	Conclusion
KS test – elevation	Massey (1951)	D = 0.230, p = 0.684	p > 0.05	Representative ✓
KS test – slope	Massey (1951)	D = 0.187, p = 0.876	p > 0.05	Representative ✓
KS test – CH	Massey (1951)	D = 0.193, p = 0.847	p > 0.05	Representative ✓
χ^2 test – forest type	Cochran (1977)	$\chi^2 = 0.062$	$\chi^2_crit = 5.991$	Representative ✓
χ^2 test – age class	Cochran (1977)	$\chi^2 = 0.041$	$\chi^2_crit = 5.991$	Representative ✓
Min. distance SPs	Spatial	177 m	≥ 120 m	Representative ✓
Spatial coverage	Nearest Hub	6.74 × 4.98 km	~7 × 3 km	Covers the area ✓

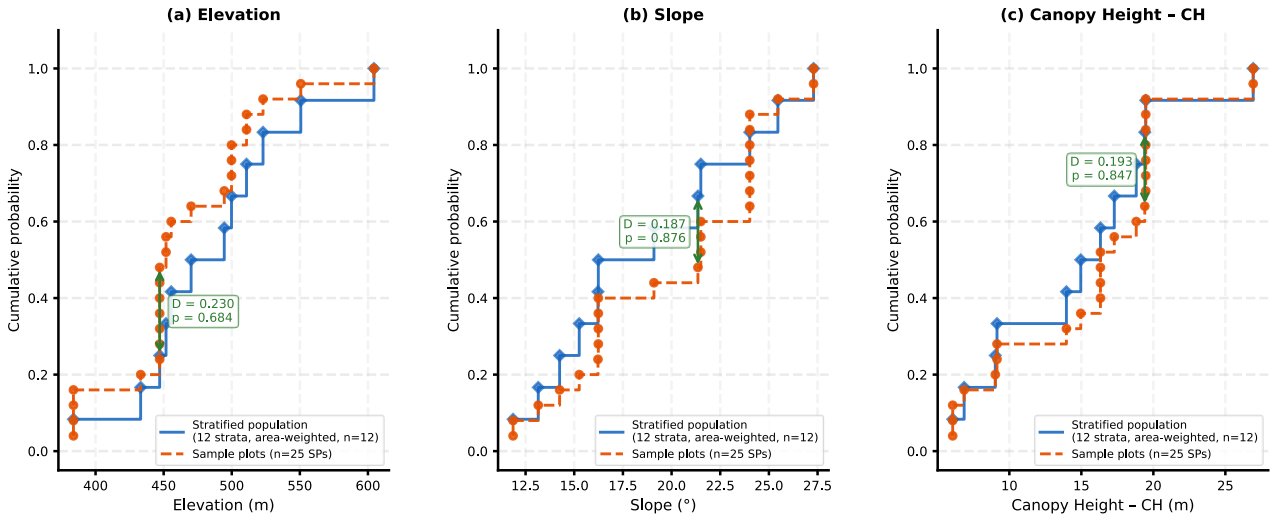

Figure 8. Empirical Cumulative Distribution Functions (ECDF) for quantitative stratification indicators: (a) elevation, (b) slope, (c) canopy height (CH). Two-sample Kolmogorov-Smirnov test: stratified reference population (12 strata, area-weighted, n = 12) vs. sample plots (SPs, n = 25). The green arrow indicates the maximum deviation D between the

two ECDF curves. Elevation: D = 0.230, p = 0.684; Slope: D = 0.187, p = 0.876; CH: D = 0.193, p = 0.847. All p-values > 0.05 confirm that the null hypothesis of homogeneity is not rejected and the sample is representative across all three indicators.

DISCUSSION

The developed GeoFireData methodology represents an integrated GIS approach for deriving a statistically representative sample of sample plots in heterogeneous forest ecosystems. Stratification based on five criteria (forest type, stage, elevation, slope, CH) provides multidimensional coverage of the key factors influencing the development of a representative model of forest ecosystems, with an emphasis on the factors that carry the greatest weight in the modeling and monitoring of wildfires.

