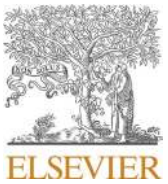
Remote Sensing Applications: Society and Environment 39 (2025) 101713



Contents lists available at ScienceDirect

Remote Sensing Applications: Society and Environment

journal homepage: www.elsevier.com/locate/rsase

Geospatial patterns of carbon storage in relation to protection status and road infrastructure in an insular forest landscape

Gabriel E. Suárez-Fernández [a,*], Savvas Zotos [b], Joaquín Martínez-Sánchez [a], Marilena Stamatiou [b], Elli Tzirkalli [b], Ioannis N. Vogiatzakis [b,c], Pedro Arias [a]

[a] CINTECX, Universidade de Vigo. Applied Geotechnologies Group, Vigo, 36310, Spain
[b] Faculty of Pure and Applied Sciences, Open University of Cyprus, Latsia, Nicosia, 2220, Cyprus
[c] Department of Soil, Plant and Food Science, University of Bari Aldo Moro, Italy

ARTICLE INFO

Keywords:
Carbon dynamics
Edge effect
Machine learning
Natura 2000
Remote sensing

ABSTRACT

Forests play a crucial role in climate change mitigation through carbon storage. Nevertheless, these ecosystems face increasing threats from human activities, such as infrastructure development and Land Use/Land Cover (LULC) changes. To date, limited research has focused on understanding how roads impact carbon stocks in forests, and how this relation is influenced by protection regimes, especially on islands. This study on the island of Cyprus aims to assess Machine Learning (ML) techniques for estimating key forest variables such as Canopy Cover (CC) and to analyze the spatial dynamics of carbon stocks around roads in relation to LULCs and protection regimes. Remote Sensing (RS) data, including Landsat imagery and orthophotos, are combined with ML to create an ensemble model for detailed LULC classifications. The Integrated Valuation of Ecosystem Services and Trade-offs (InVEST) tool is utilized to estimate carbon stocks for each LULC and statistical analysis is used to evaluate interactions between forests, roads, and protection regimes. The analysis revealed that protected sites store significantly 17 % more carbon than unprotected areas whilst proximity to roads exhibits complex effects on carbon stocks, with varying patterns depending on the protection status. The ensemble model outperforms individual models, achieving 92 % accuracy and a kappa of 0.91, showing the advantages of combining algorithms for more robust predictions. The research highlights the impact of integrating ML with ecosystem service models to improve understanding of interactions between roads, LULC, and forests. It also emphasizes the importance of conservation and roadside vegetation management for ecosystem resilience and sustainable carbon storage.

1. Introduction

Since the publication of the Sustainable Development Goals and the 2030 Agenda (United Nations General Assembly, 2015), alongside European strategies such as the Fit for 55 package (European Council. Council of the European Union, 2024), forests have gained prominence in the bioeconomy, with carbon sequestration and storage emerging as key regulating service (Dang et al., 2017;

* Corresponding author.
E-mail addresses: gabrieleduardo.suarez@uvigo.gal (G.E. Suárez-Fernández), savvas.zotos@ouc.ac.cy (S. Zotos), joaquin.martinez@uvigo.gal (J. Martínez-Sánchez), marilena.stamatiou@st.ouc.ac.cy (M. Stamatiou), elli.tzirkalli@ouc.ac.cy (E. Tzirkalli), ioannis.vogiatzakis@ouc.ac.cy (I.N. Vogiatzakis), parias@uvigo.gal (P. Arias).

https://doi.org/10.1016/j.rsase.2025.101713
Received 29 June 2025; Received in revised form 18 August 2025; Accepted 2 September 2025
Available online 3 September 2025

Zhao et al., 2022). Nonetheless, these ecosystems face growing pressure from natural and anthropogenic factors (Hu et al., 2017; Nunery and Keeton, 2010), leading to a reduction in their integrity and the benefits they provide (Ouyang et al., 2016). Changes in Land Use and Land Cover (LULC) particularly affect forests' carbon storage capacity by altering their structure and composition, notably along the interfaces between different LULC (Malik et al., 2024).

Among these dynamics, a key driver is the *Edge Effect*, which results from interactions at ecosystem boundaries, especially between forests and road networks (Buss et al., 2024). These interactions alter ecological processes, influencing the rate of forest carbon and biomass stocks (Harper et al., 2005; Kalinaki et al., 2023). Now more than ever, given their impact on vegetation dynamics (Müllerová et al., 2011), it is critical to understand and model the influence of road networks on carbon stocks. Roads affect biotic and abiotic components (Sun et al., 2025; Zhou et al., 2020), increasing light penetration, which can enhance productivity and biodiversity (Dormann et al., 2020; Pöpperl and Seidl, 2021; Vepakomma et al., 2018). Nevertheless, they also modify temperature, humidity, and soil conditions, increasing tree mortality at edges (Delgado et al., 2007; Toivio et al., 2017). These alterations influence forest dynamics and carbon stocks, but significant knowledge gaps remain regarding their effects at varying distances from roads and under different conservation conditions.

To counteract biodiversity loss, protection regimes are established to address the unfavorable conservation status (Cannizzo et al., 2024; Duncanson et al., 2023; Miranda et al., 2016). The Natura 2000 network (Evans, 2012) was established in Europe primarily to safeguard habitats and species (Kallimanis et al., 2015; Mammides et al., 2024). However, despite these efforts, over 80 % of European habitats remain in poor conservation status, and only 14 % of protected forest habitats are in favorable condition (Naumann et al., 2020). Understanding the interaction between these protected areas with carbon stocks and roads remains understudied (Graham et al., 2021). This gap is even wider in insular regions, where specific forest inventories may rely on scarce and fragmented data sources (Zenonos et al., 2025), and where limited research has analyzed the impact of road networks on carbon stocks (Cruz-Pérez et al., 2023), with minimal consideration given to protection status. Islands, especially in the Mediterranean, face additional pressures due to their ecological vulnerability and exposure to extreme climate events (Vogiatzakis et al., 2023). Therefore, distinguishing between protected and unprotected areas is essential to assess how road infrastructure differentially affects carbon stocks in these fragile ecosystems, where land-use regulation, conservation policies, and resilience to disturbance can vary substantially depending on protection status.

On the other hand, Remote Sensing (RS) has been widely used to monitor land use, forest characteristics, and vegetation dynamics (Kanjin and Alam, 2024; Teodoro and Duarte, 2022; Wang et al., 2022). Satellite imagery facilitates LULC classification (Joy et al., 2024; Kovárník and Janová, 2025) and carbon stock estimation (Chinembiri et al., 2023; Suárez-Fernández et al., 2025). While cost-effective, freely available multispectral satellite data often have medium spatial resolution—ranging from 10 to 100 m according to Gómez et al. (2016)—which can limit their ability to capture key forest inventory variables such as Canopy Cover (CC) (Bera et al., 2023). High-resolution aerial orthophotos emerge as a suitable complement to satellite data, enhancing land cover and forest structure analysis. Their integration improves LULC monitoring and ecosystem function assessments (Subedi et al., 2024).

In addition to RS, the advances of Machine Learning (ML) have also strengthened forest ecosystem analysis (Braham et al., 2023). ML techniques can outperform traditional statistical methods by capturing complex non-linear relationships between sensor reflectance and ecosystem dynamics (Tamiminia et al., 2024; Zurqani, 2025). However, single ML models may be limited by overfitting, limited generalization, or performance variability across study areas and datasets (Lei et al., 2020), while ensemble approaches, integrating multiple algorithms, enhance predictive reliability (Ayushi et al., 2024; Du et al., 2023). These models mitigate individual weaknesses, yielding more robust LULC and forest inventory predictions (Zhang et al., 2022). These techniques can be further enhanced by tools designed to assess habitat quality and ecosystem services (Rimal et al., 2019), especially when extensive and detailed ground-based inventory data are scarce, limiting direct connections between carbon stocks and ML models based on RS data (Zenonos et al., 2025). Recent studies have attempted to overcome this limitation by integrating spatial modeling tools and hybrid approaches (Almeida et al., 2025; Hernández-Guzmán et al., 2019; Khachoo et al., 2024), while also helping to bridge data gaps through spatially explicit, model-driven estimates (García-Ontiyuelo et al., 2024; Li et al., 2024). By integrating these frameworks, which enhance the reliability of LULC classifications and reinforce their link to carbon stock estimates, a deeper understanding of the complex interactions between anthropogenic infrastructure—such as roads—and natural systems is facilitated.

In this context, this study aims to: (1) evaluate individual and ensemble ML models for deriving key forest inventory variables such as CC, and (2) analyze carbon stock in relation to LULCs and protection regimes, focusing on Mediterranean island forests adjacent to roads. A key novelty is the analysis of carbon stocks across protected and unprotected habitats considering land cover and proximity to the road network.

2. Study area and materials

2.1. Study area

The study was conducted in the Troodos massif of Cyprus, the third largest island of the Mediterranean (Fig. 1). Troodos covers one third of the island's surface ranging from 300 m up to the highest peak of Olympus at 1952 m. The study area incorporates several protected sites designated under the Natura 2000 network (Fig. 1) such as Ethniko Dasiko Parko Troodous (CY5000004), Dasos Machaira (CY2000004), and Dasos Pafou (CY2000016) (European Environment Agency, 2024), which consider both protection regimes, that is, habitats and birds (Evans, 2012). The dominant tree species in the region include *Pinus brutia, Juniperus* spp., and, to a lesser extent, *Quercus alnifolia* and *Pinus nigra* subsp. *pallasiana,* the latter being found at the high altitudes (>1400m) of Troodos (Prodromou et al., 2024; Tsintides et al., 2002). These forests are embedded within a heterogeneous LULC mosaic—including

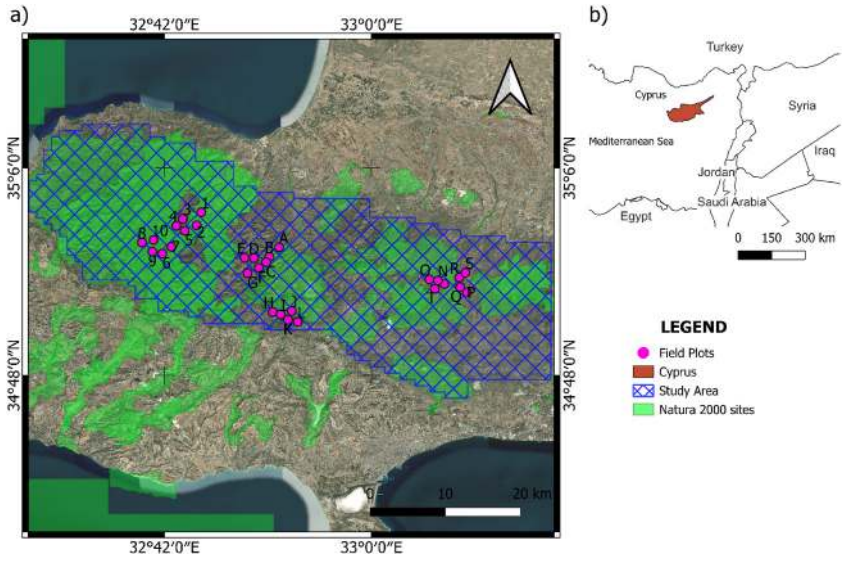


Fig. 1. Location of the study area (Visualized in EPSG:4326 – WGS84) for global context; spatial analyses were conducted in EPSG:32636). a) Distribution of fieldwork plots, Natura 2000 sites and region analyzed. b) Location of Cyprus in the eastern parts of the Mediterranean Sea.

scrublands, rocky outcrops, and agricultural land—which, when combined with the massif's topographic complexity, contributes to marked spatial variation in carbon storage.

