

Prevailing European light regimes are mismatched with conditions for high plant diversity

Wanben Wu

wanben.wu@bio.au.dk

Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), Aarhus University

<https://orcid.org/0000-0002-0581-9774>

Jens-Christian Svenning

Aarhus University <https://orcid.org/0000-0002-3415-0862>

Sean E. H. Pang

Oxford University <https://orcid.org/0000-0003-4574-6028>

Jonathan Lenoir

CNRS <https://orcid.org/0000-0003-0638-9582>

Jürgen Dengler

Vegetation Ecology Research Group, Institute of Natural Resource Management (IUNR), Zurich University of Applied Sciences (ZHAW) <https://orcid.org/0000-0003-3221-660X>

Jesper Moeslund

Aarhus University, Department of Ecoscience <https://orcid.org/0000-0001-8591-7149>

Zvezdana Stančić

Faculty of Geotechnical Engineering, University of Zagreb <https://orcid.org/0000-0002-6124-811X>

Grzegorz Swacha

Botanical Garden, University of Wrocław <https://orcid.org/0000-0002-6380-2954>

Koenraad Van Meerbeek

KU Leuven <https://orcid.org/0000-0002-9260-3815>

Idoia Biurrun

Department of Plant Biology and Ecology, University of the Basque Country UPV/EHU

<https://orcid.org/0000-0002-1454-0433>

Behlül Güler

Dokuz Eylül University <https://orcid.org/0000-0003-2638-4340>

Emmanuel Garbolino

Institut Supérieur d'Ingénierie et de Gestion de l'Environnement (ISIGE), Mines Paris PSL

<https://orcid.org/0000-0002-4954-6069>

Gianmaria Bonari

University of Siena <https://orcid.org/0000-0002-5574-6067>

Giuseppe Fenu

University of Cagliari <https://orcid.org/0000-0003-4762-5043>

Hans Henrik Bruun

University of Copenhagen <https://orcid.org/0000-0003-0674-2577>

Jiri Dolezal

Botanický ústav AV ČR, v. v. i. <https://orcid.org/0000-0002-5829-4051>

Jonathan Millett

Loughborough University

Josep Peñuelas

CREAF-CSIC <https://orcid.org/0000-0002-7215-0150>

Klaus Ecker

WSL Swiss Federal Research Institute <https://orcid.org/0000-0003-3388-1353>

Tetiana Dziuba

Department of Geobotany and Ecology, M.G. Kholodny Institute of Botany, National Academy of Sciences of Ukraine <https://orcid.org/0000-0001-8621-0890>

Borja Jiménez-Alfaro

University of Oviedo <https://orcid.org/0000-0001-6601-9597>

Ciara Dwyer

Center for Environmental and Climate Science, Lund University, Sweden, 223 62

Fabio Attorre

Sapienza University of Rome <https://orcid.org/0000-0002-7744-2195>

Francesca Napoleone

Department of Environmental Biology, Sapienza University of Rome <https://orcid.org/0000-0002-3807-7180>

Jozef Šibík

Plant Science and Biodiversity Center, Slovak Academy of Sciences, Institute of Botany
<https://orcid.org/0000-0002-5949-862X>

Renata Custerevska

University of Ss. Cyril and Methodius

Romà Ogaya

CREAF

Svetlana Aćić

University of Belgrade, Faculty of Agriculture, Department of Botany, Nemanjina 6, Belgrade-Zemun, Serbia <https://orcid.org/0000-0001-6553-3797>

Tomáš Černý

Czech University of Life Sciences, Faculty of Forestry and Wood Science, Department of Forest Ecology, Kamýcká 1176

Michele De Sanctis

Department of Environmental Biology, University Sapienza of Rome <https://orcid.org/0000-0002-7280-6199>

Ute Jandt

Martin Luther University Halle-Wittenberg <https://orcid.org/0000-0002-3177-3669>

Wolfgang Schmidt

University of Göttingen <https://orcid.org/0000-0001-5356-4684>

Željko Škvorc

University of Zagreb <https://orcid.org/0000-0002-2848-1454>

Robert Buitenwerf

Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University

<https://orcid.org/0000-0003-3356-2301>

Article

Keywords:

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10527080/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Prevailing European light regimes are mismatched with conditions for high plant diversity

Authors

Wanben Wu^{1*}, Jens-Christian Svenning¹, Sean E. H. Pang^{1,2}, Jonathan Lenoir³, Jürgen Dengler^{4,5}, Jesper E. Moeslund⁶, Zvezdana Stančić⁷, Grzegorz Swacha⁸, Koenraad Van Meerbeek^{9,10}, Idoia Biurrun¹¹, Behlül Güler¹², Emmanuel Garbolino¹³, Gianmaria Bonari¹⁴, Giuseppe Fenu¹⁵, Hans H. Bruun¹⁶, Jiri Dolezal^{17,18}, Jonathan Millett¹⁹, Josep Penuelas^{20,21}, Klaus Ecker²², Tetiana Dziuba²³, Borja Jiménez-Alfaro^{24,25}, Ciara Dwyer^{19,26}, Fabio Attorre²⁷, Francesca Napoleone²⁷, Jozef Šibík²⁸, Renata Čušterevska²⁹, Romà Ogaya²⁰, Svetlana Aćić³⁰, Tomáš Černý³¹, Michele De Sanctis²⁷, Ute Jandt^{32,33}, Wolfgang Schmidt³⁴, Željko Škvorc³⁵, Robert Buitenwerf¹

Affiliations

¹ Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), Department of Biology, Aarhus University, Aarhus, Denmark.

² Department of Biology, University of Oxford, Oxford, UK.

³ Ecologie et Dynamique des Systèmes Anthropisés (EDYSAN), Université de Picardie Jules Verne, Amiens, France.

⁴ Vegetation Ecology Research Group, Institute of Natural Resource Sciences (IUNR), Zurich University of Applied Sciences (ZHAW), Wädenswil, Switzerland.

⁵ Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Bayreuth, Germany.

⁶ Department of Ecoscience, Aarhus University, Aarhus, Denmark.

⁷ Faculty of Geotechnical Engineering, University of Zagreb, Croatia.

⁸ Botanical Garden, University of Wrocław, Wrocław, Poland.

⁹ Department of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium.

¹⁰ KU Leuven Plant Institute, KU Leuven, Leuven, Belgium.

¹¹ University of the Basque Country (UPV/EHU), Spain.

¹² Biology Education, Dokuz Eylül University, Izmir, Türkiye.

¹³ Institut Supérieur d'Ingénierie et de Gestion de l'Environnement (ISIGE), Mines Paris PSL, Fontainebleau, France.

¹⁴ Department of Life Sciences, University of Siena, Siena, Italy.

¹⁵ Department of Life and Environmental Sciences, University of Cagliari, Cagliari, Italy.

- 32 ¹⁶ Department of Biology, University of Copenhagen, Copenhagen, Denmark.
- 33 ¹⁷ Institute of Botany of the Czech Academy of Sciences, Pruhonice, Czechia.
- 34 ¹⁸ Department of Botany, Faculty of Science, University of South Bohemia, Ceske Budejovice, Czechia.
- 35 ¹⁹ Geography and Environment, Loughborough University, Loughborough, LE11 3TU, UK.
- 36 ²⁰ CSIC, Global Ecology Unit CREAF-CSIC-UAB, 08193 Bellaterra, Catalonia, Spain.
- 37 ²¹ CREAF, 08193 Cerdanyola del Vallès, Catalonia, Spain.
- 38 ²² Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland.
- 39 ²³ Department of Geobotany and Ecology, M.G. Kholodny Institute of Botany, National Academy of
40 Sciences of Ukraine, Kyiv, Ukraine.
- 41 ²⁴ Biodiversity Research Institute (CSIC/UO/PA), University of Oviedo, Mieres, Spain.
- 42 ²⁵ Department of Biology and Organism Systems, University of Oviedo, Oviedo, Spain.
- 43 ²⁶ Department of Earth and Environmental Sciences, Lund University, Lund, Sweden.
- 44 ²⁷ Department of Environmental Biology, Sapienza University of Rome, 00185 Rome, Italy.
- 45 ²⁸ Plant Science & Biodiversity Center SAS, Bratislava, Slovakia.
- 46 ²⁹ Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University, Skopje, North
47 Macedonia.
- 48 ³⁰ Faculty of Agriculture, Department of Botany, University of Belgrade, Belgrade, Serbia.
- 49 ³¹ Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czechia.
- 50 ³² Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg,
51 Halle, Germany.
- 52 ³³ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany.
- 53 ³⁴ Department of Silviculture and Forest Ecology of the Temperate Zones, University of Göttingen,
54 Göttingen, Germany.
- 55 ³⁵ University of Zagreb Faculty of Forestry and Wood Technology, Zagreb, Croatia.
- 56 *Corresponding author. Email: wanben.wu@bio.au.dk

58 **Abstract**

59 Light is a key driver of plant performance and community assembly, yet whether contemporary
60 landscapes provide the light regimes associated with high diversity remains unclear. Here, we integrate
61 more than 650,000 vegetation plots, 14,835 species-specific light indicator values from Ecological
62 Indicator Values for Europe, and Earth observation data to assess relationships between near-ground
63 light regimes and plant diversity across Europe, using separate plot subsets for light modelling and
64 biodiversity analyses. Full models incorporating light regimes increase cross-validated R^2 by 27.4–34.5%
65 at the landscape scale, relative to models containing climatic, edaphic and sampling-effort covariates
66 alone. Our results reveal a pronounced landscape-scale mismatch between prevailing light regimes and
67 those associated with high predicted plant diversity in contemporary Europe. Across three Hill numbers,
68 light-regime conditions associated with the top 10% of predicted plant diversity occurred across only
69 5.4–8.3% of European natural and semi-natural areas, indicating that these high-diversity light
70 conditions are uncommon in contemporary landscape, whereas prevailing regimes were generally
71 darker and more homogeneous. These findings call for greater attention to light regimes in efforts to
72 restore plant diversity across Europe, with a focus on restoring relatively open vegetation and spatially
73 heterogeneous near-ground light environments.