The two-stage generalization – from 50 actual combinations to 26 significant strata (≥10 ha), and then to 12 final strata (≥20 ha) – is methodologically justified. The threshold of ≥10 ha ensures that SPs of 0.5 ha are physically located with the necessary buffers. The threshold of ≥20 ha ensures statistical stability of the estimates – for smaller strata with n = 1 SPs, an intra-stratum estimate of the variance is not possible (Cochran 1977; Westfall et al. 2011). Preserving the 26-strata scheme as a separate layer allows for the comparative visualization and documentation of the area's spatial heterogeneity.

The correction of the forest type algorithm – from checking only DV1 to calculating the relative participation (threshold ≥70%) of coniferous and broadleaf species in the first floor (dominant) – led to a more accurate classification, in which coniferous and mixed strata received correct representation. Applying a buffer of 40 m (= r of the SPs) from the polygon boundary is a physically justified necessary boundary – it guarantees complete inclusion of the circle r = 40 m within the stratum boundaries. The upper limit of 800 m provides accessibility for a field team with UAV-LiDAR, SLAM scanner and forest inventory equipment.

The intensity of 1.32% relative to the allowable area is within accepted practice for detailed UAV-LiDAR surveys (Cochran 1977; Dalla Corte et al. 2022). The assessment of representativeness using the KS test, χ^2 test and spatial analysis confirms that the 27 SPs representatively cover the variability of the region across all stratification indicators.

An important methodological limitation is the exclusion of crown density from the stratification criteria. Accurate measurement requires UAV-LiDAR data with high point density (>100 points/m²), which will be generated in the next phase. The naturally regenerated forest areas (312.2 ha) are characterized solely by remote sensing indicators – sufficient for the pilot phase, but requiring verification of the forest structure in future stages.

CONCLUSION

This study presents the methodology and results of the first (preparatory) phase of the GeoFireData project, as applied to the "Vratsa Balkan" study area. An integrated geospatial layer comprising of 1,174 polygons has been developed, combining forest management attributes with topographic indicators derived from DEM and CH through spatial analyses in a GIS environment. Forest type is classified based on the relative proportion of species in the main canopy layer (threshold ≥70%).

The stratification scheme was developed in two steps: 50 actual combinations → 26 significant strata (≥10 ha, 1,019.8 ha) → 12 final strata (≥20 ha) based on five criteria. The proportional allocation designates 27 sample plots (25 managed forests + 2 naturally regenerated forests, 0.5 ha each) with an intensity of 1.32% relative to the eligible area. Representativeness was verified using the two-sample KS test ($D < 0.25$ for elevation, slope and CH), the χ^2 test ($\chi^2 < 0.07$ for forest type and stage), and spatial analysis of cover.

In the next stages of GeoFireData, the selected SPs will be characterized by UAV-LiDAR (>100 points/m²), SLAM scanning (ground fuel), multispectral UAV imaging (NDVI, NDWI, LFMC), and a complete inventory of the sample plots. The integration of the data will allow verification of the forest fire simulation using the Anderson (1982) and Scott-Burgan (Scott and Burgan 2005) classifications.

ACKNOWLEDGMENTS

This study has been supported by the National Science Fund under the research project "Mapping and analysis of geospatial data on forest ecosystems for wildfire modeling and monitoring (GeoFireData)", № КП-06-Н94/4.

REFERENCES

Agrolesproekt, Eood. 2019. *Forest Management Project of the State Forestry Enterprise "Berkovitsa" [Горскостопански план на ТП ДГС „Берковица"].* Agrolesproekt EOOD (Sofia).

Anderson, Hal E. 1982. *Aids to Determining Fuel Models for Estimating Fire Behavior.* U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station (Ogden, UT). https://www.fs.usda.gov/rm/pubs_int/int_gtr122.pdf.