As in many European islands, ecological and management contexts in Cyprus differ from mainland areas: insularity entails greater vulnerability to climate change, limited resources, and governance challenges often shaped by political and economic dependencies (Vogiatzakis et al., 2023). Within this setting, the study area shows contrasting management and disturbance histories—for instance, forested interiors such as the Paphos Forest have experienced limited anthropogenic disturbance, while other zones reflect long-term human pressure (Republic of Cyprus - The Mines Service, 2025). At a broader scale, Cyprus has one of the densest road networks in Europe (2.3 km/km^2 on average), although the few remaining roadless areas—crucial for biodiversity conservation—are mostly found within Natura 2000 sites (Zomeni and Vogiatzakis, 2014). Specifically, the area analyzed is traversed by a mix of primary and secondary paved roads (approximately 3900 km), but is predominantly shaped by an extensive network of unpaved roads, which account for approximately 11,500 km—around 75 % of the total road length.

2.2. Data sources and analytical tools

A diverse set of geospatial datasets was integrated to support the analysis, all of which fully covered the extent of the study area delineated in Fig. 1. These included high-resolution orthophotos, satellite imagery, topographic information, infrastructure layers, and protected area boundaries. All datasets were selected based on official availability, completeness, and consistency across the study area. The 2014 orthophotos represent the most recent high-resolution, full-coverage dataset publicly released by national authorities. Similarly, the 30 m Digital Elevation Model (DEM) from 2000 remains the most widely used elevation dataset with complete coverage of the island. While newer sources exist globally, more recent DEMs are not currently available for Cyprus with national coverage and open access. Accordingly, Table 1 summarizes the main input layers used in this study, including details on spatial resolution, spectral content, acquisition years, and data sources.

For data processing and statistical analysis, RStudio 2022.12.0 (Posit Software, 2025) with R version 4.1.3 (R Core Team, 2024) was used, while QGIS 3.19.9 (QGIS, 2025) facilitated visualization and generation of geospatial layers. Additionally, the Integrated Valuation of Ecosystem Services and Trade-offs (InVEST) tool (Natural Capital Project, 2024), widely recognized for ecosystem service modeling—particularly for carbon stock assessment (García-Ontiyuelo et al., 2024; Li et al., 2024)—was employed to evaluate spatial patterns of carbon storage. This approach was motivated by the absence of a comprehensive and systematically structured national forest inventory in Cyprus, which is based on infrequent field campaigns, limited sampling, and basic statistical methods (Zenonos

Table 1
Summary of input geospatial datasets and their main characteristics.

Dataset[a]	Year (s)	Resolution	Spectral Content	Temporal Coverage	Source
Orthophotos[b]	2014	0.5 m	Blue, Green, Red	Single acquisition	Public Administration and Personnel Department (2024)
Digital Elevation Model (DEM)	2000	~30 m	Elevation	Static	NASA Shuttle Radar Topography Mission (SRTM) (2013)
Road Network	2024	Vector layer	Road type	Static	United Nations Office for the Coordination of Humanitarian Affairs (OCHA) (2024)
Natura 2000 sites	2022	Vector layer	Protected area boundaries	Static	European Environment Agency (2024)
Landsat-8 imagery[c] Landsat-9 imagery[d]	2014 2024	30 m	Blue, Green, Red, NIR, SWIR1, SWIR2	Median composites per season: • Spring: Mar 15–May 31 • Summer: Jun 15–Sep 14 • Autumn: Sep 15–Nov 30	Google Earth Engine (Gorelick et al., 2017)

[a] All datasets were reprojected to EPSG:32636 – WGS 84/UTM zone 36N to ensure spatial consistency in the analysis.
[b] Orthophotos are typically acquired during summer, although late spring acquisitions may also occur. These dates generally coincide with the Landsat summer composite period (June 15 – September 14), ensuring seasonal consistency for vegetation-related analyses. Note that the 2014 dataset represents the most recent full-coverage, official orthophoto mosaic available for Cyprus.
[c] Landsat-8 and Landsat-9 images with a cloud coverage of less than 5 %.
[d] Acronyms: DEM = Digital Elevation Model; NIR = Near-Infrared; SWIR1/2 = Short-Wave Infrared bands 1 and 2.

et al., 2025), often assuming constant variables (Ministry of Agriculture Rural Development and Environment of Cyprus, 2019). These constraints preclude the direct use of RS and ML approaches to estimate carbon stocks from forest structure, while InVEST has shown strong potential for estimating carbon storage using LULC data and minimal ancillary inputs, making it a suitable alternative in data-scarce contexts (de Araujo Fonseca & da Cunha Bustamante, 2025; García-Ontiyuelo et al., 2024).

3. Methodology

The methodology is illustrated in Fig. 2 and consists of three main stages. First, satellite images and orthophotos were acquired and preprocessed alongside vegetation indices (see Table 2 for indices derived from orthophotos and Table 3 for those from Landsat imagery). Then, multiple ML models were trained and optimized for LULC classification (see the end of Section 3.1 for details on the specific models used). Based on pixel-level classification, the CC was calculated, refining the models. The three best-performing models were selected to build an ensemble, enabling LULC determination at varying CC levels. Finally, matching and statistical analyses evaluated the impact of both Natura 2000 sites and road networks on carbon stocks, integrating field data, classification results, and carbon models.

3.1. Land Use/Land Cover classification models with orthophotos

ML models—described in detail later—identified five land cover classes at the pixel level: forest (“Forest”), ground (“Ground”), dwellings (“House/Urban”), roadway(“Road”), and water bodies (“Water”). This initial LULC classification was based on 0.5 m high-resolution visible-spectrum orthophotos from 2014—the most recent official dataset available—which were complemented with Landsat-8 infrared bands from the summer 2014 composite to ensure seasonal and same-year consistency. Due to the disparity in spatial resolution (30 m for Landsat-8 and 0.5 m for orthophotos), the Lanczos interpolation method (Turkowski, 1990) was employed to rescale Landsat-8 infrared band images to a target resolution of 0.5 m, as it better preserves edge definition and reduces aliasing compared to others, while maintaining spectral information (Sales et al., 2023). In addition, vegetation spectral indices were calculated from the orthophoto visible bands using scaled reflectance values (Table 2) to enhance vegetation detection. These indices were specifically selected for their applicability to visible-spectrum imagery, consistent with the orthophoto data; thus, indices requiring near-infrared bands, such as NDVI, were not prioritized in this stage.

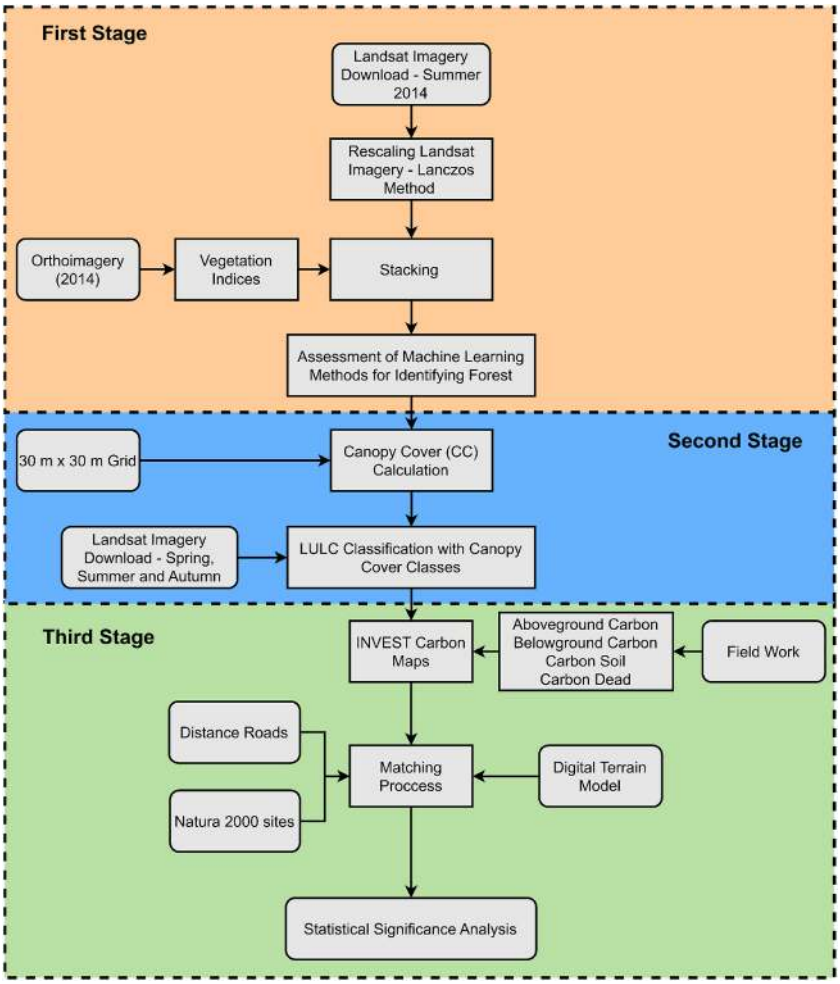


Fig. 2. Workflow of the methodology. The first step, involving the acquisition and preprocessing of RS data, is highlighted in orange. The second step, which focuses on training and testing classification models, is shown in blue. Finally, the third stage, encompassing ecological services modeling and statistical analyses, is represented in green.

Table 2
Vegetation indices based on the visible light spectrum (Red, Green, Blue) for classification using aerial orthoimages[a].

Index	Equation	Source
Green Leaf Index (GLI)	$\frac{(2*Green) - Red - Blue}{(2*Green) + Red + Blue}$	Louhaichi et al. (2001)
Normalized Difference Green-Red Index (NGRDI)	$\frac{Green - Red}{Green + Red}$	Gitelson et al. (2002)
Visible Atmospherically Resistant Index (VARI)	$\frac{Green - Red}{Green + Red - Blue}$	Gitelson et al. (2003)
Green-Red Ratio Index (GRRI)	$\frac{Green}{Red}$	Gamon and Surfus (1999)
Modified Green-Red Vegetation Index (MGRVI)	$\frac{(Green^2) - (Red^2)}{(Green^2) + (Red^2)}$	Bendig et al. (2015)
Vegetative Index (VEG)	$\frac{Green}{(Red^{0.667}) + (Blue^{(1-0.667)})}$	Hague et al. (2006)

[a] This table presents vegetation indices calculated from scaled reflectance values of orthophotos using only bands in the visible spectrum (Red, Green, Blue). These indices were selected for their suitability with very high-resolution imagery lacking near-infrared or shortwave infrared information.

Subsequently, feature selection was performed using the Boruta method (Kursa and Rudnicki, 2010) and Variance Inflation Factor (VIF) (Galvão and Araújo, 2009) to retain relevant predictors, eliminating irrelevant and redundant variables. For the VIF, values greater than 10, which indicate high collinearity and correlation with other variables, were selected for elimination (Ayushi et al., 2024), resulting in a final set of 10 predictors. Then, the training and testing of the models for the classification of LULCs at a 0.5 m pixel resolution, and the subsequent identification of the presence of vegetation, were performed using a total of 1750 manually selected points on the orthophotos, distributed equally (350 points per class) in the five categories mentioned above. These points were randomly distributed and allocated into a 70/30 ratio for training and testing of the models, respectively, ensuring that the distribution was representative for both the area and the explanatory variables.

Consequently, several ML models, including Support Vector Machine (SVM) (Cervantes et al., 2020), Random Forest (RF) (Sun et al., 2024), k-Nearest Neighbor (KNN) (Shi et al., 2022), Artificial Neural Network (ANN) (Park and Lek, 2016), Multivariate Adaptive Regression Splines (MARS) (López-Serrano et al., 2016), Gradient Boosting (GB) (Yang et al., 2020) and Penalized Regression (PR) (Adhikari et al., 2023) were applied. The hyperparameter optimization was crucial to enhance accuracy and robustness, preventing overfitting or underfitting (Bhungeni et al., 2024). Thus, for each model, a random search strategy was used to explore different combinations of hyperparameters. In total, just over 3500 runs were executed across all models to identify the optimal configurations, all implemented in R (via RStudio) to ensure reproducibility and efficient processing, with computations performed on an HP OMEN laptop equipped with an 11th Gen Intel Core i7-11800H processor and 32 GB of RAM. The best-performing setup was selected based on cross-validated accuracy and consistency between predictions and observations. Finally, the model with the highest accuracy and agreement between predictions and observations was selected. The assessment was carried out using the confusion matrix and the Kappa coefficient (Equation S1 in Supplementary Material), which measures the agreement between predictions and observations, adjusting performance relative to random expectations. A higher Kappa indicates better model agreement and provides a more robust evaluation.