INTRODUCTION

Light regulates ecosystem function and plant growth, influencing primary productivity¹, water cycling², nutrient turnover³, and plant responses across ecosystems⁴. Yet its role in biodiversity research has often been reduced to a one-dimensional gradient of mean light availability. Real light regimes near the ground are more complex, varying not only in mean intensity but also in spatial heterogeneity. These two dimensions may shape plant diversity through distinct but complementary mechanisms: greater light availability can increase resource supply^{5,6} and population sizes⁷, whereas higher spatial heterogeneity can expand niche space and promote local coexistence among species with contrasting light requirements⁸. Together, light availability and spatial heterogeneity can define a light regime that organizes plant diversity across scales, but this hypothesis remains untested over large areas. Understanding these relationships also requires determining how widely the light regimes associated with high diversity are represented in contemporary landscapes.

Empirical evidence for light-diversity relationships remains inconsistent. Previous studies have reported positive^{9,10}, neutral¹¹, and negative^{12,13} associations between light availability and plant diversity. Evidence for the effects of spatial heterogeneity in light availability on plant diversity is similarly mixed and largely restricted to studies of forest understories^{4,9}. These contrasting findings may partly arise because mean light availability and spatial heterogeneity are often analysed separately, despite their likely interactive effects. Such inconsistencies among previous findings may also reflect the limited spatial grains and extents covered by most studies, where microclimate, vegetation structure, soil conditions, disturbance history and topography can strongly mediate plant species distribution and thus diversity patterns¹⁴. Local-scale observations may therefore fail to reveal broader ecological and biogeographic relationships between light regimes and plant diversity. We therefore ask whether consistent relationships between light regimes and plant diversity emerge across environmental gradients at a continental extent.

Light-diversity relationships may be expected to reflect evolutionary context. For example, more than 80% of European forest plant species are associated with relatively high light conditions¹⁵, suggesting that much of the contemporary forest flora retains affinities with semi-open habitats and heterogeneous light regimes. Paleo-ecological evidence further indicates that open vegetation and light woodlands were widespread across parts of Europe before intensive human land use and the late-Quaternary decline of large herbivores¹⁶. Since the early Holocene, changes in herbivore populations^{17,18} and the expansion of woody vegetation have altered canopy structure across Europe. More recently, plantation forestry¹⁹, land abandonment²⁰ and shifts in disturbance and management regimes have further accelerated woody vegetation expansion and canopy closure across Europe. These changes are expected to reduce light availability and alter spatial heterogeneity, potentially reshaping the light regimes experienced by plant communities across Europe. Woody vegetation expansion and canopy closure may reduce the prevalence of relatively open and spatially heterogeneous

light conditions. If these conditions support high plant diversity, such changes could create a mismatch between prevailing European light regimes and those associated with diverse plant communities.

A major obstacle to understanding light–diversity relationships at broad spatial extents has been the difficulty of quantifying light regimes consistently across large regions and long time periods. Field-based approaches, including hemispherical photography⁵ and radiation sensors²¹, provide accurate local estimates but are limited in spatial and temporal coverage. Remote sensing has expanded the capacity to characterize vegetation structure and infer canopy light conditions^{22,23}, but available data often represent contemporary snapshots that may not capture the long-term light environments experienced by plant communities. However, the long-term environmental conditions at a site can be inferred from the composition of the plant communities²⁴. Because local communities reflect long-term filtering, species-specific light preferences can be used to infer biologically meaningful light regimes integrated over ecological timescales.

Here, we combine species-specific light indicator values from Ecological Indicator Values for Europe (EIVE)²⁴ with data from more than 650,000 vegetation plots in the European Vegetation Archive (EVA)²⁵ and multi-source Earth observations to assess light regimes across Europe. This allows us to quantify spatial variation in mean light availability and light heterogeneity, and to test how their individual and interactive effects structure plant diversity from local to landscape scales. We tested three predictions: (1) light-regime predictors improve plant diversity models beyond macroclimatic and edaphic variables; and (2) mean light availability and spatial heterogeneity interact to shape plant diversity more strongly than either dimension alone. (3) there is a mismatch between contemporary European light regimes and those associated with high plant diversity.

RESULTS

Light availability model performance and feature importance

To map the geographic variation of light availability, we trained random-forest models using 328 layers of spectral, SAR backscatter and topographic predictors (*Supplementary Materials Table S1*). Models were applied to all natural and semi-natural areas, as defined by the ESA WorldCover 10 m 2020 layer²⁶, after excluding cropland (code 40), built-up areas (code 50), bare land (code 60), snow and ice (code 70), and permanent water bodies (code 80). The final model output represents mean EIVE for light availability.

Across the eight European biogeographic regions, model R^2 ranged from 0.56 to 0.73, and RMSE ranged from 0.32 to 0.80 (*Fig. 1a, b*). Across Europe, an 80/20 train-test split gave a similar overall performance, with an R^2 of 0.69 (*Fig. 1d*). Tenfold cross-validation yielded a mean Pearson's r of 0.80, an R^2 of 0.64, an RMSE of 0.60 and an MAE of 0.42 (*Fig. 1e*). Feature-importance analysis identified Sentinel-1 VH backscatter (10.7%), kernel Normalized Difference Vegetation Index (kNDVI; 10.5%) and ALOS PALSAR HV backscatter (7.5%) as the strongest predictors of light availability (*Fig. 1c*).

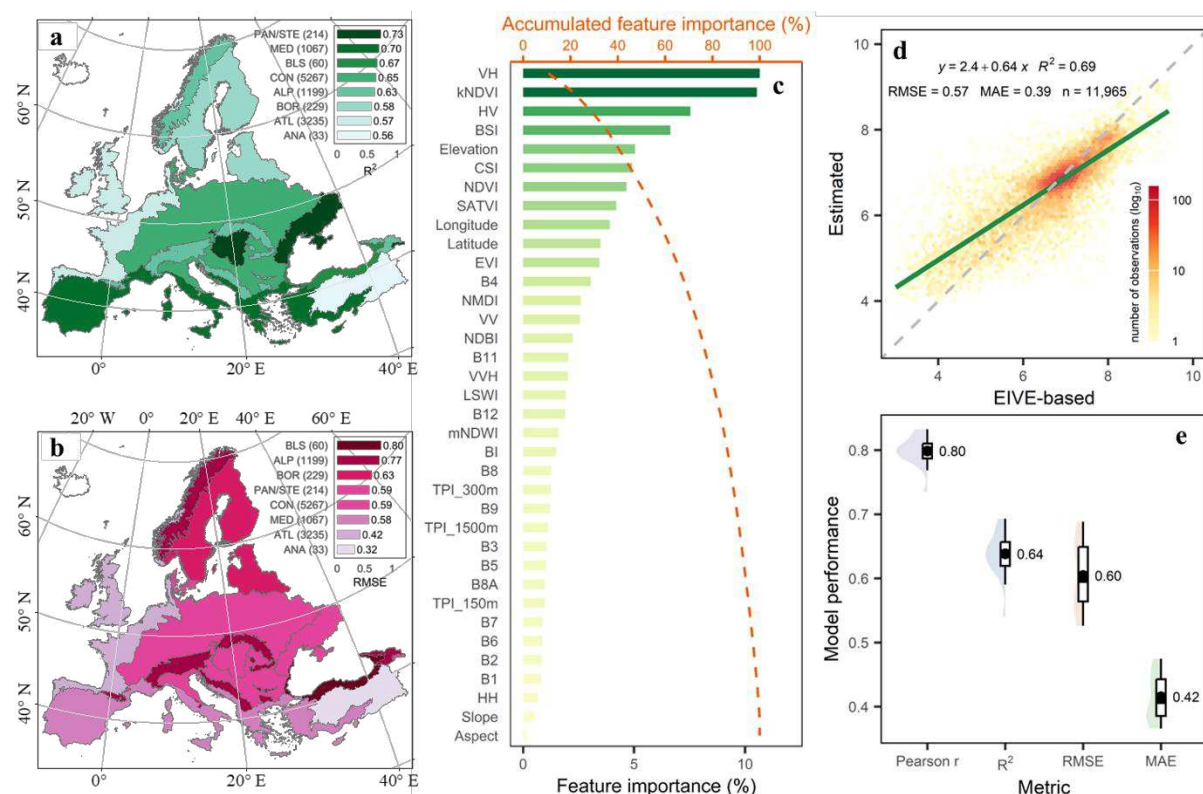


Fig.1 Validation and feature importance analysis of the continental light-availability model.

a,b, Regional model performance measured by R^2 (a) and RMSE (b) across eight European biogeographic regions. **c**, Relative importance of predictor groups derived from multi-source Earth observation data and used to reconstruct biologically inferred light availability regimes across Europe. Values represent the summed contribution of all spatial and temporal metrics within each predictor group. **d**, Observed versus predicted light availability for an independent 20% holdout dataset ($R^2 = 0.69$). **e**, Distributions of Pearson's r , R^2 , RMSE and MAE across cross-validation folds.

Geospatial patterns of light regimes across Europe

We produced a pan-European map of the first pan-European map of ecologically inferred light availability near the ground at 10-m resolution (Fig.2). Light availability follows a clear latitudinal pattern, peaking in southern and northern Europe and reaching its lowest values in the European mid-latitudinal belt (~40 - 55°) (Fig.2a–b). Substantial variation exists among regions, with mean light availability values declining from the Anatolian region (mean light availability = 7.7) to the Continental region (mean light availability = 6.16) (Supplementary Materials Fig.S1).

At the local scale, light heterogeneity (Fig.2c–d), quantified as the standard deviation of light availability within a 50-m moving window, exhibits an opposing latitudinal gradient relative to mean light availability, with lower values in southern and northern Europe. Across biogeographic regions, the Black Sea region shows the highest light heterogeneity (0.23), whereas the Anatolian (0.09) and Arctic (0.06) regions exhibit the lowest levels.

Four example locations (*I–IV* in Fig.2) highlight the ecological detail captured by the European map. Examples range from the mosaic landscapes of the Mols rewilding site in Denmark (*I*), characterized by relatively high light availability and heterogeneity, to forests heavily affected by bark beetle outbreaks in the Harz Mountains, Germany (*II*), where light availability is very high but its spatial heterogeneity remains low due to severe disturbance. In contrast, the near-primary forests of the Carpathian Mountains in western Ukraine (*III*) exhibit very low light availability and low spatial heterogeneity in light conditions, while intensively managed forest landscapes in France (*IV*) display a highly heterogeneous light environment with important local contrasts between low and high light availability. Together, these examples demonstrate that our light regime mapping approach effectively captures ecologically meaningful variation across contrasting vegetation types.