Bohn, Udo, and Rudolf Neuhäusl. 2000. Karte der natürlichen Vegetation Europas / Map of the Natural Vegetation of Europe. Münster: Landwirtschaftsverlag.

Chuvieco, Emilio, and Russell G. Congalton. 1989. "Application of remote sensing and geographic information systems to forest fire hazard mapping." *Remote Sensing of Environment* 29 (2): 147-159. https://doi.org/https://doi.org/10.1016/0034-4257(89)90023-0. https://www.sciencedirect.com/science/article/pii/0034425789900230.

Chuvieco, Emilio, Florent Mouillot, Guido R. van der Werf, Jesús San Miguel, Mihai Tanase, Nikos Koutsias, Mariano García, Marta Yebra, Marc Padilla, Ioannis Gitas, Angelika Heil, Todd J. Hawbaker, and Louis Giglio. 2019. "Historical background and current developments for mapping burned area from satellite Earth observation." *Remote Sensing of Environment* 225: 45-64. https://doi.org/https://doi.org/10.1016/j.rse.2019.02.013. https://www.sciencedirect.com/science/article/pii/S0034425719300689.

Cochran, William G. 1977. *Sampling Techniques*. 3rd ed.*Wiley series in probability and mathematical statistics*. New York: John Wiley & Sons.

Dalla Corte, Ana Paula, Ernandes M. da Cunha Neto, Franciel Eduardo Rex, Deivison Souza, Alexandre Behling, Midhun Mohan, Mateus Niroh Inoue Sanquetta, Carlos Alberto Silva, Carine Klauberg, Carlos Roberto Sanquetta, Hudson F. P. Veras, Danilo Roberti Alves de Almeida, Gabriel Prata, Angelica M. A. Zambrano, Jonathan W. Trautenmüller, Anibal de Moraes, Mateus A. Karasinski, and Eben North Broadbent. 2022. "High-Density UAV-LiDAR in an Integrated Crop-Livestock-Forest System: Sampling Forest Inventory or Forest Inventory Based on Individual Tree Detection (ITD)." *Drones* 6 (2): 48. https://doi.org/10.3390/drones6020048.

European Commission, Joint Research Centre. 2025. EFFIS Burnt Areas [Dataset]. In: *European Forest Fire Information System (EFFIS)*: Copernicus Emergency Management Services (CEMS).

Haydutov, I., S. Yanev, D. Tronkov, T. Nikolov, I. Sapunov, P. Tchoumatchenco, Tz Tzankov, R. Dimitrova, and N. Popov. 1995. *Explanatory note for the geological map of Bulgaria in scale 1:100 000, Berkovitsa map sheet*. Sofia: Committee on Geology, Company for Geophysical Surveys and Geological Mapping.

Hoffrén, Raúl, María T. Lamelas, and Juan de la Riva. 2024. Classification and Mapping of Fuels in Mediterranean Forest Landscapes Using a UAV-LiDAR System and Integration Possibilities with Handheld Mobile Laser Scanner Systems. *Remote Sensing* 16 (18): 3536. https://doi.org/10.3390/rs16183536.

Holvoet, Justin, Markus P. Eichhorn, Francesca Giannetti, Daniel Kükenbrink, Xinlian Liang, Martin Mokroš, Jan Novotný, Timo P. Pitkänen, Stefano Puliti, Mitja Skudnik, Krzysztof Stereńczak, Louise Terryn, Cédric Vega, and Chiara Torresan. 2025. "Terrestrial and mobile laser scanning for national forest inventories: From theory to implementation." *Remote Sensing of Environment* 329: 114947.
https://doi.org/https://doi.org/10.1016/j.rse.2025.114947. https://www.sciencedirect.com/science/article/pii/S0034425725003517.

Horn, Berthold K. P. 1981. "Hill shading and the reflectance map." *Proceedings of the IEEE* 69 (1): 14-47. https://doi.org/10.1109/PROC.1981.11918.