3.2. Ensemble model with Landsat imagery

Landsat-8 and Landsat-9 images from multiple periods were used to update LULC classification from 2014 to 2024, refining forest categories based on CC, defined as the ground area covered by the vertical projection of tree crowns (Jennings, 1999). The analysis defined as ground area a 30 m × 30 m grid matching Landsat pixels, whilst canopy area was derived from orthophoto-based LULC classifications (0.5 m × 0.5 m). To ensure purely forested areas, the 30 m grids were selected using the 2012 CORINE (Coordination of Information on the Environment) LULC classification (European Environment Agency and Joint research centre, n.d.)—a standardized European LULC dataset—as its year closely aligns with the 2014 orthophoto-derived LULC data. The classification from orthophotos was then binarized, assigning 1 to "forest" and 0 to "soil" pixels, and the proportion of "1″ pixels was computed for each 30 m grid identified as forest using CORINE, representing the CC due to the high resolution of orthophotos.

Forest grids were defined, in accordance with the official criteria, as areas with a CC of at least 10 % (Ministry of Agriculture Rural Development and Environment of Cyprus, 2019). These grids were then categorized into three classes based on CC: Hollow (<50 %), Incomplete (50 %–80 %), and Complete (>80 %), reflecting thresholds commonly applied in LULC classification standards (Di Gregorio; Antonio & Jansen, 2005) and RS studies (Tang et al., 2019). Although slight variations exist in other established thresholds (FAO, 2001), these categories are widely used in the literature, fall within a comparable range, and reflect ecologically meaningful distinctions. Additionally, other non-forest landscape classes, including those representing crops ("Crop"), dwellings ("House/Urban"), water bodies (Water), and barren areas ("Bare"), were also classified, enabling structured analysis by coverage type. In the following step, the vegetation indices shown by Bera et al. (2023) and detailed in Table 3 were calculated using the seasonal average satellite images—Summer, Autumn, and Spring.

In order to address typical multicollinearity in RS data, a Principal Component Analysis (PCA) was applied to spectral bands and vegetation indices, reducing redundancy while preserving at least 99 % of the original data variability (Jolliffe, 2002). At this stage, the aforementioned ML algorithms were applied to this new vector space, with their hyperparameters tuned using the randomized search approach. A total of randomly-selected 350 points were used for each of the classes (Bare, Complete, Crop, Hollow, House, Incomplete, Water), preserving the 70/30 ratio for model training and validation. After this, the three models that presented the best evaluation metrics, based on their accuracy and generalization, were selected.

Finally, the probabilities from these models were integrated into a new ensemble model, whose hyperparameters were re-optimized using the randomized search approach, resulting in just over 3500 model runs. The model with the best testing metrics was then applied to all satellite images from 2014 to 2024, obtaining an updated LULC classification, including CC classes. The performance of both the final and intermediate models was evaluated using the Kappa coefficient.

Table 3
Vegetation indices based on the multispectral spectrum (including Red, Green, Blue, Near-Infrared, and Shortwave Infrared) used with Landsat satellite imagery[a].

Index	Equation[b]	Source
Normalized Burn Ratio (NBR)	$\frac{NIR - SWIR2}{NIR + SWIR2}$	Key and Benson (2006)
Normalized Multi-band Drought Index (NMDI)	$\frac{NIR - (SWIR1 - SWIR2)}{NIR + (SWIR1 + SWIR2)}$	Wang and Qu (2007)
Normalized Canopy Index 1 (NCI1)	$\frac{SWIR1 - Green}{SWIR1 + Green}$	Vescovo and Gianelle (2008)
Normalized Canopy Index 2 (NCI2)	$\frac{SWIR2 - Green}{SWIR2 + Green}$	
Short-Wave Infrared Ratio 21 (SWIR21)	$\frac{SWIR2}{SWIR1}$	Guerschman et al. (2009)
Normalized Green (NG)	$\frac{Green}{(NIR + Red + Green)}$	Sripada et al. (2006)
Optimized Soil-Adjusted Vegetation Index (OSAVI)	$\frac{(SWIR2 - Red)}{(NIR + Red + 0.16)}$	Rondeaux et al. (1996)
Green Difference Vegetation Index (GDVI)	$NIR - Green$	Tucker et al. (1979)
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR - Red}{NIR + Red}$	Rouse et al. (1973)
Enhanced Vegetation Index (EVI)	$2.5 * \frac{NIR - Red}{NIR + 6\,Red - 7.5\,Blue + 1}$	Huete et al. (2002)
Visible Atmospherically Resistant Index Green (VARIg)	$\frac{Green - Red}{Green + Red - Blue}$	Gitelson et al. (2002)

[a] This table includes vegetation indices commonly applied to satellite multispectral data. Calculations incorporate bands across the full spectral range, including near-infrared and shortwave infrared, allowing for enhanced detection of vegetation and biophysical properties.
[b] Acronyms: NIR = Near-Infrared; SWIR1/2 = Short-Wave Infrared bands 1 and 2.

3.3. Analysis of the relationships between Natura 2000 areas, roads and carbon stocks

To estimate carbon stored in the study area, which included Above-Ground Carbon (AGC), Below-Ground Carbon (BGC), dead carbon, and soil carbon, 10 representative circular plots (12m radius) were selected for each CC class—Complete, Incomplete, and Hollow— totaling 30 plots (see pink dots in Fig. 1 and Table S1 in Supplementary Material for coordinates). As previously noted, the CC thresholds—10 %, 50 %, and 80 %—are commonly used in LULC classifications, RS studies, and technical criteria for forest inventory reporting (Di Gregorio; Antonio & Jansen, 2005; Ministry of Agriculture Rural Development and Environment of Cyprus, 2019; Tang et al., 2019). The tree species recorded were *Pinus brutia*, *Pinus nigra*, *Cedrus brevifolia*, *Quercus alnifolia*, and *Quercus coccifera*. *Pinus brutia* was the most abundant species, consistent with previous species distribution maps for the region (Prodromou et al., 2024). To minimize potential biases in mean carbon estimates, species-specific allometric equations were applied whenever available, and all individuals—both dominant and suppressed—were measured across mono- and mixed-species plots to ensure a representative assessment of stand-level biomass. Diameter at breast height (dbh) and total height (h) were measured to quantify Above-Ground Biomass (AGB), while the Below-Ground Biomass (BGB) was calculated based on the AGB value, using the equation defined by the Intergovernmental Panel on Climate Change (IPCC) (Penman et al., 2003) (see Table S2 in Supplementary Material for equations). Dead biomass in island regions was estimated at around 8 % of AGB, according to forest inventories of several Mediterranean countries (Augustynczik et al., 2024; Gasparini et al., 2022). These values were converted to carbon using a 0.5 conversion factor (Petersson et al., 2012), obtaining AGC, BGC and dead carbon. Soil carbon was derived from Camera et al. (2017) using soil bulk density from Panagos et al. (2024).

The InVEST tool was used to map the carbon storage across the study area using LULC classification layers from 2024 Landsat-9 images and field data, generating a carbon distribution map as in García-Ontiyuelo et al. (2024) and Li et al. (2024). As commented in Section 2.2, this tool is suitable for generating consistent carbon estimates from limited inputs, especially when extensive forest inventory data are lacking, which restricts the use of reflectance-based ML models for estimating carbon stocks. In order to analyze bias-free carbon distribution patterns relative to Natura 2000 sites and road networks, a matching analysis was conducted to compare points with similar characteristics (Ho et al., 2011; Mammides et al., 2024). A total of 17500 points were randomly selected in forested areas, each located within a 4 x 4 Landsat-9-pixel grid (120 m × 120 m). This grid size was chosen to approximate the 100 m spatial resolution of the CORINE dataset, while remaining compatible with Landsat's 30 m pixel structure. Slightly enlarging the unit also helps reduce geolocation errors and spectral noise, allowing each sample to more reliably capture local forest conditions rather than isolated pixel anomalies. Subsequently, for each point, the following average data was extracted: slope, roughness, altitude, and distances to paved roads, unpaved roads, all roads (paved and unpaved), and whether the pixel fell within a Natura 2000 site. Due to high correlation between slope and roughness (Fig. S1 in Supplementary Material), roughness was excluded. Similarly, as distance to unpaved roads was highly correlated with distance to all roads (since unpaved roads represented 75 % of the road network), it was excluded, as the distance to all roads more effectively assesses the impact of the road network.

Following the best practices described by Schleicher et al. (2020), several matching algorithms—specifically full, genetic, nearest neighbor, optimal, and subclass matching—were evaluated, and the one providing the best balance between treated and control points

was selected. Standard metrics, including standardized mean differences, variance ratios, and distributional balance (Ho et al., 2011), were used. As a result, the optimal algorithm was "nearest neighbor" with Mahalanobis distance (Supplementary Material Tables S3 and S4 and Figs. S2–S6), commonly applied in protected areas research (Mammides et al., 2024).

After matching, carbon stocks inside and outside Natura 2000 sites were analyzed using Generalized Additive Models (GAMs) with a Gamma distribution and a logarithmic link function—an approach well-suited for modeling continuous, positive, and right-skewed response variables (Ng and Cribbie, 2019), such as total carbon (Fig. S7 of the Supplementary Material). A schematic summary of the GAM model structure, including smooth terms and interaction effects, is provided in Specification S1 of the Supplementary Material. GAMs provide flexibility to capture non-linear relationships between predictors and the response variable, which are common in ecological data (Clark and Wells, 2023; Guisan et al., 2002). In this context, the modeling approach was used to describe patterns in the data, rather than to make predictions, focusing on the functional form and statistical significance of covariate effects, especially distance to roads and Natura 2000 status. Consequently, model adequacy was assessed using standard diagnostics, including residual plots, evaluation of variance structure, identification of influential observations, and assessment of smooth term complexity based on effective degrees of freedom and k-index values.

The GAM incorporated Natura 2000 status and its interaction with distance to roads, allowing for a comprehensive evaluation of potential effects. Furthermore, a Rosenbaum's sensitivity analysis (Rosenbaum, 2002) was performed to check robustness against unobserved biases, and the Wilcoxon test (Rey and Neuhäuser, 2011) was used to assess statistical significance of carbon differences between areas inside and outside Natura 2000. Finally, total carbon was calculated for each distance band from roads, and g-computation (Snowden et al., 2011) was utilized to estimate Natura 2000's impact based on road proximity, addressing residual imbalances post-matching.

4. Results

4.1. LULC classification with orthophotos and Landsat ensemble models

The model adjustment results and metrics for LULC classification using orthophotos and Landsat ensemble models are presented. For predictor selection in LULC classification with orthophotos, the Boruta method confirmed the significance of all predictors, while VIF analysis identified high multicollinearity (VIF >10) in two vegetation indices (NGRDI and MGRVI), leading to their exclusion (Table S5 in Supplementary Material).

Regarding the ML models using orthophotos, Table 4 summarizes the accuracy and optimal hyperparameters for each model. The SVM was the best-performing model, achieving a Kappa coefficient of 0.88 in the test set, and 0.93 during training. Thus, SVM was selected for its higher performance.

To calculate CC, the SVM classification model was applied to the entire study area, as exemplified in Fig. 3. The classification was then binarized, assigning 1 to the forest class (Section 3.2) and, using the 30 m grid belonging to forested areas, as shown in Fig. 3c, the CC was determined and classified into Complete, Incomplete, and Hollow.

In relation to PCA, a cumulative variance of 99.23 % was reached by PC17. Note that PC1 and PC2 together accounted for 75.61 % and the scatterplot of samples along these two components—shown in Fig. S8 in Supplementary Material—illustrates partial overlap among vegetation classes, indicating that additional PCs were necessary for better class discrimination. Consequently, after selecting 17 PCs, ML models were trained and optimized. As can be seen in Table 5, the best-performing models according to test performance metrics were ANN, SVM, and GB. Therefore, their predicted probabilities were used as inputs for the new ensemble model.