At the landscape scale (10 km grid cells), the spatial distributions of mean light availability and light heterogeneity near the ground closely mirror local-scale patterns, indicating that large-scale patterns of European light regimes are similar at 10-m and 10-km resolutions (*Supplementary Materials Fig.S2a–b*). When projected into climatic space, light availability is higher in warmer and drier climates, whereas the spatial heterogeneity of light conditions is more pronounced in cooler and wetter regions (*Supplementary Materials Fig.S2c–d*). Across landscapes, light availability and its local spatial heterogeneity exhibit a pronounced nonlinear relationship (*Supplementary Materials Fig.S2e*), with heterogeneity peaking at intermediate-to-high levels.

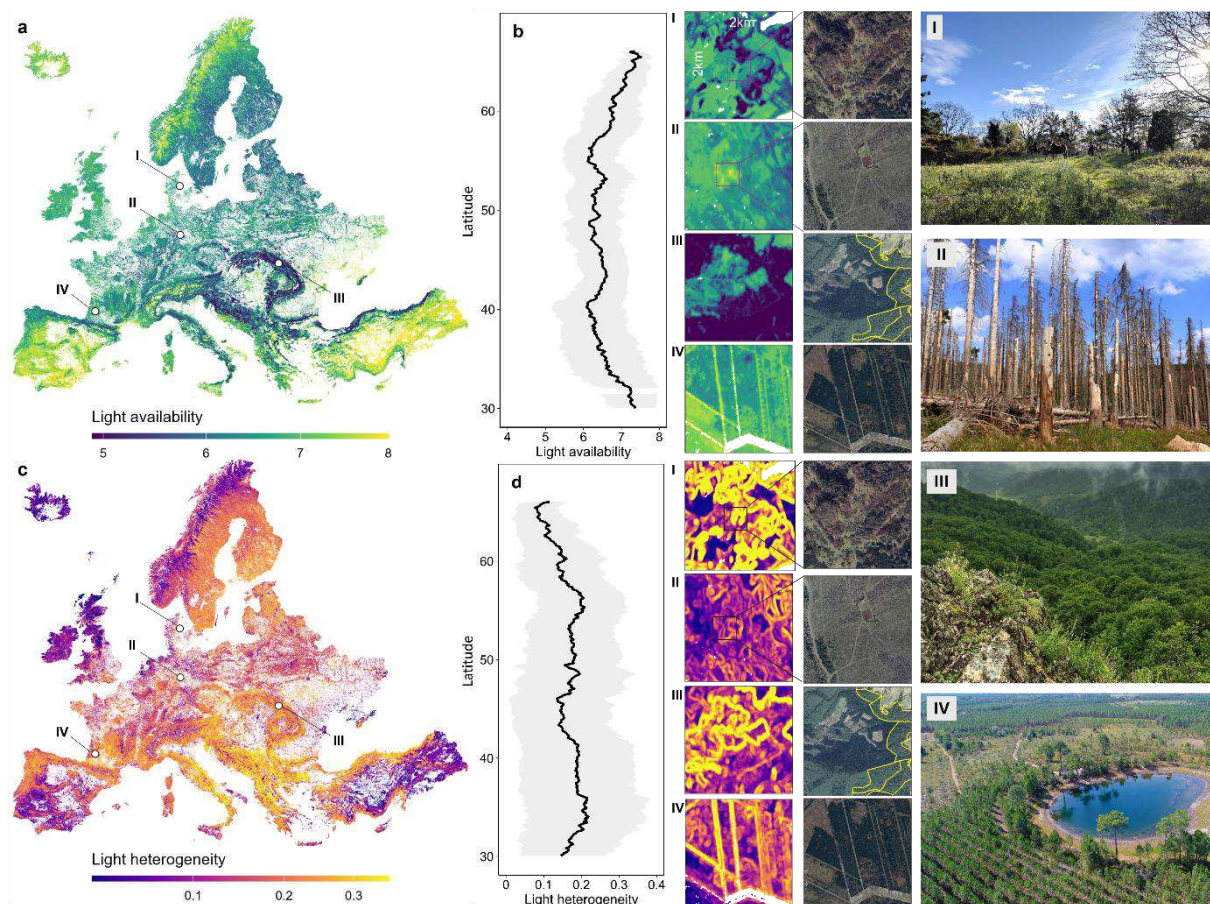


Fig.2 European light-regime maps near the ground.

a, Spatial distribution of light availability near the ground across natural and semi-natural areas in Europe. **c**, Local-scale light heterogeneity near the ground across natural and semi-natural areas, quantified using a 50 m moving window. Natural and semi-natural areas were defined based on the ESA WorldCover 10 m 2020 layer²⁶, excluding cropland (code 40), built-up areas (code 50), bare land (code 60), snow and ice (code 70), and permanent water bodies (code 80). **b**, **d**, Corresponding latitudinal trends in the spatial distribution of mean light availability and spatial heterogeneity, aggregated within 0.1° latitude bands; shaded envelopes indicate ± 1 SD around the mean. Maps are displayed at 100 m resolution for visualization, with insets I–IV showing native 10 m resolution subsets. Insets illustrate contrasting ecological contexts: **I**: the mosaic landscapes of the Mols rewilding site in Denmark; **II**: forests heavily disturbed by bark beetle outbreaks in the Harz Mountains, Germany; **III**: near-primary forests in the Carpathian Mountains, Ukraine; and **IV**: intensively managed plantation forests in France. The last two columns show high-resolution true-color imagery provided by Google Maps, along with corresponding landscape photographs. The full 10 m resolution map is available via an interactive online platform (<https://ee-wwanben1994.projects.earthengine.app/view/eulr-10m-explorer>).

Habitat-specific distribution of light regimes

To evaluate the ecological realism of the modeled light regime, we overlaid the light regime maps with EUNIS habitat types²⁷ and quantified habitat-specific distributions of both mean light availability and light heterogeneity (Fig.3). Clear and systematic differences emerged among habitat types.

Forest habitats (*T1* - *T4*) were predominantly associated with lower mean light availability, with median values ranging from 5.88 to 6.71. The lowest light levels were observed in broadleaved deciduous forests (*T1*), with *Alnus cordata* forests (*T1G*) exhibiting the lowest light availability (EIVE-L = 5.44) (Supplementary Materials Fig.S3), consistent with the closed-canopy conditions characteristic of mature stands. In contrast, the linear tree formations, small anthropogenic forests, recently felled areas, early successional stands, and coppice systems (*T4*) exhibited substantially higher light availability. Recently felled area (*T43*) exhibited the highest light availability (EIVE-L = 6.72).

Grassland and tall forbs from open to semi-open habitats (*R1* - *R6*) exhibited consistently higher light availability (EIVE-L = 6.23 - 6.80), reflecting more open vegetation states. While some overlap in light availability was observed among habitat types, clear shifts in both mean light availability and distribution density suggest that our mean light availability metric effectively captures broad structural differences across European ecosystems.

In contrast, the spatial heterogeneity of light conditions near the ground varied over a narrower numerical range across habitat types but exhibited coherent directional patterns. Forest habitats, particularly types *T1* - *T3*, tended to show lower light heterogeneity (median values ≈ 0.20 - 0.22), consistent with relatively uniform canopy structures. Higher light heterogeneity was more frequently observed in structurally complex

or disturbance-associated habitats, including several riparian and semi-open types (R1 - R5; median values $\approx 0.23 - 0.27$), reflecting greater spatial variation in canopy openness.

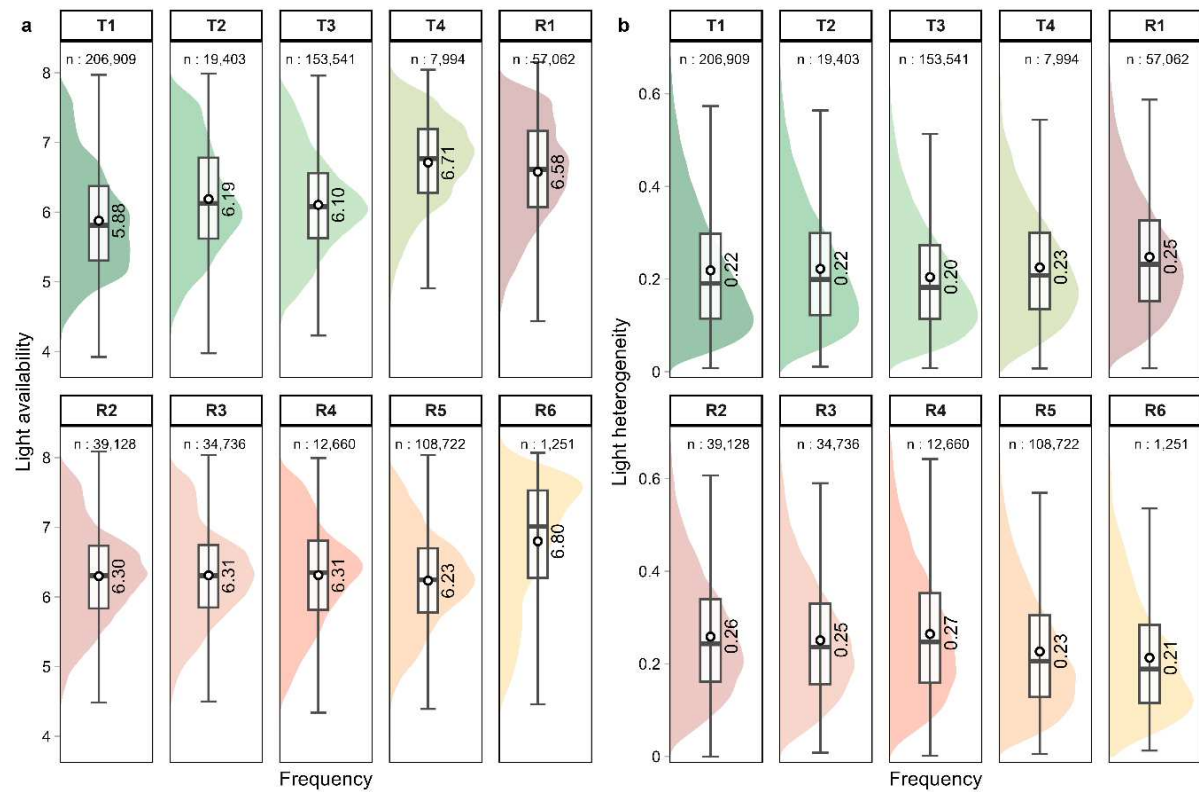


Fig.3 Distributions of modeled light regimes across EUNIS habitat types.