Howe, Alexander A., Sean A. Parks, Brian J. Harvey, Saba J. Saberi, James A. Lutz, and Larissa L. Yocom. 2022. Comparing Sentinel-2 and Landsat 8 for Burn Severity Mapping in Western North America. *Remote Sensing* 14 (20): 5249. https://doi.org/10.3390/rs14205249.

Koinov, V., H. Trashliev, M. Yolevski, T. Andonov, N. Ninov, A. Hadzhiyanakiev, E. Angelov, T. Boyadzhiev, E. Fotakieva, S. Krastanov, and Y. Staykov. 1968. Soil map of Bulgaria at a scale of 1:400,000. Sofia: Institute of Soil Science Nikola Poushkarov, Bulgarian Academy of Sciences Press, General Office of Geography and Mapping (GUGK).

Liu, Peng, Yongxue Liu, Xiaoxiao Guo, Wanjing Zhao, Huansha Wu, and Wenxuan Xu. 2023. "Burned area detection and mapping using time series Sentinel-2 multispectral images." *Remote Sensing of Environment* 296: 113753. https://doi.org/https://doi.org/10.1016/j.rse.2023.113753. https://www.sciencedirect.com/science/article/pii/S0034425723003048.

Lundbäck, Mikael, Henrik Persson, Carola Häggström, and Tomas Nordfjell. 2021. "Global analysis of the slope of forest land." *Forestry: An International Journal of Forest Research* 94 (1): 54-69. https://doi.org/10.1093/forestry/cpaa021.

Marcozzi, Anthony, Lucas Wells, Russell Parsons, Eric Mueller, Rodman Linn, and J. Kevin Hiers. 2025. "FastFuels: Advancing wildland fire modeling with high-resolution 3D fuel data and data assimilation." *Environmental Modelling & Software* 183: 106214.

https://doi.org/https://doi.org/10.1016/j.envsoft.2024.106214.
https://www.sciencedirect.com/science/article/pii/S1364815224002755.

Martin-Ducup, Olivier, Jean-Luc Dupuy, Maxime Soma, Juan Guerra-Hernandez, Eva Marino, Paulo M. Fernandes, Ariadna Just, Jordi Corbera, Marion Toutchkov, Charlie Sorribas, Jerome Bock, Alexandre Piboule, Francesco Pirotti, and François Pimont. 2025. "Unlocking the potential of Airborne LiDAR for direct assessment of fuel bulk density and load distributions for wildfire hazard mapping." *Agricultural and Forest Meteorology* 362: 110341. https://doi.org/https://doi.org/10.1016/j.agrformet.2024.110341. https://www.sciencedirect.com/science/article/pii/S0168192324004544.

Massey, Frank J. 1951. "The Kolmogorov-Smirnov Test for Goodness of Fit." *Journal of the American Statistical Association* 46 (253): 68-78. https://doi.org/10.2307/2280095. http://www.jstor.org/stable/2280095.

Open Source Geospatial Foundation Project. 2024. QGIS Geographic Information System Version 3.40.11-Münster (LTR). Accessed 20.02.2026. https://qgis.org.

Roteta, Ekhi, Aitor Bastarrika, Marc Padilla, and Emilio Chuvieco. 2019. "Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa." *Remote Sensing of Environment,222:- 17.*

Rothermel, R.C. 1972. *A Mathematical Model for Predicting Fire Spread in Wildland Fuels*. Intermountain Forest & Range Experiment Station, Forest Service, U.S. Department of Agriculture.

Roy, David P., Haiyan Huang, Luigi Boschetti, Louis Giglio, Lin Yan, Hankui H. Zhang, and Zhongbin Li. 2019. "Landsat-8 and Sentinel-2 burned area mapping - A combined sensor multi-temporal change detection approach." *Remote Sensing of Environment* 231: 111254. https://doi.org/https://doi.org/10.1016/j.rse.2019.111254. https://www.sciencedirect.com/science/article/pii/S0034425719302731.