Finally, to obtain the final ensemble model for LULC classification, the hyperparameters of the selected base models—ANN, SVM,

Table 4
Results for ML models assessed with orthoimagery, including training and testing metrics, and optimal hyperparameters.

	Hyperparameters	Training		Testing	
Model[a]	Best Adjustment	Accuracy	Kappa	Accuracy	Kappa
SVM	cost: 2^11 gamma: 0.0625	0.9426	0.9283	0.9042	0.8802
RF	ntree: 1000 mtry: 7	0.9213	0.9017	0.8716	0.8395
KNN	k: 13	0.9107	0.8884	0.8620	0.8275
ANN	NeuronsperLayer: 24 threshold: 0.15	0.9033	0.8791	0.9022	0.8778
MARS	degree: 2 nprune: 15	0.8493	0.8116	0.8429	0.8036
GB	n.trees: 400 interaction.depth: 1 shrinkage: 0.06 n.minobsinnode: 16	0.9131	0.8914	0.8735	0.8419
PR	alpha: 0.857 lambda: 0	0.8591	0.8239	0.8582	0.8227

[a] Acronyms: SVM = Support Vector Machine; RF = Random Forest; KNN = k-Nearest Neighbors; ANN = Artificial Neural Network; MARS = Multivariate Adaptive Regression Splines; GB = Gradient Boosting; PR = Penalized Regression.

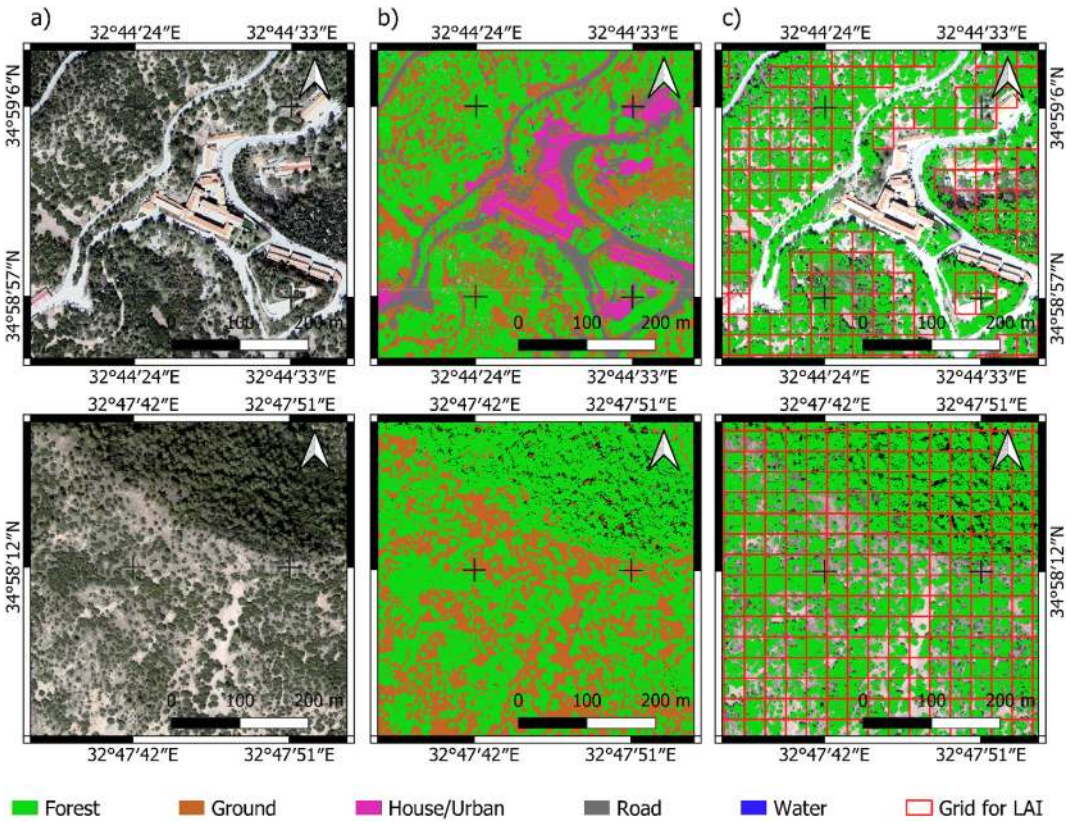


Fig. 3. Example of LULC classification from orthoimagery and binarization process for CC estimation (Visualized in EPSG:4326 – WGS84) for global context; classification and analyses performed in EPSG:32636). a) Orthoimagery. b) LULC classification. c) Forest class with 30 m × 30 m Landsat grid for CC estimation.

Table 5
Results for the ML models assessed with Landsat imagery, including training and testing metrics, and optimal hyperparameters.

	Hyperparameters	Training		Testing	
Model[a]	Best Adjustment	Accuracy	Kappa	Accuracy	Kappa
SVM	cost: 2^3 gamma: 0.0625	0.9428	0.9332	0.8840	0.8647
RF	ntree: 800 mtry: 16	0.9166	0.9027	0.8485	0.8233
KNN	k: 7	0.9583	0.9513	0.8472	0.8217
ANN	NeuronsperLayer: 58 threshold: 0.605	0.9526	0.9447	0.8840	0.8647
MARS	degree: 1 nprune: 25	0.8300	0.8017	0.8240	0.7946
GB	n.trees: 100 interaction.depth: 3 shrinkage: 0.0421 n.minobsinnode: 11	0.9272	0.9151	0.8594	0.8360
PR	alpha: 0.98 lambda: 0	0.8758	0.8551	0.8485	0.8233

[a] Acronyms: SVM = Support Vector Machine; RF = Random Forest; KNN = k-Nearest Neighbors; ANN = Artificial Neural Network; MARS = Multivariate Adaptive Regression Splines; GB = Gradient Boosting; PR = Penalized Regression.

and GB—and commonly used complementary models—KNN and PR— as in Ayushi et al. (2024), were optimized. As shown in Table 6, GB achieved the best test performance, with a kappa coefficient of 0.91 and an accuracy of 92 % (see Table S6 in the Supplementary Material for class-specific performance metrics), and with a suitable difference between training and testing results, indicating an acceptable generalization capacity. Consequently, the GB-based model was applied to classify the seven LULC categories in the study area for both 2014 and 2024 (shown in Fig. S9 Supplementary Material).

Table 6
Results for the ensemble models for final LULC classification, including training and testing metrics, and optimal hyperparameters.

	Hyperparameters	Training		Testing	
Model[a]	Best Adjustment	Accuracy	Kappa	Accuracy	Kappa
SVM	cost: 2^11 gamma: 0.001	0.9345	0.9236	0.9079	0.8926
GB	n.trees: 100 interaction.depth: 4 shrinkage: 0.0230 n.minobsinnode: 6	0.9604	0.9538	0.9202	0.9069
KNN	k: 8	0.9331	0.9220	0.9079	0.8926
ANN	NeuronsperLayer: 28 threshold: 0.136	0.9167	0.9029	0.9141	0.8997
PR	alpha: 0.7171 lambda: 0.0037	0.9031	0.8869	0.9018	0.8854

[a] Acronyms: SVM = Support Vector Machine; GB = Gradient Boosting; KNN = k-Nearest Neighbors; ANN = Artificial Neural Network; PR = Penalized Regression.

Table 7
Summary of carbon stocks used for estimation by InVEST modeling for each LULC category withing the study area.

LULC name	C above (t/ha)	C below (t/ha)	C soil (t/ha)	C dead (t/ha)	Total (tC/ha)	Area (ha)	tC Total
Bare	0	0	50	0	50	15081.03	754051.5
Complete	125	28	53	10	216	29577.42	6388722.7
Crop	0	0	43	0	43	16437.42	706809.1
Hollow	18	5	49	1	73	30319.92	2213354.2
City	0	0	49	0	49	2772.09	135832.4
Incomplete	54	13	55	4	126	63115.02	7952492.5
Water	0	0	35	0	35	3073.77	107582.0
Total	197	46	334	15	592	160376.67	18258844.3

4.2. Invest estimation and carbon relationships

Regarding the carbon stock estimates, each LULC category from the Landsat-9 classification, which provided the mapped surface areas, was linked to its carbon capacity using Table 7. This table was constructed following the input requirements of the InVEST model, combining carbon pool estimates derived from fieldwork sampling plots (see Table S1 in Supplementary Material) with additional data sources (see section 3.3). For the entire Troodos Mountain range, the total carbon stock estimated using InVEST was 18258844 t of carbon, or approximately 66949095 t of CO_2 equivalent, based on the conversion factor provided by the United States Environmental Protection Agency (EPA, 2024). This corresponds to an economic value of 4371 million Euros, calculated using the 2024 average carbon price in the European Trading System of 65.29 €/tCO_2 (SENDECO2, n.d.). Additionally, Fig. 4 specifically shows

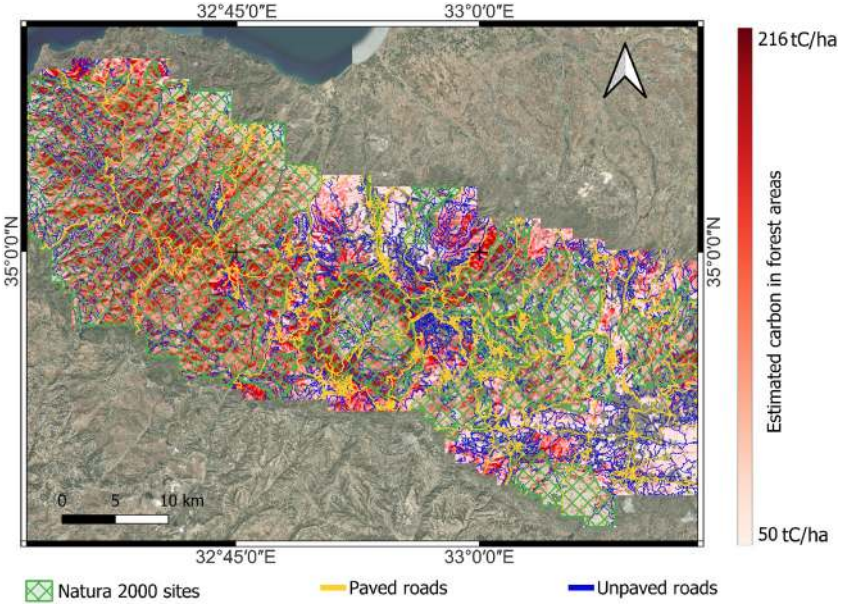


Fig. 4. Spatial distribution of estimated carbon stocks across the Troodos massif, overlaid with Natura 2000 sites and road infrastructure (Visualized in EPSG:4326 – WGS84 for global context; analyses performed in EPSG:32636).

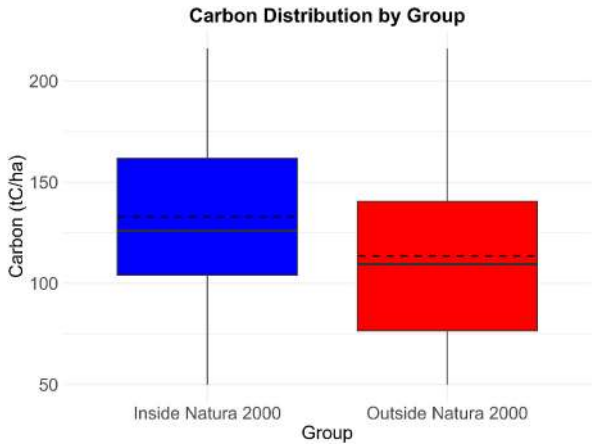


Fig. 5. Boxplots of carbon values in forest areas inside and outside Natura 2000 sites. Blue represents protected sites (inside Natura 2000), and red corresponds to non-protected (outside) areas. Boxes indicate the interquartile range (IQR), with the median shown as a solid horizontal line and the mean as a dashed line. Whiskers extend to 1.5 times the IQR, highlighting data variability and potential outliers.

the spatial distribution of carbon stocks across the Troodos massif, highlighting areas of high and low carbon density in relation to Natura 2000 sites and both paved and unpaved roads.