a, Light availability, and **b**, light heterogeneity (standard deviation within a 50-m-radius window), across forest (T1–T4) and grassland/tall forb (R1–R6) habitats. Half violin plots show kernel density distributions (frequency) of pixel-level values; embedded boxplots indicate medians (bold line), interquartile ranges (box), and full ranges (whiskers); white dots denote means. Sample sizes (number of 10 m pixels) are shown above each panel. Forest types generally exhibit lower light availability and heterogeneity, while grassland and structurally complex habitats show higher values for both metrics. Habitat codes are as follows: T1, broadleaved deciduous forests; T2: Broadleaved evergreen forests; T3: Coniferous forests; T4: Lines of trees, small anthropogenic forests, recently felled forest, early-stage forest and coppice. R1: Dry grasslands; R2: Mesic grasslands; R3: Seasonally wet and wet grasslands; R4: Alpine and subalpine grasslands; R5: Woodland fringes and clearings and tall-forb stand; R6: Inland salt steppes. Supplementary Materials Fig.S3 shows the light regime patterns across EUNIS level 3 forest (T) and grassland (R) habitat types.

Light regime as predictor of plant diversity

To quantify the contribution of light regime to plant-diversity predictions, we fitted three hierarchical spatial and spatiotemporal generalized linear mixed models (GLMMs) using the package sdmTMB²⁸ in R (version 4.01): (1) an environment-only model including macroclimate, soil properties and plot area; (2) a light-added model further incorporating mean light availability and its spatial heterogeneity; and (3) a light interaction model, including the interaction between both light variables. Model performance differed markedly between spatial scales (Table 1). At the local scale, models excluding and including light explained relatively

little variation in diversity ($R^2 = 0.134 - 0.138$), whereas explanatory power was substantially higher at the landscape scale ($R^2 = 0.307 - 0.483$). Incorporating light-regime variables improved model performance at both scales, but the effect was considerably stronger at the landscape scale. Relative to the environment-only model, improvements at the local scale were minimal, with explained variance increasing by only 3% ($\Delta R^2 = 0.004$). In contrast, relative to the environment-only model, the interaction model increased R^2 by 0.104 for N_0 , 0.107 for N_1 and 0.106 for N_2 at the landscape scale, corresponding to relative improvements of 27.4%, 33.2% and 34.5%, respectively. Most of these gains were already achieved by the additive light model; the interaction formulation provided only small further improvements. Together, these results identify light regimes near the ground as a major determinant of broad-scale diversity patterns while suggesting a more limited role in explaining local-scale variation. Tenfold spatial cross-validation further demonstrated strong agreement between observed and predicted diversity, particularly at the landscape scale (*Supplementary Materials Fig.S5*). Moran's I tests indicated little residual spatial autocorrelation in the final interaction models (Moran's I < 0.05, P > 0.05; *Supplementary Materials Table S2*), suggesting that the spatial random field adequately captured spatial dependence.

Table 1 | Cross-validated predictive performance of alternative light-regime models across spatial scales.

Values represent the mean $\pm$ standard error (s.e.) across five spatial cross-validation folds. The Environment-only model included climatic, edaphic and plot-level covariates but excluded light-regime predictors; the light-added model included additive effects of the two light regime predictors only; the light-interaction model included additive effects of the two light predictors and their interactions. Model performance was evaluated using the coefficient of determination (R^2) and root mean square error (RMSE). Bold values indicate the best-performing model for each diversity metric and spatial scale.

Scale	Response variable	Model	R^2	RMSE
Landscape	N_0	Environment-only model	0.379 ± 0.025	0.715 ± 0.013
		Light-added model	0.477 ± 0.018	0.656 ± 0.014
		Light interaction model	0.483 ± 0.017	0.653 ± 0.013
	N_1	Environment-only model	0.322 ± 0.023	0.623 ± 0.012
		Light-added model	0.427 ± 0.014	0.573 ± 0.010
		Light interaction model	0.429 ± 0.014	0.572 ± 0.009

		<i>Environment-only model</i>	0.307 ± 0.018	0.579 ± 0.007
	N_2	<i>Light-added model</i>	0.409 ± 0.014	0.535 ± 0.008
		<i>Light interaction model</i>	0.413 ± 0.014	0.534 ± 0.008
Local	N_0	<i>Environment-only model</i>	0.134 ± 0.007	0.600 ± 0.003
		<i>Light-added model</i>	0.137 ± 0.005	0.600 ± 0.004
		<i>Light interaction model</i>	0.138 ± 0.005	0.599 ± 0.004

To assess the relative importance of individual predictors, we quantified the increase in prediction error following permutation of each predictor variable in the light-added models under blocked spatial cross-validation (see methods). Predictor importance differed markedly between spatial scales (*Fig.4*). At the local scale, plot area dominated model performance, accounting for 69.2% of total permutation importance ($\Delta\text{RMSE} = 0.54$), followed by annual precipitation (17.1%; $\Delta\text{RMSE} = 0.13$). Mean light availability and light heterogeneity contributed comparatively little, accounting for 2.7% ($\Delta\text{RMSE} = 0.02$) and 3.5% ($\Delta\text{RMSE} = 0.03$) of total importance, respectively. In contrast, at the landscape scale, light heterogeneity emerged as the most influential predictor, accounting for 36.8% of total importance for species richness (N_0), 75.6% for Shannon diversity (N_1), and 74.3% for Simpson diversity (N_2). These results reveal a pronounced shift in the drivers of plant diversity across spatial scales, from local dominance of sampling area and climatic conditions to a strong influence of light heterogeneity at the landscape scale.

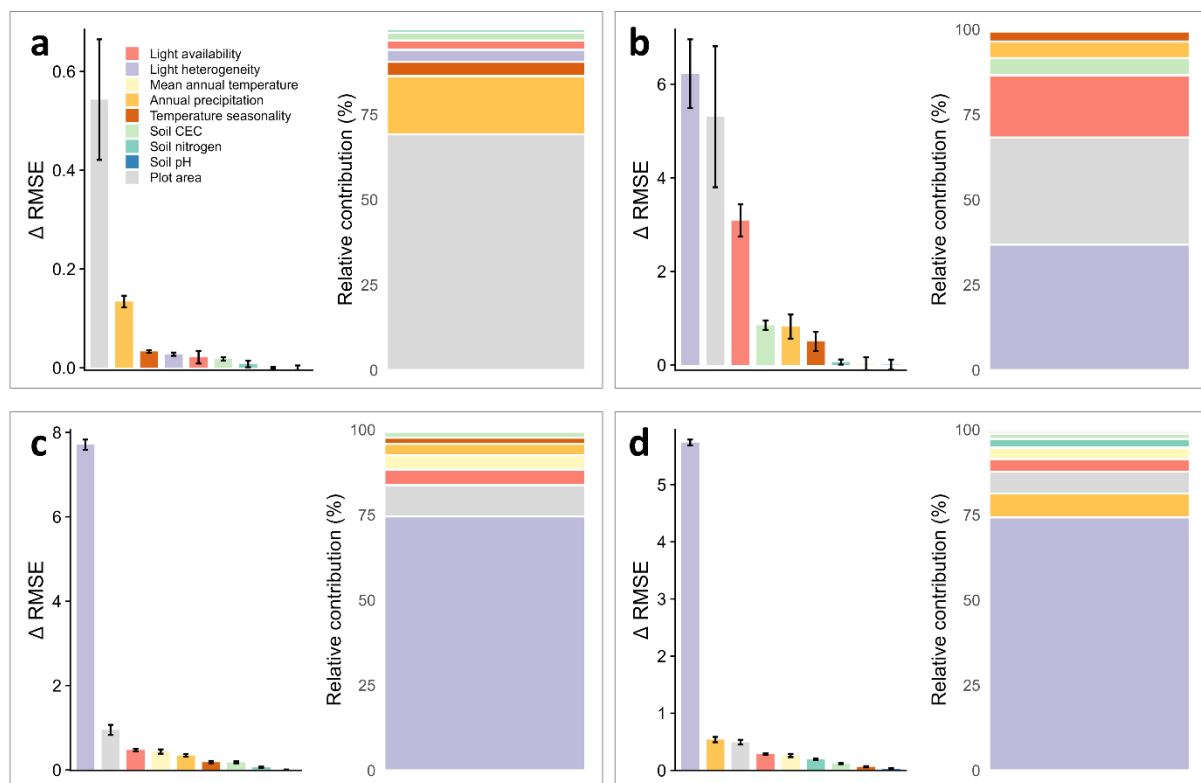


Fig.4 Relative contributions of light-regime, climatic and edaphic predictors to plant diversity across spatial scales.

Panels show results for species richness (N_0) at the local scale (**a**) and for species richness (N_0 ; **b**), Shannon diversity (N_1 ; **c**) and Simpson diversity (N_2 ; **d**) at the landscape scale (10×10 km). Light-regime predictors include mean light availability and light heterogeneity. Environmental predictors comprise mean annual temperature, annual precipitation, temperature seasonality, soil cation-exchange capacity (CEC), soil nitrogen, soil pH and sampling area (plot area).

Plant diversity across the light regime space

Using the interaction models, we identified combinations of mean light availability and spatial heterogeneity associated with high predicted plant diversity across Europe (Fig.5). Plant diversity exhibited a structured and nonlinear distribution across the joint gradients of light availability and heterogeneity, with the highest values consistently concentrated in relatively non-shaded and heterogeneous environments. However, the shape and extent of this high-diversity space differed markedly between spatial scales.

At the local scale, species richness showed only a weak dependence on the joint light regime (Fig.5a). High-diversity plots occurred across a broad range of light conditions, indicating that local richness was not confined to a narrowly defined combination of light availability and local heterogeneity. Nevertheless, the highest richness values were generally associated with moderately high light availability and high spatial heterogeneity in light conditions.