Pix4D SA. 2024. PIX4Dmapper 4.5.6. Lausanne, Switzerland. Accessed 19.02.2026. https://www.pix4d.com/product/pix4dmapper-photogrammetry-software.

Sakellariou, Stavros, Athanassios Sfougaris, Olga Christopoulou, and Stergios Tampekis. 2022. "Integrated wildfire risk assessment of natural and anthropogenic ecosystems based on simulation modeling and remotely sensed data fusion." *International Journal of Disaster Risk Reduction* 78: 103129. https://doi.org/https://doi.org/10.1016/j.ijdrr.2022.103129. https://www.sciencedirect.com/science/article/pii/S221242092200348X.

Santos, Filippe L. M., Gonçalo Rodrigues, Miguel Potes, Flavio T. Couto, Maria J. Costa, Susana Dias, Maria J. Monteiro, Nuno D. Ribeiro, and Rui Salgado. 2024. Moisture Content Vegetation Seasonal Variability Based on a Multiscale Remote Sensing Approach. *Remote Sensing* 16 (23): 4434. https://doi.org/10.3390/rs16234434.

Scott, Joe H., and Robert E. Burgan. 2005. *Standard Fire Behavior Fuel Models: A Comprehensive Set for Use with Rothermel's Surface Fire Spread Model.* U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station (Fort Collins, CO). https://www.fs.usda.gov/rm/pubs/rmrs_gtr153.pdf.

Westfall, James A., Paul L. Patterson, and John W. Coulston. 2011. "Post-stratified estimation: within-strata and total sample size recommendations." *Canadian Journal of Forest Research* 41 (5): 1130-1139. https://doi.org/10.1139/x11-031. https://cdnsciencepub.com/doi/abs/10.1139/x11-031.

BIOGRAPHY

Davis Dinkov is an associate professor of GIS and Cartography at the Bulgarian Academy of Sciences, National Institute of Geophysics, Geodesy and Geography. His research and professional work focus on geospatial solutions and technologies, professional cartography, GIS, and web-based information systems. He is the author and team leader for the development and publication of over 30 cartographic and geoinformation products, as well as numerous scientific articles and studies. He is a member of the Bulgarian Cartographic Society. Team leader and participant in national and international scientific and educational projects.

Atanas Kitev is an associate professor of GIS and Cartography in the Department of Geography at the National Institute of Geophysics, Geodesy and Geography of the Bulgarian Academy of Sciences, Applied Geoinformatics Section. His main research interests are in Geographic Information Systems, Cartography, Remote Sensing, Landscape structure analysis, and Ecology. He has participated in several research projects, both international and national with emphasis on GIS and Remote Sensing applications.

Desislava Hristova works as a geographer at the National Geoinformation Center (NGIC) at the National Institute of Geophysics, Geodesy and Geography, Bulgarian Academy of Sciences (NIGGG-BAS). She is also a PhD Candidate at the NIGGG-BAS, majoring in “Cartography and GIS”. Desislava completed her master's degree in 2018 at Sofia University “St. Kliment Ohridski” (at the Faculty of Geology and Geography), majoring in “Physical Geography and Landscape Ecology”. Her scientific interests are focused in the field of: GIS, Spatial Analysis, Cartography, Remote Sensing, Ecosystem Ecology, Ecosystem Services, Ecosystem Accounting.

Lyubka Pashova is an associate professor of Geodesy at the Bulgarian Academy of Sciences, National Institute of Geophysics, Geodesy and Geography. Her research work and professional and teaching experience in geodetic science include topics related to geodata processing using modern mathematical methods, geodetic coordinate reference systems, regional and local geodynamic studies, monitoring, analysis and assessment of natural hazards, incl. disaster risk reduction. She is a member of the Bulgarian Cartographic Association. Leads and participates in several projects at the national and international levels. Since October 2014, she has been the GEO Principal Alternate for Bulgaria in the Group on Earth Observations (GEO).