Subsequently, using the 10220 matched points (see Fig. S10 in Supplementary Material for an example of the matching technique), the mean value was calculated for the sites belonging to and not belonging to the Natura 2000 network. A mean value of 132.958 tC/ha was obtained for the treated group, or Natura sites group, and a mean value of 113.596 tC/ha for the control group, or non-Natura sites group. The Wilcoxon test revealed a statistically significant difference between the mean carbon values of both groups (p-value <0.001), and the g-computation calculation also indicated a significant difference in estimates between in and out of Natura 2000) sites of 16.25 tC/ha with an error of 1.57 tC/ha. These results are visually summarized in Fig. 5, which shows the distribution of carbon values across both groups using boxplots. The figure highlights higher median and mean carbon values within Natura 2000 sites compared to non-protected areas, suggesting a positive effect of protection status on carbon storage. Furthermore, the sensitivity analysis demonstrated that matching results were robust to unobserved biases up to a level of $\Gamma = 2$ (Table S7 in Supplementary Material), indicating that even in the presence of moderate selection bias—where treatment probabilities could double due to unobserved variables—the treatment effect remained significant.

The GAM was then fitted using a Gamma distribution with a logarithmic link function, providing a flexible and robust approach to capture potential nonlinear patterns and relationships in the data (section 3.3). Diagnostic analyses indicated that model assumptions were reasonably met, as residuals were approximately normally distributed, centered around zero, and displayed no major patterns or heteroscedasticity when plotted against fitted values or the linear predictor (see diagnostic plots in Fig. S11, Supplementary Material). In terms of predictor effects, smooth terms revealed important nonlinear patterns (see Fig. S12). Notably, both the distance to paved roads and its interaction with Natura 2000 designation were significant and showed nonlinear effects, with increasing distances generally associated with higher carbon stocks. In contrast, while the main effect of distance to all roads was not significant (Table S8), its interaction with Natura 2000 areas was, reinforcing the idea that proximity to infrastructure influences carbon differently within protected areas. Additionally, slope and altitude were both significant and exhibited complex nonlinear effects, while aspect showed a clear U-shaped relationship. These results highlight the utility of GAMs in uncovering nuanced spatial effects on carbon storage, also confirming that the model provided an adequate and well-calibrated fit to the data.

The role of road-related variables in shaping carbon stocks is further illustrated in Fig. 6, which presents the smooth functions estimated by the GAM for areas surrounding infrastructure, accompanied by their associated uncertainty and linear trend lines. As shown, Natura 2000 areas (blue) generally exhibited higher carbon stocks than non-protected zones (red). For distance to paved roads (left panel), carbon stocks tend to increase within Natura 2000 sites and decrease outside them, based on the linear trend over the first 1000 m. In contrast, when considering all roads—both paved and unpaved—carbon stocks decrease with distance, being more pronounced within Natura 2000 sites. Additionally, Table S9 (Supplementary Material) provides more details on carbon storage per buffer around paved roads, total stored carbon, and the area of each buffer.

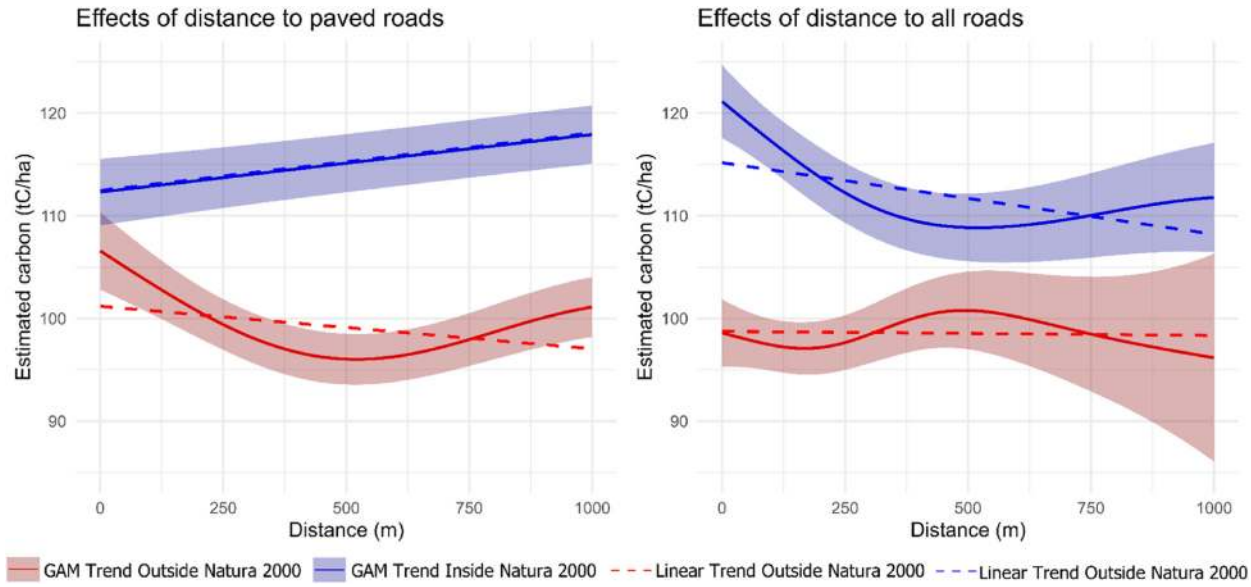


Fig. 6. Variation in carbon stocks in relation to roads. Left: Considering only distances to paved roads. Right: Considering distances to all types of roads.

5. Discussion

The analysis revealed that forested sites under Natura 2000 protection store significantly more carbon than unprotected areas, with an average increase of 17 %, supporting the conservation value of such areas. Particularly, this difference may partly reflect the long-standing protection of forested lands in the Troodos massif, which were subject to legal and customary conservation efforts well before their formal inclusion in the Natura 2000 network, thus reflecting the effectiveness of long-term conservation measures in enhancing carbon storage capacity. For instance, early forest legislation such as the 1879 Ordinance and the 1881 Forest Law brought all forests under state control, establishing legal frameworks that restricted exploitative uses and regulated activities like grazing, which was progressively phased out in state forests by 1950. Combined with additional protections—such as the declaration of Paphos Forest as a game reserve in 1940—these long-term policies helped preserve forest structure and integrity over decades (FAO, 1952; Kyriacou, 2006). More recently, European-funded projects have reinforced these conservation outcomes through restoration actions and adaptive forest management within Natura 2000 sites (European Commission, 2025; Georghiou et al., 2008).

In light of this context, the observed difference—although minor, possibly due to the specific characteristics of the Mediterranean region analyzed—aligns with the findings of Duncanson et al. (2023), who reported that AGC was 28 % higher in protected areas compared to globally matched unprotected sites. This difference in carbon storage between protection regimes could be attributed to two main factors. On the one hand, the effectiveness of protection frameworks facilitates the implementation of adaptive management measures and restrictions that minimize degrading activities, such as deforestation or urbanization, as demonstrated by Kallimanis et al. (2015) or Mammides et al. (2024). These strategies enhance the resilience of forest ecosystems, ensuring vegetation regeneration and contributing to increased carbon stocks. On the other hand, as noted by Miranda et al. (2016), protected areas are often located in remote or inaccessible regions, which are typically less affected by human activities and land-use intensification. Such spatial characteristics contribute to the preservation of natural ecosystem structure and function, facilitating continuous tree growth and biomass accumulation. This pattern appears to be particularly relevant in certain parts of the study area, such as the Paphos Forest, where limited anthropogenic disturbance has likely supported higher carbon retention, even though other regions have historically experienced varying degrees of human pressure (Republic of Cyprus - The Mines Service, 2025). Therefore, these findings highlight the critical role of effectively managed protected areas in mitigating climate change and fostering long-term environmental sustainability, as also emphasized by Cannizzo et al. (2024).

The analysis of road networks and carbon stocks in forested areas revealed complex interactions influenced by the type of road and the protection status. Proximity to roads, considering the entire network—both paved and unpaved—was associated with an increase in carbon stocks, aligning with Vepakomma et al. (2018) and Kalinaki et al. (2023), with this effect being especially pronounced in protected areas, highlighting the statistical significance of the interaction between Natura 2000 status and distance to all roads. This pattern may be linked to the predominance of unpaved roads, which account for 75 % of the network analyzed (over 11,500 km vs. 3900 km of paved roads). These roads extend deeper into natural landscapes, increasing light penetration, which may enhance productivity and regeneration, thus promoting higher carbon storage along the roadsides. Nevertheless, while increased light availability may stimulate biomass accumulation, roadside edges could simultaneously induce structural degradation, microclimatic alterations, or species turnover, potentially compromising ecological integrity (Sun et al., 2025; Zhou et al., 2020). In addition, including this type of distance to the entire road network in the analysis also introduced greater uncertainty with increasing distance, primarily due to

fewer data points. Conversely, when distance to paved roads was considered, an opposite trend was observed between protected and non-protected areas. Within Natura 2000 sites, carbon stocks increased with greater distance from roads, in accordance with the results of Hu et al. (2017), whereas outside Natura 2000 sites, carbon stocks tend to decrease with distance, possibly due to differences in LULC between protection regimes, with more forest cover, particularly complete CC class, in Natura 2000 sites, while outside, cropland and forests with hollow CC class are more prevalent (see Supplementary Material Figs. S13–S15). It is also important to note that all these findings also align with prior research on edge effects, which presents varying perspectives. For example, Pöpperl and Seidl (2021) noted that edges can degrade forest structure but also promote biodiversity, while Delgado et al. (2007) and Dormann et al. (2020) highlighted improvements due to increased light availability. Therefore, further research is needed to better understand these dynamics, especially considering how protection regimes and management interact with road networks, rather than focusing solely on road type, as suggested by Braham et al. (2023) or Vepakomma et al. (2018). Notably, to our knowledge, this is the first study to examine edge effects on carbon storage across multiple road distances while explicitly accounting for protection regimes, highlighting the need for a more detailed approach in future analyses.

As for the top-performing ML models used in this study—ANN, SVM, and GB—demonstrated strong performance in classifying LULCs using orthophotos and Landsat images, effectively estimating key variables such as CC. The final ensemble model, combining these top-performing algorithms (ANN, SVM, and GB), outperformed individual models, achieving a Kappa coefficient of 0.91 and 92 % accuracy. This ensemble approach leveraged each algorithm's strengths while mitigating weaknesses, as demonstrated by Du et al. (2023) and Subedi et al. (2024), who reported comparable accuracies of 96 % and 94 %, respectively. In contrast, Kovárník and Janová (2025) obtained comparable accuracies using simple ML models. Consequently, for this region, traditional methods, such as the SVM-based classification from orthophotos (section 3.1), achieved satisfactory results, with an accuracy of 90 % and a Kappa coefficient of 0.88 in the test set, while ensemble models excelled in identifying complex patterns in heterogeneous environments and diverse classes, such as CC in Mediterranean island ecosystems.

Concerning the limitations encountered, the use of InVEST stands out due to its inherent uncertainties in representing complex forest ecosystem processes. Certain factors such as droughts, pests, silvicultural treatments, and fires are not dynamically incorporated, potentially altering projections during ecological stability, as suggested by Nunery and Keeton (2010). Carbon stock estimates, which assume constant values for the various LULC categories, simplify the influence of these biophysical factors. This assumption may overlook critical biophysical processes that significantly impact growth and carbon storage. Additionally, the linear approach of the InVEST model disregards variables such as nonlinear tree growth and soil chemistry changes (Dang et al., 2017), potentially underestimating or overestimating carbon stocks (Ouyang et al., 2016). Although the model outputs do not provide explicit error margins, efforts were taken to minimize estimation biases by applying species-specific allometric equations when available and measuring both dominant and suppressed individuals across mono- and mixed-species plots, ensuring a representative assessment of stand-level biomass. Thus, future studies should integrate nonlinear models, accounting for climatic and abiotic factors, such as those developed in Ayushi et al. (2024), Suárez-Fernández et al. (2025) or Zurqani (2025), to improve carbon estimates, better quantify potential uncertainties, and refine comparisons with InVEST-based predictions by evaluating possible deviations.