In contrast, landscape-scale diversity displayed a much stronger and more clearly defined association with light regime (Fig.5b - d). Across all diversity metrics, landscapes exceeding the 90th percentile of diversity were primarily concentrated at mean light-availability values of approximately 6.3 - 7.2 and at spatial heterogeneity values above 0.5. The most diverse landscapes (top 5%) occupied an even narrower region centered around light availability values of ~6.7 - 7.0 and spatial heterogeneity values of ~0.7 - 1.0, revealing a well-defined optimum light-regime space for landscape-scale plant diversity.

The extent of this favorable light-regime space differed among diversity metrics. Species richness (N_0) occupied the broadest range of suitable light conditions, whereas Shannon diversity (N_1) and Simpson diversity (N_2) were more tightly clustered around the optimum light-regime window. Thus, spatially heterogeneous light environments supported species accumulation across a relatively wide range of conditions, whereas communities characterized by high evenness and low dominance occurred only under a more restricted combination of mean light availability and spatial heterogeneity. Conditional partial effects supported the interaction analyses (Figs. 5e - h). At the local scale, species richness showed a hump-shaped response to mean light availability, peaking at intermediate-to-high light levels, whereas responses to light heterogeneity were comparatively weak. At the landscape scale, all diversity metrics exhibited unimodal responses to light availability and strong positive responses to light's spatial heterogeneity. Diversity increased rapidly with increasing spatial heterogeneity before approaching saturation, highlighting spatial light heterogeneity as a key determinant of landscape-scale plant diversity.

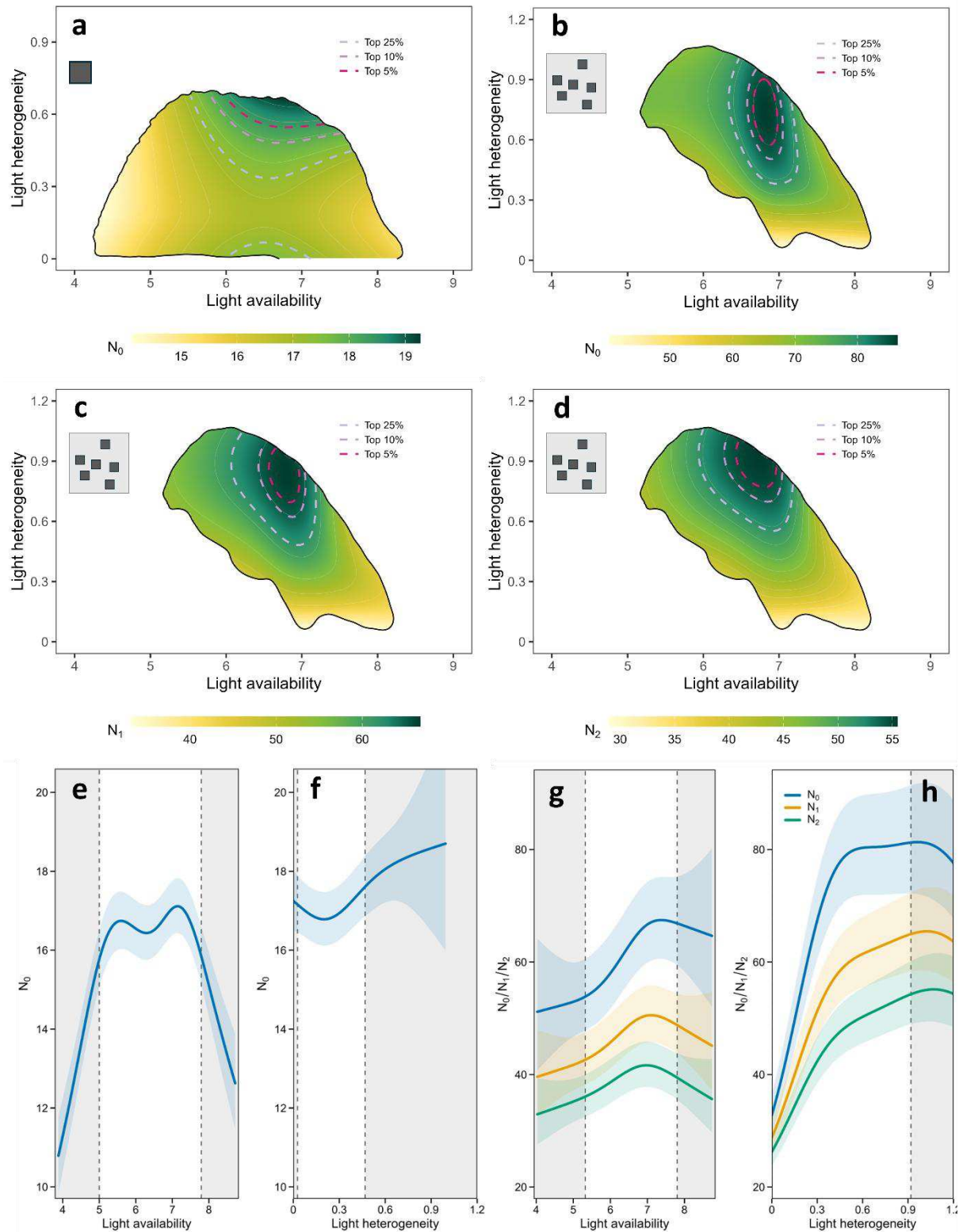


Fig.5 | Predicted plant-diversity responses to light regimes near the ground.

a-d, Conditional effects of mean light availability and spatial light heterogeneity on predicted plant diversity from *sdmTMB* models. Heat maps show local-scale species richness (N_0 ; a), landscape-scale species richness (N_0 ; b), exponential Shannon diversity (N_1 ; c) and inverse Simpson diversity (N_2 ; d). Light heterogeneity was quantified as the standard deviation of light availability within a 50-m-radius window at the local scale and within 10-km grain size at

the landscape scale. Greener colours indicate higher predicted diversity. Coloured dashed contours delineate the upper 5%, 10% and 25% of predicted diversity values. Black solid contours indicate realistic light-regime space, defined by the 99% kernel-density boundary of observed combinations of mean light availability and heterogeneity. **e-f**, Conditional partial effects of local-scale mean light availability (e) and light heterogeneity (f) on plant diversity. **g-h**, Corresponding effects of landscape-scale mean light availability (g) and light heterogeneity (h). Partial-dependence relationships were estimated while holding all other environmental predictors, including climate, soil properties, plot area, spatial covariates and the remaining light-regime variable, at their median values. Shaded bands denote 95% confidence intervals, and grey points show observed data. Hill numbers N_0 , N_1 and N_2 represent species richness, exponential Shannon diversity and inverse Simpson diversity, respectively.

Mismatch with high-diversity light regimes

At the landscape scale, projecting the European light regime into a two-dimensional space defined by mean light availability and spatial heterogeneity in light conditions revealed a pronounced continent-wide mismatch between realized light conditions and those associated with high plant diversity (*Fig. 6a*). Across Europe, plant diversity was maximized under intermediate-to-high light availability combined with high spatial heterogeneity. The light-regime niches associated with the top 10% of plant diversity occupied only 8.3%, 6.3%, and 5.4% of European natural and semi-natural areas for species richness, Shannon diversity, and Simpson diversity, respectively. The progressively smaller niche occupancy from species richness to Simpson diversity suggests that the light environments supporting communities not dominated by a few abundant species are particularly restricted across Europe. Most European landscapes clustered in darker or more homogeneous regions of the light-regime space, consistent with structurally closed and weakly heterogeneous vegetation, indicating a systematic deviation from the light regimes that best support high and evenly distributed plant diversity.

As an exploratory assessment of one potential contributor to this broader pattern, we examined how current management strategies were associated with fine-scale light regimes. Management categories occupied distinct regions of the light-regime space at the 10-m resolution (*Fig. 6b*). Unmanaged forests had the highest mean light availability, whereas forests under “close-to-nature” management had the lowest. These differences may have implications for local plant diversity, although the contribution of light-regime variables to local-scale predictions was modest. These contrasts illustrate that forest management can modify local light regimes and may therefore contribute to broader variation in habitat conditions. However, this forest-focused comparison represents only one possible component of the continent-wide mismatch observed across natural and semi-natural habitats.

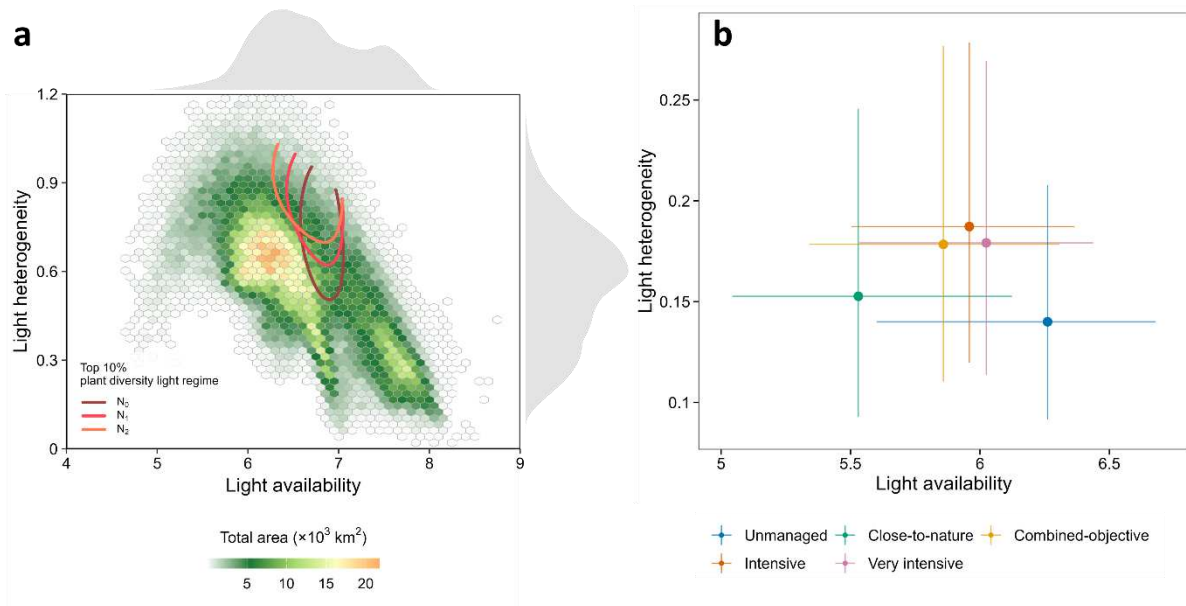


Fig.6 Mismatch between contemporary light regimes and those associated with high plant diversity (a), and differences among forest management types (b).

a, Continental mismatch between realized light regimes and high-diversity light-regime space. The realized space was derived from a 10-m map of biologically inferred light availability aggregated to a 10-km grid across Europe. Coloured hexagons indicate the total area ($\times 10^3 \text{ km}^2$) associated with each combination of mean light availability and spatial heterogeneity. Contours delineate the light-regime space occupied by the top 10% of grid cells for species richness (N_0), exponential Shannon diversity (N_1) and inverse Simpson diversity (N_2). Grey marginal density distributions show the continental distributions of mean light availability (top) and light heterogeneity (right). **b**, Light regimes under different forest-management types. Light availability was estimated at 10-m resolution, and local heterogeneity was calculated within a 50-m-radius moving window and assigned to individual 10-m pixels. Distributions show the light-regime configurations of individual management types and all types combined; points indicate their centroids.

DISCUSSION

At the landscape scale, we find a continent-wide mismatch between prevailing light regimes and those associated with high predicted plant diversity. High predicted diversity was associated with intermediate-to-high light availability and pronounced spatial heterogeneity, yet these combinations were uncommon across contemporary Europe. We defined high-diversity light regimes as combinations of mean light availability and spatial heterogeneity associated with the top 10% of predicted diversity values on the modelled response surfaces. These regimes covered only 8.3%, 6.3% and 5.4% of European natural and semi-natural areas for species richness, exponential Shannon diversity and inverse Simpson diversity, respectively.

One possible explanation for the limited representation of these light regimes is a historical shift towards woody densification²⁹, associated with land abandonment³⁰, and the loss or reduced abundance of large herbivores³¹. Where woody densification reduces both near-ground light availability and its spatial heterogeneity, landscapes may shift away from the light regimes associated with high predicted plant diversity. Consequently, biodiversity-oriented management should consider not only forest extent or canopy cover, but also the spatial configuration of light conditions across landscapes.

The shape of the diversity response surfaces suggests complementary roles for light availability and spatial heterogeneity. Under more open canopy conditions, diversity increased, consistent with the view that reduced light limitation expands the range of viable plant strategies and increases the populations that can be maintained⁹. However, predicted diversity peaked where intermediate-to-high light availability coincided with pronounced spatial heterogeneity, suggesting that mean light availability alone does not account for the observed pattern. Such heterogeneity may broaden the range of light niches, reduce the dominance of any single competitive strategy and allow species with contrasting light requirements to coexist across landscapes³³. This interpretation is consistent with resource-based perspectives linking resource availability to species richness³⁴, and with coexistence theory identifying spatial resource variation as one mechanism that can promote coexistence³⁵.

The evolutionary context of European vegetation also provides a complementary explanation for why high diversity was associated with relatively open and spatially heterogeneous light regimes. A possible historical explanation is that the European species pool reflects, in part, adaptation to mosaics of open habitats, semi-open woodlands and closed-canopy patches maintained by herbivory and other disturbances^{16,36}. Such historical influences could help explain why contemporary diversity is associated with relatively open and heterogeneous light regimes. Our results are therefore compatible with contributions from both contemporary environmental variation and historical influences on the regional species pool.

The scale dependence of the predictive contribution of light regimes highlights the importance of considering near-ground light beyond individual microsites. At the local scale, adding light availability and spatial heterogeneity provided relatively little improvement in predictions beyond macroclimate, soil conditions

and plot area. At the landscape scale, however, including light regimes substantially improved cross-validated predictions of plant diversity. Heterogeneous landscapes may support more species simply because they contain a wider range of habitat types, successional stages and competitive environments. Variation in light conditions among habitat patches may favour species with different light requirements, contributing to compositional differences among communities and greater diversity across the landscape. This interpretation are consistent with heterogeneity–diversity theory^{4,37} and metacommunity perspectives³⁸. Finally, landscapes with heterogenous light regimes may harbor more species, because they better fit the requirements of the species pool by reflecting the prevailing conditions under which it evolved^{39,40}, in line with the species pool hypothesis⁴¹.

The exploratory comparison among forest-management categories is consistent with this interpretation but should not be treated as a causal test. “Close-to-nature” forestry was associated with darker and more homogeneous light conditions (*Fig.6b*), suggesting that this management classification alone does not guarantee the heterogeneous light environments associated with high predicted plant diversity. Similarly, while land abandonment may present important opportunities for nature restoration, it promotes canopy closure and eventually landscape homogenization in the absence of natural disturbance processes from e.g. large herbivores. Depending on local ecological conditions, naturalistic grazing⁴², trophic rewilding⁴³, coppicing⁴⁴ or prescribed burning⁴⁵, may help maintain or restore a mosaic of open, semi-open and habitats thereby sustaining heterogeneous light environments associated with high predicted plant diversity.

By integrating mean light availability and its spatial heterogeneity, the light-regime framework substantially improves landscape-scale predictions of plant diversity beyond climate and soil predictors. This framework provides a basis for investigating how changes in vegetation structure and disturbance regimes may alter light environments relevant to plant diversity. Relating the diversity response surfaces to the continental distribution of light regimes reveals the limited geographic representation of conditions associated with high predicted plant diversity. Maintaining or restoring mosaics of open, semi-open habitats would expand the availability of heterogeneous light environments favourable to diverse plant communities.

CONCLUSION

The integration of European vegetation-plot data with high-resolution Earth observation allowed us to quantify light availability as experienced by plant species near the ground and assess how light regimes shape plant diversity across spatial scales. This integration produced the first continental-scale, 10-m resolution map of light availability, providing a new basis for linking vegetation structure, disturbance and plant biodiversity across Europe. Our analysis identifies mean light availability and spatial heterogeneity as complementary predictors of plant diversity. Plant diversity was maximized under light regimes combining relatively intermediate-to-high light availability with pronounced spatial heterogeneity. This pattern may reflect the combined effects of resource availability, niche differentiation, and historical adaptation to heterogeneous landscapes. Importantly, mosaics of open habitats, semi-open woodlands and closed-canopy patches provide a range of light conditions near the ground. We also observed strong scale dependence, with light-regime variables contributed little additional predictive power at the local scale but substantially improved predictions at the landscape scale. By projecting contemporary landscapes into light-regime space, we further show that many European landscapes are shifted toward darker and more homogeneous conditions than those associated with high plant diversity. Forest management types differed in their associated light regimes. “Close-to-nature” forestry was associated with particularly low mean light availability, whereas more intensive management types showed higher light heterogeneity. Our findings establish the light regime as an underappreciated dimension of biodiversity regulation, and highlight the need to manage semi-natural habitats to promote the heterogeneous light environments that sustain plant diversity, rather than implement wide-scale dense tree cover establishment.

METHODS

The European Vegetation Archive (EVA) data

The raw vegetation plot data were obtained from the European Vegetation Archive (EVA)²⁵ and accessed in April 2025 (Project ID: 240). Each vegetation plot includes complete records of vascular plant species composition, geospatial coordinates, plot size, and the year of observation. The EVA dataset served two main purposes: (1) to provide training and validation plots for light-availability modeling, and (2) to explore the relationship between light availability and plant diversity.

After filtering for observations collected between 2010 and 2024 and excluding plots with a location uncertainty greater than 50 m, a total of 652,028 plots encompassing 12,870 taxa were retained for analysis. These plots were then split into two subsets: (1) Model development. To enhance sample heterogeneity and reduce spatial autocorrelation, we applied a 1 km × 1 km grid filter, retaining only one plot (the most recent observation) per grid cell. When multiple surveys were available for the same plot, only the most recent survey was kept, resulting in an average survey year of 2016. The final modeling dataset comprised 59,733 plots representing 8,658 taxa, which were used to train a random forest (RF) model of light availability. (2) Plant diversity analysis. The remaining 592,295 plots, containing 12,566 taxa, were used to calculate plant diversity metrics and to assess the relationship between modeled light availability and plant diversity. Importantly, plant diversity responses were quantified using plots that were entirely independent from those used to calibrate the light model, thereby preventing circularity and ensuring robust inference

Ecological Indicator Values for Europe (EIVE)

The database of Ecological Indicator Values for Europe (EIVE)²⁴ is currently the most comprehensive ecological indicator value (EIV) system available for European vascular plants. The EIVE database includes harmonized ecological indicator values for 14,835 vascular plant taxa from Europe, covering 31 indicator value systems related to five primary plant niche dimensions: soil moisture, soil nitrogen, soil reaction, light regime, and temperature. These EIVs have been standardized using taxonomic names from the Euro+Med PlantBase⁴⁶. The EIVE light indicator reflects the amount of light availability near the ground after transmission through the canopy, with values ranging from 0 (lowest light availability) to 10 (highest light availability). In this study, we utilized EIVE light indicator values (available for 14,054 taxa) to derive plot-level light regime metrics.

Plot-level light availability values

Plot-scale light regime values were derived from the EIVE database. Species-level aggregates were used when available; otherwise, unweighted means across species within plots were calculated, as cover-weighting has been shown to reduce the predictive performance of mean EIVs⁴⁷.

473 Sentinel-1 backscatter data

474 Sentinel-1 Ground Range Detected scenes were used to estimate light regime. Sentinel-1 provides 10 m
475 spatial resolution imagery with a revisit period of 6–12 days. In this study, we utilized two polarization bands,
476 the backscatter coefficient in VH (vertical-horizontal) and VV (vertical-vertical) polarization, derived from
477 a dual-polarization C-band Synthetic Aperture Radar (SAR) instrument⁵. A total of 169,580 Sentinel-1
478 scenes from January 2019 to December 2021 were processed across Europe. All Sentinel-1 data were
479 acquired and processed using Google Earth Engine (GEE)⁴⁸. The distribution of total Sentinel-1 observations
480 is shown in *Supplementary Materials Fig.S6a*.