Another limitation was the scope of field sampling, which focused on an insular-Mediterranean ecosystem. While this may limit extrapolation to other biomes or regions, it is the first study of its kind on islands, offering a novel reference for similar contexts. As with any sampling approach, there is an inherent limitation in fully capturing the spatial heterogeneity of forest stands. In this case, carbon stocks may vary with factors such as altitude, aspect, and forest age, which might not be adequately reflected in the existing plots. Despite this, the findings provide valuable insights applicable to other Mediterranean islands—often underrepresented in scientific literature—such as Corsica, Sardinia, or Sicily. Future research should include more diverse sampling locations, extend data collection to continental ecosystems, and incorporate slopes and altitudinal gradients across various CC categories to enhance representativeness and validate methodologies across broader environmental contexts.

Finally, it is important to note that limitations related to the spatial and temporal resolution of the RS data were identified. Landsat images—spatial resolution of 30 m— may struggle to capture small-scale changes, leading to potential misclassification in complex LULC mosaics. This could affect the accuracy of LULC classifications and, therefore, carbon stock estimates in transition zones. Higher-resolution data, such as Sentinel-2, offer better boundary delineation but are limited temporally, as images are only available since 2015, preventing alignment with reference-period orthophotos. Moreover, persistent cloud cover and limited revisit frequency—particularly during key seasons—could further restrict the availability of optimal scenes, especially in mountainous regions. Future studies should prioritize higher spatial resolution data with consistent temporal and spatial coverage, and incorporate the most recent officially released datasets whenever available.

Despite the methodological limitations noted, the results provided valuable insights for ecological protection and land management on Mediterranean islands such as Cyprus, where forest inventories are often deficient and rely on limited sampling approaches (Zenonos et al., 2025). This hampers their effectiveness in directly linking forest structure and carbon stocks with RS data, thereby limiting accurate and up-to-date environmental planning. In this context, the application of the InVEST model enabled a robust estimation of carbon storage (García-Ontiyuelo et al., 2024), offering an objective analytical framework for regions with scarce geospatial forest information. In addition, the increasing availability of cloud-based platforms—such as Google Earth Engine (GEE), Amazon Web Services (AWS), and the Deep Earth System Data Lab (DeepESDL)—is further enhancing the capacity to perform large-scale, reproducible carbon stock assessments by facilitating access to multi-source RS data and scalable computational tools (Khachoo et al., 2024; Zhao et al., 2022). The findings indicated that Natura 2000 areas hold higher carbon levels, possibly due, as previously discussed, to a greater presence of fully covered forest stands. This pattern—along with other documented environmental benefits of these protected areas, such as improved soil conservation (Hagyó and Tóth, 2018) and reduced land conversion to artificial surfaces (Mammides et al., 2024)—reinforces the requirement for local governments to prioritize the expansion, proper management,

and interconnection of such areas. In particular, integrating carbon storage criteria into spatial planning, road infrastructure development, and forest management strategies could substantially strengthen climate change mitigation efforts by providing decision-makers with tools based on objective territorial characteristics.

In essence, the findings demonstrate the crucial role of Natura 2000 sites in conserving and enhancing ecosystem services, particularly their effectiveness in carbon storage. They also emphasize the influence of road networks on carbon stocks, which depends on both the distance to roads and the applied protection regime. This underscores the need for proper management of roadside vegetation in protected areas to maximize environmental benefits, and in which future research will continue to explore these dynamics to develop more effective and adaptive conservation strategies.

6. Conclusion

This study emphasizes the critical role of combining advanced ML techniques with environmental modeling tools to improve forest ecosystem management and understanding. The integration of an ensemble-based LULC classification approach with ecosystem service models allowed for a highly accurate subdivision of forest classes based on CC, enabling a more detailed assessment of carbon stocks. The findings underscore the crucial role of Natura 2000 sites in increasing carbon storage and reveal the complex interactions between road networks, protection regime and carbon stocks. Specifically, carbon distribution varies depending on both the distance to roads and the applied protection regime. These insights emphasize the need for adaptive conservation strategies that integrate ML, RS, and ecosystem modeling to support sustainable forest management. Future research should further explore these dynamics to refine land-use policies and enhance conservation and management planning.

CRediT authorship contribution statement

Gabriel E. Suárez-Fernández: Writing – review & editing, Writing – original draft, Validation, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Savvas Zotos:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Methodology. **Joaquín Martínez-Sánchez:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Investigation, Funding acquisition. **Marilena Stamatiou:** Writing – review & editing, Supervision, Resources. **Elli Tzirkalli:** Writing – review & editing, Supervision, Resources. **Ioannis N. Vogiatzakis:** Writing – review & editing, Writing – original draft, Supervision. **Pedro Arias:** Writing – review & editing, Writing – original draft, Supervision.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Ethical statement

We declare that all ethical practices have been followed in relation to the development, writing, and publication of the article.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gabriel E Suarez-Fernandez reports financial support was provided by Gobierno de Espana Ministerio de Educación Cultura y Deporte. Gabriel E Suarez-Fernandez reports financial support was provided by European Cooperation in Science and Technology (COST). Joaquin Matinez-Sanchez reports financial support was provided by State Agency of Research. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by "Ayudas para la Formación de Profesorado Universitario (FPU)", Gobierno de Espana Ministerio de Educación Cultura y Deporte [FPU21/03038] and by COST Action (CA21158) Enhancing Small-nMedium IsLads resilience by securing the sustainability of Ecosystem Services (SMILES). The publication is part of the project PLEC2023-010215, funded by MCIN/AEI/10.13039/501100011033. Funding for open access charge: Universidade de Vigo/CISUG.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rsase.2025.101713.

Data availability

The developed codes and results are available at https://doi.org/10.5281/zenodo.16894162. Additionally, satellite images, along with other input data, can be obtained by contacting the corresponding author or through the references listed in the text.

References

Adhikari, A., Montes, C.R., Peduzzi, A., 2023. A comparison of modeling methods for predicting forest attributes using lidar metrics. Remote Sens. 15 (5), 1284. https://doi.org/10.3390/rs15051284.

Almeida, B., Monteiro, L., Tiengo, R., Gil, A., Cabral, P., 2025. Spatially explicit assessment of carbon storage and sequestration in forest ecosystems. Remote Sens. Appl.: Soc. Environ. 38, 101544. https://doi.org/10.1016/j.rsase.2025.101544.

Augustynczik, A.L.D., Gusti, M., di Fulvio, F., Lauri, P., Forsell, N., Havlík, P., 2024. Modelling the effects of climate and management on the distribution of deadwood in European forests. J. Environ. Manag. 354. https://doi.org/10.1016/j.jenvman.2024.120382.

Ayushi, K., Babu, K.N., Ayyappan, N., Nair, J.R., Kakkara, A., Reddy, C.S., 2024. A comparative analysis of machine learning techniques for aboveground biomass estimation: a case study of the Western Ghats, India. Ecol. Inform. 80. https://doi.org/10.1016/j.ecoinf.2024.102479.

Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M.L., Bareth, G., 2015. Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. Int. J. Appl. Earth Obs. Geoinf. 39, 79–87. https://doi.org/10.1016/j.jag.2015.02.012.

Bera, D., Das Chatterjee, N., Bera, S., Ghosh, S., Dinda, S., 2023. Comparative performance of Sentinel-2 MSI and Landsat-8 OLI data in canopy cover prediction using Random Forest model: comparing model performance and tuning parameters. Adv. Space Res. 71 (11), 4691–4709. https://doi.org/10.1016/j.asr.2023.01.027.

Bhungeni, O., Ramjatan, A., Gebreslasie, M., 2024. Evaluating machine-learning algorithms for mapping LULC of the uMngeni catchment area, KwaZulu-Natal. Remote Sens. 16 (12), 2219. https://doi.org/10.3390/rs16122219.

Braham, N., Valeria, O., Imbeau, L., 2023. Characterization of vegetation dynamics on linear features using airborne laser scanning and ensemble learning. Forests 14 (3). https://doi.org/10.3390/f14030511.

Buss, J., Dabros, A., Higgins, K.L., Hammond, H.E.J., Pinzon, J., Langor, D.W., 2024. Comparison of edge effects from well pads and industrial roads on mixed upland boreal forest vegetation in Alberta. Plant Ecol. https://doi.org/10.1007/s11258-023-01393-3.

Camera, C., Zomeni, Z., Noller, J.S., Zissimos, A.M., Christoforou, I.C., Bruggeman, A., 2017. A high resolution map of soil types and physical properties for Cyprus: a digital soil mapping optimization. Geoderma 285, 35–49. https://doi.org/10.1016/j.geoderma.2016.09.019.

Cannizzo, Z.J., Belle, E.M.S., Smith, R.B., Mommsen, T.P., 2024. Climate change—protected areas as a tool to address a global crisis. In: Managing Protected Areas. Springer International Publishing, pp. 295–325. https://doi.org/10.1007/978-3-031-40783-3_16.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A., 2020. A comprehensive survey on support vector machine classification: applications, challenges and trends. Neurocomputing 408, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118.

Chinembiri, T.S., Mutanga, O., Dube, T., 2023. Hierarchical Bayesian geostatistics for C stock prediction in disturbed plantation forest in Zimbabwe. Ecol. Inform. 73. https://doi.org/10.1016/j.ecoinf.2022.101934.

Clark, N.J., Wells, K., 2023. Dynamic Generalised Additive Models (DGAMs) for forecasting discrete ecological time series. Methods Ecol. Evol. 14 (3), 771–784. https://doi.org/10.1111/2041-210X.13974.

Cruz-Pérez, N., Santamarta, J.C., Rodríguez-Martín, J., García-Gil, A., Márquez, J.H., Morales, C.G.R., Castañeda, I.D., Yuste, R.J.P., Rayo, M.E., 2023. Net carbon balance study for selected roads in Tenerife, Canary Islands. Integrated Environ. Assess. Manag. 19 (4), 1023–1030. https://doi.org/10.1002/ieam.4719.

Dang, X., Liu, G., Zhao, L., Zhao, G., 2017. The response of carbon storage to the age of three forest plantations in the Loess Hilly Regions of China. Catena 159, 106–114. https://doi.org/10.1016/j.catena.2017.08.013.

de Araujo Fonseca, F.N., da Cunha Bustamante, M.M., 2025. The importance of indigenous territories for the provision of ecosystem services: a case study in the Brazilian Cerrado-Amazon Transition. Ecosyst. Serv. 72, 101706. https://doi.org/10.1016/j.ecoser.2025.101706.

Delgado, J.D., Arroyo, N.L., Arévalo, J.R., Fernández-Palacios, J.M., 2007. Edge effects of roads on temperature, light, canopy cover, and canopy height in laurel and pine forests (Tenerife, Canary Islands). Landsc. Urban Plann. 81 (4), 328–340. https://doi.org/10.1016/j.landurbplan.2007.01.005.

Di Gregorio, Antonio, Jansen, L.J.M., 2005. Land Cover Classification System Classification Concepts and User Manual Software Version (2). Food and Agriculture Organization of the United Nations (FAO).

Dormann, C.F., Bagnara, M., Boch, S., Hinderling, J., Janeiro-Otero, A., Schäfer, D., Schall, P., Hartig, F., 2020. Plant species richness increases with light availability, but not variability, in temperate forests understorey. BMC Ecol. 20 (1). https://doi.org/10.1186/s12898-020-00311-9.

Du, H., Li, M., Xu, Y., Zhou, C., 2023. An ensemble learning approach for land use/land cover classification of arid regions for climate simulation: a case study of Xinjiang, Northwest China. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 16, 2413–2426. https://doi.org/10.1109/JSTARS.2023.3247624.

Duncanson, L., Liang, M., Leitold, V., Armston, J., Krishna Moorthy, S.M., Dubayah, R., Costedoat, S., Enquist, B.J., Fatoyinbo, L., Goetz, S.J., Gonzalez-Roglich, M., Merow, C., Roehrdanz, P.R., Tabor, K., Zvoleff, A., 2023. The effectiveness of global protected areas for climate change mitigation. Nat. Commun. 14 (1). https://doi.org/10.1038/s41467-023-38073-9.

European Commission, 2025. Integrated conservation management of priority habitat type 9590* in the Natura 2000 site Koilada Kedron-Kampos - LIFE-KEDROS. LIFE Public Database.

European Council. Council of the European Union, 2024. Fit for 55. https://www.consilium.europa.eu/en/policies/fit-for-55/.

European Environment Agency, 2024. Natura 2000 data - the European network of protected sites. https://doi.org/10.2909/0801ff11-24ae-4c6b-852f-029965df5bc5.

European Environment Agency, & Joint research centre. (n.d.). CORINE Land Cover. Retrieved August 1, 2024, from https://land.copernicus.eu/en/products/corine-land-cover.

Evans, D., 2012. Building the European Union's natura 2000 network. Nat. Conserv. 1, 11–26. https://doi.org/10.3897/natureconservation.1.1808.

FAO, 1952. Forestry in the Middle East. In: Unasylva, third ed., vol. 6. FAO https://www.fao.org/4/x5364e/x5364e00.htm#Contents.

FAO, 2001. Global Forest Resources Assessment 2000.

Galvão, R.K.H., Araújo, M.C.U., 2009. Variable selection. In: Comprehensive Chemometrics. Elsevier, pp. 233–283. https://doi.org/10.1016/B978-044452701-1.00075-2.

Gamon, J.A., Surfus, J.S., 1999. Assessing leaf pigment content and activity with a reflectometer. New Phytol. 143 (1), 105–117. https://doi.org/10.1046/j.1469-8137.1999.00424.x.

García-Ontiyuelo, M., Acuña-Alonso, C., Valero, E., Álvarez, X., 2024. Geospatial mapping of carbon estimates for forested areas using the InVEST model and Sentinel-2: a case study in Galicia (NW Spain). Sci. Total Environ. 922. https://doi.org/10.1016/j.scitotenv.2024.171297.

Gasparini, P., Di Cosmo, L., Floris, A., De Laurentis Editors, D., 2022. Italian National Forest Inventory—Methods and Results of the Third Survey Inventario Nazionale delle Foreste e dei Serbatoi Forestali di Carbonio—Metodi e Risultati della Terza Indagine. Springer. https://doi.org/10.1007/978-3-030-98678-0.

Georghiou, K., Delipetrou, P., Andreou, M., Kardakari, N., Zotos, S., 2008. Conservation management in NATURA 2000 sites of Cyprus. In: Layman's Report. https://webgate.ec.europa.eu/life/publicWebsite/project/LIFE04-NAT-CY-000013/conservation-management-in-natura-2000-sites-of-cyprus.

Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80 (1), 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9.

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett. 30 (5). https://doi.org/10.1029/2002GL016450.

Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. ISPRS J. Photogrammetry Remote Sens. 116, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.
Graham, V., Geldmann, J., Adams, V.M., Negret, P.J., Sinovas, P., Chang, H.C., 2021. Southeast Asian protected areas are effective in conserving forest cover and forest carbon stocks compared to unprotected areas. Sci. Rep. 11 (1). https://doi.org/10.1038/s41598-021-03188-w.
Guerschman, J.P., Hill, M.J., Renzullo, L.J., Barrett, D.J., Marks, A.S., Botha, E.J., 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. Remote Sens. Environ. 113 (5), 928–945. https://doi.org/10.1016/j.rse.2009.01.006.
Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157 (2–3), 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1.
Hague, T., Tillett, N.D., Wheeler, H., 2006. Automated crop and weed monitoring in widely spaced cereals. Precis. Agric. 7 (1), 21–32. https://doi.org/10.1007/s11119-005-6787-1.
Hagyó, A., Tóth, G., 2018. The impact of environmental policy on soil quality: organic carbon and phosphorus levels in croplands and grasslands of the European Natura 2000 network. J. Environ. Manag. 223, 9–15. https://doi.org/10.1016/j.jenvman.2018.06.003.
Harper, K.A., Macdonald, S.E., Burton, P.J., Chen, J., Brosofske, K.D., Saunders, S.C., Euskirchen, E.S., Roberts, D., Jaiteh, M.S., Esseen, P.A., 2005. Edge influence on forest structure and composition in fragmented landscapes. Conserv. Biol. 19 (3), 768–782. https://doi.org/10.1111/j.1523-1739.2005.00045.x.
Hernández-Guzmán, R., Ruiz-Luna, A., González, C., 2019. Assessing and modeling the impact of land use and changes in land cover related to carbon storage in a western basin in Mexico. Remote Sens. Appl.: Soc. Environ. 13, 318–327. https://doi.org/10.1016/j.rsase.2018.12.005.
Ho, D.E., Imai, K., King, G., Stuart, E.A., 2011. MatchIt: nonparametric preprocessing for parametric causal inference. JSS J. Statist. Software 42. http://www.jstatsoft.org/.
Hu, X., Zhang, L., Ye, L., Lin, Y., Qiu, R., 2017. Locating spatial variation in the association between road network and forest biomass carbon accumulation. Ecol. Indic. 73, 214–223. https://doi.org/10.1016/j.ecolind.2016.09.042.
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83 (1–2), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.
Jennings, S., 1999. Assessing forest canopies and understorey illumination: canopy closure, canopy cover and other measures. Forestry 72 (1), 59–74. https://doi.org/10.1093/forestry/72.1.59.
Jolliffe, I.T., 2002. Principal Component Analysis, second ed. Springer-Verlag. https://doi.org/10.1007/b98835.
Joy, M.S., Jha, P., Yadav, P.K., Bansal, T., Rawat, P., Begam, S., 2024. Forest fragmentation and forest cover dynamics: mining induced changes in the West Singhbhum District of Jharkhand. Remote Sens. Appl.: Soc. Environ. 36, 101350. https://doi.org/10.1016/j.rsase.2024.101350.
Kalinaki, K., Malik, O.A., Lai, D.T.C., Sukri, R.S., Wahab, R.B.H.A., 2023. Spatial-temporal mapping of forest vegetation cover changes along highways in Brunei using deep learning techniques and Sentinel-2 images. Ecol. Inform. 77. https://doi.org/10.1016/j.ecoinf.2023.102193.
Kallimanis, A.S., Touloumis, K., Tzanopoulos, J., Mazaris, A.D., Apostolopoulou, E., Stefanidou, S., Scott, A.V., Potts, S.G., Pantis, J.D., 2015. Vegetation coverage change in the EU: patterns inside and outside Natura 2000 protected areas. Biodivers. Conserv. 24 (3), 579–591. https://doi.org/10.1007/s10531-014-0837-9.
Kanjin, K., Alam, B.M., 2024. Assessing changes in land cover, NDVI, and LST in the Sundarbans mangrove forest in Bangladesh and India: a GIS and remote sensing approach. Remote Sens. Appl.: Soc. Environ. 36, 101289. https://doi.org/10.1016/j.rsase.2024.101289.
Key, C.H., Benson, N.C., 2006. Landscape assessment (LA): sampling and analysis methods. In: Lutes, D.C. (Ed.), FIREMON: Fire Effects Monitoring and Inventory System. USDA Forest Service. https://www.researchgate.net/publication/241688462.
Khachoo, Y.H., Cutugno, M., Robustelli, U., Pugliano, G., 2024. Impact of Land Use and Land Cover (LULC) changes on carbon stocks and economic implications in Calabria using Google Earth Engine (GEE). Sensors 24 (17), 5836. https://doi.org/10.3390/s24175836.
Kovárník, R., Janová, J., 2025. Validation of sentinel 2 based machine learning models for Czech National Forest Inventory. Ecol. Inform. 87. https://doi.org/10.1016/j.ecoinf.2025.103133.
Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. JSS J. Statist. Software 36. http://www.jstatsoft.org/.
Kyriacou, K., 2006. Protected forest areas in Europe – analysis and harmonisation (PROFOR) - Country report - Cyprus. In: COST Action E27. https://bfw.ac.at/020/profor/.
Lei, G., Li, A., Bian, J., Yan, H., Zhang, L., Zhang, Z., Nan, X., 2020. OIC-MCE: a practical land cover mapping approach for limited samples based on multiple classifier ensemble and iterative classification. Remote Sens. 12 (6), 987. https://doi.org/10.3390/rs12060987.
Li, Y., Yuan, L., Song, Z., Yu, S., Zhang, X., Tian, B., Liu, M., 2024. Salt marsh carbon stock estimation using deep learning with Sentinel-1 SAR of the Yangtze River estuary, China. Int. J. Appl. Earth Obs. Geoinf. 133. https://doi.org/10.1016/j.jag.2024.104138.
López-Serrano, P., Corral-Rivas, J., Díaz-Varela, R., Álvarez-González, J., López-Sánchez, C., 2016. Evaluation of radiometric and atmospheric correction algorithms for aboveground forest biomass estimation using Landsat 5 TM data. Remote Sens. 8 (5), 369. https://doi.org/10.3390/rs8050369.
Louhaichi, M., Borman, M.M., Johnson, D.E., 2001. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. Geocarto Int. 16 (1), 65–70. https://doi.org/10.1080/10106040108542184.
Malik, A.D., Arief, M.C.W., Withaningsih, S., Parikesit, P., 2024. Modeling regional aboveground carbon stock dynamics affected by land use and land cover changes. Global J. Environ. Sci. Manag. 10 (1), 245–266. https://doi.org/10.22034/gjesm.2024.01.16.
Mammides, C., Zotos, S., Martini, F., 2024. Quantifying the amount of land lost to artificial surfaces in European habitats: a comparison inside and outside Natura 2000 sites using a quasi-experimental design. Biol. Conserv. 293. https://doi.org/10.1016/j.biocon.2024.110556.
Ministry of Agriculture Rural Development and Environment of Cyprus, 2019. Cyprus' National Forestry Accounting Plan in Accordance with Regulation (EU) 2018/41.
Miranda, J.J., Corral, L., Blackman, A., Asner, G., Lima, E., 2016. Effects of protected areas on forest cover change and local communities: evidence from the Peruvian Amazon. World Dev. 78, 288–307. https://doi.org/10.1016/j.worlddev.2015.10.026.
Müllerová, J., Vítková, M., Vítek, O., 2011. The impacts of road and walking trails upon adjacent vegetation: effects of road building materials on species composition in a nutrient poor environment. Sci. Total Environ. 409 (19), 3839–3849. https://doi.org/10.1016/j.scitotenv.2011.06.056.
NASA Shuttle Radar Topography Mission (SRTM), 2013. Shuttle Radar Topography Mission (SRTM) global. OpenTopography. https://doi.org/10.5069/G9445JDF.
Natural Capital Project, 2024. InVEST 0.0. Stanford University, University of Minnesota, Chinese Academy of Sciences. In: The Nature Conservancy, World Wildlife Fund. Stockholm Resilience Centre and the Royal Swedish Academy of Sciences. https://naturalcapitalproject.stanford.edu/software/invest.
Naumann, S., Noebe, R., Gaudillat, Zelmina, Stein, U., Röschel, L., Ittner, S., Davis, M., Staneva, A., Rutherford, C., Romão, C., 2020. State of nature in the EU - results from reporting under the nature directives 2013-2018. https://doi.org/10.2800/705440.
Ng, V.K.Y., Cribbie, R.A., 2019. The gamma generalized linear model, log transformation, and the robust Yuen-Welch test for analyzing group means with skewed and heteroscedastic data. Commun. Stat. Simulat. Comput. 48 (8), 2269–2286. https://doi.org/10.1080/03610918.2018.1440301.
Nunery, J.S., Keeton, W.S., 2010. Forest carbon storage in the northeastern United States: net effects of harvesting frequency, post-harvest retention, and wood products. For. Ecol. Manag. 259 (8), 1363–1375. https://doi.org/10.1016/j.foreco.2009.12.029.
Ouyang, Z., Zheng, H., Xiao, Y., Polasky, S., Liu, J., Xu, W., Wang, Q., Zhang, L., Xiao, Y., Rao, E., Jiang, L., Lu, F., Wang, X., Yang, G., Gong, S., Wu, B., Zeng, Y., Yang, W., Daily, G.C., 2016. Improvements in ecosystem services from investments in natural capital. Science 352 (6292), 1455–1459. https://doi.org/10.1126/science.aaf2295.
Panagos, P., De Rosa, D., Liakos, L., Labouyrie, M., Borrelli, P., Ballabio, C., 2024. Soil bulk density assessment in Europe. Agric. Ecosyst. Environ. 364. https://doi.org/10.1016/j.agee.2024.108907.
Park, Y.-S., Lek, S., 2016. Artificial Neural Networks, pp. 123–140. https://doi.org/10.1016/B978-0-444-63623-2.00007-4.