481 Sentinel-2 spectral data

482 We used 597,956 Sentinel-2 Level-2A surface-reflectance scenes acquired across Europe between January
483 2019 and December 2021. Clouds and cirrus were masked using the QA60 band and the Sentinel-2 Cloud
484 Probability product. Pixels with cloud probabilities above 25% were excluded. The distribution of total
485 Sentinel-2 observations used in the final analysis is shown in *Supplementary Materials Fig.S6b*.

486 In addition to Sentinel spectral bands, twelve other spectral indices were calculated and used as independent
487 variables, including Normalized Difference Vegetation Index (NDVI)⁴⁹, Enhanced Vegetation Index (EVI)⁵⁰,
488 kernel Normalized Difference Vegetation Index (kNDVI)⁵¹, Land Surface Water Index (LSWI)⁵², Modified
489 Normalized Difference Water Index (MNDWI)⁵³, Green Chlorophyll Vegetation Index (GCVI)⁵⁴,
490 Normalized Multi-band Drought Index (NMDI)⁵⁵, Combinational Shadow Index (CSI)⁵⁶, Brightness Index
491 (BI), Bare Soil Index (BSI)⁵⁷, and Soil-Adjusted Total Vegetation Index (SATVI)⁵⁸. All Sentinel-2 data were
492 acquired and processed using GEE.

$$493 \quad NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \quad (1)$$

$$494 \quad EVI = \frac{2.5(\rho_{NIR} - \rho_{red})}{(\rho_{NIR} + 6\rho_{red} - 7.5\rho_{blue} + 1)} \quad (2)$$

$$495 \quad kNDVI = \tanh\left(\left(\frac{\rho_{NIR} - \rho_{red}}{2\sigma}\right)^2\right) \quad (3)$$

$$496 \quad LSWI = \frac{\rho_{NIR} - \rho_{swir1}}{\rho_{NIR} + \rho_{swir1}} \quad (4)$$

$$497 \quad MNDWI = \frac{\rho_{green} - \rho_{swir1}}{\rho_{green} + \rho_{swir1}} \quad (5)$$

$$498 \quad GCVI = \left(\frac{\rho_{NIR}}{\rho_{green}}\right) - 1 \quad (6)$$

$$499 \quad NMDI = \left(\frac{\rho_{NIR} - (\rho_{swir1} - \rho_{swir2})}{\rho_{NIR} + (\rho_{swir1} - \rho_{swir2})}\right) \quad (7)$$

$$500 \quad CSI = \begin{cases} SEI - \rho_{NIR}, & \text{if } \rho_{NIR} \geq NDWI \\ SEI - NDWI, & \text{else} \end{cases} \quad (8)$$

$$\text{Where } SEI = \frac{(\rho_{aerosols} + \rho_{water\ vapor}) - (\rho_{green} + \rho_{NIR})}{(\rho_{aerosols} + \rho_{water\ vapor}) + (\rho_{green} + \rho_{NIR})}$$

$$NDWI = \frac{\rho_{green} - \rho_{NIR}}{\rho_{green} + \rho_{NIR}} \quad (9)$$

$$BI = \frac{\sqrt{(\rho_{red}^2 + \rho_{green}^2)/2}}{2a} \quad (10)$$

$$BSI = \frac{(\rho_{swir2} + \rho_{red}) - (\rho_{NIR} + \rho_{blue})}{(\rho_{swir2} + \rho_{red}) + (\rho_{NIR} + \rho_{blue})} \quad (11)$$

$$SATVI = \left(\frac{\rho_{swir1} - \rho_{red}}{\rho_{swir1} + \rho_{red} + 0.5} \right) \times 1.5 - \frac{\rho_{swir2}}{2} \quad (12)$$

where ρ_{red} , ρ_{green} , ρ_{blue} , ρ_{NIR} , ρ_{swir1} , ρ_{swir2} , $\rho_{aerosols}$, $\rho_{water\ vapor}$ are the surface reflectance of the red, green, blue, near infrared, shortwave infrared band1, shortwave infrared band2, aerosols and water vapor bands of the Sentinel-2 MSI sensor.

509 PALSAR data

510 ALOS PALSAR is an active microwave sensor with a multi-polarization configuration. In this study, we
 511 used the global 25-m PALSAR/PALSAR-2 mosaic from 2019 to 2021⁵⁹, which includes Horizontal-
 512 Horizontal (HH) and Horizontal-Vertical (HV) polarization backscattering coefficients, for light availability
 513 estimation. The dataset includes pre-applied slope correction and radiometric calibration. For further light
 514 regime modeling and mapping, we used the average HH and HV values from the three-year mosaic data. To
 515 match the spatial resolution of the predictor variables, the HH and HV backscatter layers were resampled
 516 from 25 m to 10 m.

517 Topographic predictors

518 Topographic predictors were derived from the 30 m resolution GLO-30 digital elevation model⁶⁰ to represent
 519 terrain controls on local light availability. Elevation data were used to compute slope and aspect using the
 520 Terrain algorithm in GEE. To quantify relative landform position and fine-scale variability in topography,
 521 we calculated multi-scale Topographic Position Indices (TPI)⁶¹ as the difference between the original
 522 elevation and its focal mean within circular neighborhoods of 150 m, 300 m, and 1500 m radii. All
 523 topographic predictors were subsequently resampled to a 10-m spatial resolution using nearest-neighbor
 524 resampling to match the target resolution of the light-availability model, ensuring spatial consistency across
 525 datasets.

526 EUNIS Habitat Maps

527 We used the EUNIS habitat maps, which provide spatially explicit predictions of 260 terrestrial habitat types
 528 (Level 3) across Europe at 100 m resolution²⁷. The dataset was produced using ensemble machine-learning
 529 models trained on over 590,000 vegetation plots from the European Vegetation Archive and a suite of

climatic, topographic, soil, and remote-sensing predictors. Model evaluation with spatial block cross-validation yielded an average F1-score of 0.78 for Level 1 formations and 0.80 for Level 3 habitat types.

Forest management strategies

To characterize light regimes across various forest management regimes across Europe, we used the recently developed continental-scale forest management map⁶². This dataset integrates multi-source indicators of management intensity, including long-term harvesting disturbance frequency and extent, forest age structure, distributions of fast-growing tree species, and the European Primary Forest Database⁶³. Using a rule-based decision tree, forest area was classified into five management classes with a spatial resolution of 1-km, unmanaged forest, “close-to-nature” forestry, combined-objective forestry, intensive forestry, and very intensive forestry. The class proportions were calibrated to match the estimates reported in FOREST EUROPE (2020)⁶⁴ and showed broad agreement with official statistics across regions (within a 5–10% range).

Preparation of prediction variables

To enhance the temporal and spectral information content, we aggregated all Sentinel-1 and Sentinel-2 observations from 2019 to 2021 following the principles of spectral-temporal variability analysis⁶⁵. Specifically, from these time series, we derived summary metrics for each variable and its derivatives, including the maximum, minimum, and standard deviation. To further capture the influence of vegetation structural complexity, we extracted local statistics integrating spectral and radar signals, specifically the maximum, mean, and standard deviation, within circular windows of 50 m radius. For PALSAR bands, only the spatial aggregation approach was applied. All predictors with different spatial resolutions were resampled to a common 10-m spatial resolution using the nearest-neighbor method.

Climate

To account for climatic effects, we included climate variables derived from WorldClim v2.1, representing long-term average conditions during 1970–2000 at a spatial resolution of 1 km⁶⁶. We selected four bioclimatic variables that capture major dimensions of climatic energy and water availability while minimizing collinearity among predictors: annual mean temperature (bio01), temperature seasonality (bio04), annual precipitation (bio12), and precipitation of the driest quarter (bio17). Climate values were extracted for each vegetation plot from the corresponding 1-km grid cell. For landscape-scale analyses, climatic conditions were represented by the mean climate values across all plots within each 10-km grid cell.

Soil

Soil properties were obtained from SoilGrids 2.0, a global soil information system providing spatially explicit predictions of soil characteristics at 250-m resolution⁶⁷. To characterize edaphic conditions relevant to plant diversity and vegetation structure, we considered a suite of soil variables describing soil physical,

chemical and nutrient properties, including cation exchange capacity (CEC), total nitrogen (Nitrogen), soil pH (PH), and soil organic carbon (SOC). For each variable, depth-specific estimates from five standard soil layers (0–5, 5–15, 15–30, 30–60 and 60–100 cm) were aggregated into a single profile-weighted value representing conditions within the upper 1 m of soil. Layer-specific values were weighted according to their depth intervals before averaging.

At the local scale, soil variables were extracted from the raster cell corresponding to each vegetation plot. For landscape-scale analyses, soil conditions were represented by the mean values across all plot locations within each 10-km grid cell. To reduce multicollinearity among environmental predictors while retaining major gradients in soil fertility, texture and chemistry, a subset of soil variables was selected for the final analyses.

Plant diversity measures

At the local scale, we used species richness (N_0) as the plant diversity response variable. At landscape scale, we used effective numbers of species, also known as Hill numbers⁶⁸. Hill numbers are parameterized by the order q , which determines the sensitivity of the diversity measure to species relative abundances and provides a unified, continuous framework for quantifying biodiversity. Different values of q assign varying weights to species frequencies, thereby emphasizing the contributions of rare species ($q = 0$), common species ($q = 1$), and dominant species ($q = 2$) to community diversity^{69,70}. In this study, species abundance was approximated using plot occupancy. Specifically, species occurrences were recorded as presence–absence data at the plot level, and the abundance of each species within a 10-km grid cell was represented by the number of plots in which the species occurred. Thus, species frequencies reflected landscape-scale occupancy rather than local abundance. Hill numbers N_q provide a unified framework for measuring diversity and are defined as:

$$N_q = \left(\sum_{i=1}^S p_i^q \right)^{\frac{1}{1-q}},$$

where p_i is the relative abundance of species i , S is the total number of species, and q determines the sensitivity of the index to common versus rare species.

For $q = 0$, $N_0 = S$, equals species richness.

For $q = 1$, the limit of N_q as $q \rightarrow 1$ yields the exponential of Shannon entropy:

$$N_1 = \exp \left(- \sum_{i=1}^S p_i \ln p_i \right)$$

which weights species in proportion to their abundance and reflects the diversity of common species.