Penman, J., Gytarsky, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Wagner, F., 2003. Good Practice Guidance for Land Use, Land-Use Change and Forestry. Published by the Institute for Global Environmental Strategies for the IPCC.

Petersson, H., Holm, S., Ståhl, G., Alger, D., Fridman, J., Lehtonen, A., Lundström, A., Mäkipää, R., 2012. Individual tree biomass equations or biomass expansion factors for assessment of carbon stock changes in living biomass – a comparative study. For. Ecol. Manag. 270, 78–84. https://doi.org/10.1016/j.foreco.2012.01.004.

Pöpperl, F., Seidl, R., 2021. Effects of stand edges on the structure, functioning, and diversity of a temperate mountain forest landscape. Ecosphere 12 (8). https://doi.org/10.1002/ecs2.3692.

Posit Software, 2025. RStudio desktop. https://posit.co/download/rstudio-desktop/.

Prodromou, M., Theocharidis, C., Gitas, I.Z., Eliades, F., Themistocleous, K., Papasavvas, K., Dimitrakopoulos, C., Danezis, C., Hadjimitsis, D., 2024. Forest habitat mapping in Natura2000 regions in Cyprus using sentinel-1, sentinel-2 and topographical features. Remote Sens. 16 (8). https://doi.org/10.3390/rs16081373.

Public Administration and Personnel Department, 2024. *National open data portal* . Ministry of finance. https://data.gov.cy/en.

QGIS, 2025. QGIS overview. https://qgis.org/project/overview/.

R Core Team, 2024. The R Project for Statistical Computing. R Foundation. https://www.r-project.org/.

Republic of Cyprus - The Mines Service, 2025. The Mines services. https://www.moa.gov.cy/moa/Mines/MinesSrv.nsf/dmlindex_en/.

Rey, D., Neuhäuser, M., 2011. Wilcoxon-signed-rank test. In: International Encyclopedia of Statistical Science. Springer Berlin Heidelberg, pp. 1658–1659. https://doi.org/10.1007/978-3-642-04898-2_616.

Rimal, B., Sharma, R., Kunwar, R., Keshtkar, H., Stork, N.E., Rijal, S., Rahman, S.A., Baral, H., 2019. Effects of land use and land cover change on ecosystem services in the Koshi River Basin, Eastern Nepal. Ecosyst. Serv. 38. https://doi.org/10.1016/j.ecoser.2019.100963.

Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55 (2), 95–107. https://doi.org/10.1016/0034-4257(95)00186-7.

Rosenbaum, P.R., 2002. Observational Studies. Springer, New York. https://doi.org/10.1007/978-1-4757-3692-2.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the great plains with ERTS. 3rd ERTS Symposium, NASA SP-351, 309–317.

Sales, V., Marques, A., Racolte, G., Nunes, A., Guimaraes, T., Zanotta, D., Spigolon, A., Gonzaga, L., Roberto Veronez, M., 2023. Evaluation of resampling techniques to provide better synthesized input data to super-resolution deep learning model training. International Geoscience and Remote Sensing Symposium (IGARSS) 7368–7371. https://doi.org/10.1109/IGARSS52108.2023.10281470, 2023-July.

Schleicher, J., Eklund, J., Barnes, M.D., Geldmann, J., Oldekop, J.A., Jones, J.P.G., 2020. Statistical matching for conservation science. Conserv. Biol. 34 (3), 538–549. https://doi.org/10.1111/cobi.13448.

SENDECO2. (n.d.). Precios CO2. co2 Prices. Retrieved January 9, 2025, from https://www.sendeco2.com/es/precios-co2.

Shi, Y., Yang, K., Yang, Z., Zhou, Y., 2022. Primer on artificial intelligence. In: Mobile Edge Artificial Intelligence. Elsevier, pp. 7–36. https://doi.org/10.1016/B978-0-12-823817-2.00011-5.

Snowden, J.M., Rose, S., Mortimer, K.M., 2011. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. Am. J. Epidemiol. 173 (7), 731–738. https://doi.org/10.1093/aje/kwq472.

Sripada, R.P., Heiniger, R.W., White, J.G., Meijer, A.D., 2006. Aerial color infrared photography for determining early in-season nitrogen requirements in corn. Agron. J. 98 (4), 968–977. https://doi.org/10.2134/agronj2005.0200.

Suárez-Fernández, G.E., Martínez-Sánchez, J., Arias, P., 2025. Enhancing carbon stock estimation in forests: integrating multi-data predictors with random forest method. Ecol. Inform., 102997 https://doi.org/10.1016/j.ecoinf.2025.102997.

Subedi, M.R., Portillo-Quintero, C., McIntyre, N.E., Kahl, S.S., Cox, R.D., Perry, G., Song, X., 2024. Ensemble machine learning on the fusion of sentinel time series imagery with high-resolution orthoimagery for improved land use/land cover mapping. Remote Sens. 16 (15), 2778. https://doi.org/10.3390/rs16152778.

Sun, M., Li, W., Zhu, L., Guo, Z., Zhao, Z., Meng, N., Han, M., Wang, N., Zhang, X., 2025. Degradation in edge forests caused by forest fragmentation. Carbon Res. 4 (1), 38. https://doi.org/10.1007/s44246-025-00206-8.

Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., Liang, X., 2024. An improved random forest based on the classification accuracy and correlation measurement of decision trees. Expert Syst. Appl. 237, 121549. https://doi.org/10.1016/j.eswa.2023.121549.

Tamiminia, H., Salehi, B., Mahdianpari, M., Goulden, T., 2024. State-wide forest canopy height and aboveground biomass map for New York with 10 m resolution, integrating GEDI, Sentinel-1, and Sentinel-2 data. Ecol. Inform. 79. https://doi.org/10.1016/j.ecoinf.2023.102404.

Tang, H., Armston, J., Hancock, S., Marselis, S., Goetz, S., Dubayah, R., 2019. Characterizing global forest canopy cover distribution using spaceborne lidar. Remote Sens. Environ. 231, 111262. https://doi.org/10.1016/j.rse.2019.111262.

Teodoro, A.C., Duarte, L., 2022. The role of satellite remote sensing in natural disaster management. In: Nanotechnology-Based Smart Remote Sensing Networks for Disaster Prevention. Elsevier, pp. 189–216. https://doi.org/10.1016/B978-0-323-91166-5.00015-X.

Toivio, J., Helmisaari, H.S., Palviainen, M., Lindeman, H., Ala-Ilomäki, J., Sirén, M., Uusitalo, J., 2017. Impacts of timber forwarding on physical properties of forest soils in southern Finland. For. Ecol. Manag. 405, 22–30. https://doi.org/10.1016/j.foreco.2017.09.022.

Tsintides, T.C., Hadjikyriakou, G.N., Christodoulou, C.S., Kourtellarides, L., Cristodoulou, Ch s., Hadjikyriakou, G.N., Fragman, O., Levy-Yamamori, R., 2002. *Trees and Shrubs in Cyprus* (First). Leventis Foundation &Cyprus Forest Association.

Tucker, C.J., Elgin, J.H., McMurtrey, J.E., Fan, C.J., 1979. Monitoring corn and soybean crop development with hand-held radiometer spectral data. Remote Sens. Environ. 8 (3), 237–248. https://doi.org/10.1016/0034-4257(79)90004-X.

Turkowski, K., 1990. Filter for common resampling tasks. In: Glassner, A.S. (Ed.), Graphics Gems. Academic Press, pp. 147–165.

United Nations General Assembly, 2015. Transforming Our World: the 2030 Agenda for Sustainable Development.

United Nations Office for the Coordination of Humanitarian Affairs (OCHA), 2024. Cyprus roads. Humanitarian Data Exchange HDX. https://data.humdata.org/dataset/hotosm_cyp_roads.

United States Environmental Protection Agency (EPA), 2024. Greenhouse gas equivalencies calculator. https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator#results.

Vepakomma, U., Kneeshaw, D.D., De Grandpré, L., 2018. Influence of natural and anthropogenic linear canopy openings on forest structural patterns investigated using LiDAR. Forests 9 (9). https://doi.org/10.3390/f9090540.

Vescovo, L., Gianelle, D., 2008. Using the MIR bands in vegetation indices for the estimation of grassland biophysical parameters from satellite remote sensing in the Alps region of Trentino (Italy). Adv. Space Res. 41 (11), 1764–1772. https://doi.org/10.1016/j.asr.2007.07.043.

Vogiatzakis, I., Balzan, M., Drakou, E., Katsanevakis, S., Padoa-Schioppa, E., Tzirkalli, E., Zotos, S., Álvarez, X., Külvik, M., Fonseca, C., Moustakas, A., Martínez-López, J., Mackelworth, P., Mandzukovski, D., Ricci, L., Srdjevic, B., Tase, M., Terkenli, T., Zemah-Shamir, S., et al., 2023. Enhancing small-medium IsLands resilience by securing the sustainability of ecosystem services: the SMILES cost action. Research Ideas and Outcomes 9. https://doi.org/10.3897/rio.9.e116061.

Wang, L., Qu, J.J., 2007. NMDI: a normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. Geophys. Res. Lett. 34 (20). https://doi.org/10.1029/2007GL031021.

Wang, X., Ou, T., Zhang, W., Ran, Y., 2022. An overview of vegetation dynamics revealed by remote sensing and its feedback to regional and global climate. Remote Sens. 14 (20), 5275. https://doi.org/10.3390/rs14205275.

Yang, L., Liang, S., Zhang, Y., 2020. A new method for generating a global forest aboveground biomass map from multiple high-level satellite products and ancillary information. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 13, 2587–2597. https://doi.org/10.1109/JSTARS.2020.2987951.

Zenonos, A., Li, S., Brandt, M., Sciare, J., Ciais, P., 2025. AI-powered estimation of tree covered area and number of trees over the Mediterranean island of Cyprus. Front. Remote Sens. 6. https://doi.org/10.3389/frsen.2025.1498217.

Zhang, Y., Ma, J., Liang, S., Li, X., Liu, J., 2022. A stacking ensemble algorithm for improving the biases of forest aboveground biomass estimations from multiple remotely sensed datasets. GIScience Remote Sens. 59 (1), 234–249. https://doi.org/10.1080/15481603.2021.2023842.

Zhao, J., Liu, D., Zhu, Y., Peng, H., Xie, H., 2022. A review of forest carbon cycle models on spatiotemporal scales. J. Clean. Prod. 339, 130692. https://doi.org/10.1016/j.jclepro.2022.130692.

Zhou, T., Luo, X., Hou, Y., Xiang, Y., Peng, S., 2020. Quantifying the effects of road width on roadside vegetation and soil conditions in forests. Landsc. Ecol. 35 (1), 69–81. https://doi.org/10.1007/s10980-019-00930-8.

Zomeni, M., Vogiatzakis, I.N., 2014. Roads and Roadless areas in Cyprus: implications for the natura 2000 network. J. Landscape Ecol.(Czech Republic) 7 (1), 75–90. https://doi.org/10.2478/jlecol-2014-0010.

Zurqani, H.A., 2025. A multi-source approach combining GEDI LiDAR, satellite data, and machine learning algorithms for estimating forest aboveground biomass on Google Earth Engine platform. Ecol. Inform. 86. https://doi.org/10.1016/j.ecoinf.2025.103052.