For $q = 2$, Hill numbers reduce to the inverse Simpson index:

593

$$N_2 = \left(\sum_{i=1}^S p_i^2 \right)^{-1}$$

594

which emphasizes dominant species and is less sensitive to rare taxa.

595

Hill numbers integrate species richness (N_0), abundance-weighted diversity (N_1) and diversity among

596

dominant species (N_2), enabling consistent comparisons of plant diversity responses to mean light

597

availability and spatial light heterogeneity across landscapes.

598

Modeling light availability

599

We modeled the light regime as a function of the spectral, SAR backscatter, and topographic variables and

600

geographic coordinates, with a total of 328 predictors using the Random Forest (RF)⁷¹ machine learning

601

algorithm. The final modeling dataset included 59,733 plots for model training and validation. The RF model

602

was first trained using 80% of the plots and validated on the remaining 20% and was then re-trained using

603

all 59,733 plots to produce the final model.

604

To improve predictive accuracy and robustness, we evaluated 768 combinations of random-forest

605

hyperparameters. These including: number of trees (300, 500, 800, 1200), number of variables considered

606

at each split (mtry: default, 55, 164), minimum leaf population (5, 10, 20, 50), bag fraction (0.6, 0.7, 0.8,

607

0.9), and maximum number of terminal nodes (300, 500, 800, 1200). The final model adopted the

608

hyperparameter combination that minimized the Root Mean Square Error (RMSE). The optimized settings

609

included 500 trees, 164 variables considered at each split, a minimum leaf population of 20, a bag fraction

610

of 0.9, and a maximum of 1,200 terminal nodes. This optimized model was subsequently used to generate

611

spatial predictions across the pan-European region. The relative importance of each variable was evaluated

612

using the Mean Decrease in Gini (MDG) method⁷¹. We scaled the variable importance values per covariate

613

to a 0 - 1 scale for comparison purposes. All calculations were implemented in GEE, a cloud-based platform

614

for remote sensing data processing.

615

We evaluated predictive performance using the coefficient of determination (R^2), root mean square error

616

(RMSE) and mean absolute error (MAE). Model evaluation was conducted in two ways. First, we used a

617

randomly selected 20% subset of the data as an independent validation set. Second, we applied a tenfold

618

spatial cross-validation approach, in which models were evaluated using spatially separated folds to account

619

for potential spatial autocorrelation and to provide a more robust assessment of predictive performance.

620

$$R^2 = 1 - \frac{\sum_{i=1}^n (lE_i - lO_i)}{\sum_{i=1}^n (lE_i - \underline{lO})^2}$$

621

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (lE_i - lO_i)^2}{n}}$$

622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654

$$MAE = \frac{\sum_{i=1}^n |lE_i - lO_i|}{n}$$

Quantifying light-diversity relationships

Predictor selection and response variables

We selected predictors to represent climate, soil conditions and sampling effort, while mean light availability and light heterogeneity were retained as focal predictors throughout all analyses. Candidate climate predictors initially included the four bioclimatic variables, whereas candidate soil predictors included cation exchange capacity (CEC), soil nitrogen content, soil pH and soil organic carbon. To reduce redundancy among environmental predictors, variables were selected based on ecological relevance and screened for multicollinearity using pairwise correlation analysis and variance inflation factors (VIFs). All pairwise correlations were moderate ($|r| < 0.6$; *Supplementary Materials Fig.S4*), and all VIF values were below 5, indicating low collinearity among predictors. The final predictor set comprised mean annual temperature (bio01), annual precipitation (bio12), temperature seasonality (bio04), cation exchange capacity, soil nitrogen, soil pH and cumulative plot area, together with mean light availability and light heterogeneity.

Plant diversity was quantified using Hill numbers corresponding to species richness (N_0), the exponential of Shannon entropy (N_1) and the inverse Simpson index (N_2). Diversity responses were transformed as $\ln(N_q + 1)$ prior to modelling. For each cross-validation fold, predictor variables were standardized using the mean and standard deviation of the training data, and the same transformation was subsequently applied to the corresponding test data.

Spatial modelling framework

We evaluated the contribution of light regime to plant diversity using spatial generalized linear mixed models fitted with the package *sdmTMB*²⁸ in R (version 4.01). Models assumed Gaussian errors and incorporated a spatial Gaussian random field represented through a stochastic partial differential equation (SPDE) mesh. Spatial meshes were constructed from projected coordinates using a cutoff distance of 5 km at local scale, and 50 km at landscape scale.

To assess the contribution of light regime beyond environmental controls, we compared three hierarchical model formulations. The environment-only model included only climate, soil and sampling-effort covariates. The light-added model additionally incorporated separate smooth functions for mean light availability and light heterogeneity. The light interaction model replaced these independent smooths with a two-dimensional tensor-product smooth describing their joint effects on diversity. Smooth terms were fitted using penalized regression splines with basis dimension $k = 5$.

Model performance was evaluated using tenfold spatial block cross-validation implemented with *blockCV*⁷². Spatial folds were generated using 100-km blocks to reduce spatial dependence between training and test observations. For each fold, models were fitted to four training subsets and evaluated on the withheld subset.

Predictive performance was quantified using the coefficient of determination (R^2) and root mean squared error (RMSE) on the log-transformed response scale. The contribution of light regime was quantified as the difference in predictive performance between the light-added model and the environment-only model, whereas the additional contribution of light-regime interactions was quantified as the difference between the light interaction and light-added models.

Permutation importance

To assess the relative importance of individual predictors, we used the light-added model and implemented a permutation procedure within the spatial cross-validation framework. For each fold, baseline predictions were first generated for the withheld test data. Individual predictors were then randomly permuted while all remaining predictors were left unchanged, and predictions were recalculated using the fitted model. This procedure was repeated 20 times for each predictor and cross-validation fold.

Permutation importance was quantified as the increase in RMSE (ΔRMSE) and decrease in R^2 (ΔR^2) relative to the unpermuted predictions. Larger ΔRMSE and ΔR^2 values indicate greater dependence of model performance on a given predictor. Predictor-level importance was summarized across all folds and permutations using mean values and associated variability statistics.

Partial-response relationships

To visualize marginal associations between diversity and the two light-regime components, partial-response curves were estimated from the light-added model. For each focal predictor, a sequence of 200 values spanning the observed range was generated while all remaining predictors were held constant at their median standardized values.

Predictions were generated from the fixed-effects component of the model, excluding the spatial random field. Predicted values and associated standard errors were obtained on the log-transformed scale and converted to approximate 95% confidence intervals using ± 1.96 standard errors. Predictions and confidence intervals were subsequently back-transformed to the original Hill-diversity scale using $\exp(x) - 1$. The lower and upper 5% tails of the observed predictor distributions were indicated to identify regions with limited empirical support.

Mapping light-regime space

To characterize the joint effects of mean light availability and light heterogeneity, predictions from the interaction model were projected onto a regular two-dimensional light-regime grid spanning the observed ranges of both variables. Climate, soil and sampling-effort covariates were fixed at their median standardized values. Predictions were generated from fixed effects only and back-transformed to the original diversity scale.

To avoid extrapolation into poorly represented combinations of light availability and heterogeneity, predictions were restricted to the realized European light-regime space. This space was estimated using two-dimensional kernel density estimation of observed grid-cell light-regime values. We retained the region containing 99% of the estimated probability density and masked predictions outside this boundary.

To identify regions of exceptionally high predicted diversity, we calculated the 75th, 90th and 95th percentiles of predicted diversity values within the realized light-regime space. These thresholds were visualized as contour lines representing the upper 25%, 10% and 5% of predicted diversity conditions, respectively.

Model validation and residual spatial autocorrelation

Model validation was based on the same tenfold spatial block cross-validation framework used for model comparison. For each diversity metric, cross-validated predictions from the final interaction model were pooled across folds and compared with withheld observations. Model performance was summarized using the squared Pearson correlation coefficient (R^2), and RMSE on the original diversity scale. Observed-versus-predicted relationships were used to assess overall predictive agreement.

Spatial autocorrelation was explicitly accounted for during model fitting through the spatial Gaussian random field implemented in sdmTMB. To evaluate remaining residual spatial structure, Pearson residuals were extracted from the final interaction model, and Moran's I was calculated using an eight-nearest-neighbor spatial network based on projected coordinates. A row-standardized spatial weights matrix was used to quantify residual autocorrelation and assess whether substantial spatial structure remained after accounting for environmental covariates, light-regime variables and the spatial random field.

Data availability

The pan-European light regime map for 2020 has been deposited in Zenodo and will be made publicly available upon acceptance of the manuscript. The map can be explored interactively in this browser application: <https://ee-wwanben1994.projects.earthengine.app/view/eulr-10m-explorer>. The complete vegetation plot data, including species records, used in this study are stored within the EVA database repository at <https://doi.org/10.58060/250x-we61>, they can be accessed through a request to the EVA Coordinating Board at <https://euroveg.org/eva-database>. The Ecological Indicator Values for Europe (EIVE) 1.0 is freely available at <https://www.doi.org/10.5281/zenodo.7534792>. Sentinel-1 data, Sentinel-2 Level-2A surface-reflectance data and the Sentinel-2 Cloud Probability product are publicly available through the GEE Data Catalog. The 25-m PALSAR/PALSAR-2 annual mosaics are available from JAXA, and the Copernicus DEM GLO-30 elevation data are available through the GEE Data Catalog. EUNIS habitat maps are available from <https://zenodo.org/records/11108226>. WorldClim v2.1 bioclimatic data are available at <https://www.worldclim.org/data/worldclim21.html>. SoilGrids 2.0 soil-property data are available from ISRIC at <https://isric.org/explore/soilgrids>. The European forest management map is provided through <https://doi.org/10.34894/HQIJN5>. ESA WorldCover 10 m 2020 data is available in GEE Data Catalog.

Code availability

The code used for data processing, model fitting, and figure generation is publicly available on GitHub at: <https://github.com/WanbenWu/European-light-regime>.

Acknowledgements

This work was supported by the Horizon Europe project wildE (grant agreement No. 101081251). We also consider this work a contribution to the Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), funded by the Danish National Research Foundation (grant DNRF173). We thank the European Vegetation Archive (<https://euroveg.org/resurvey/>) for providing data (Project ID: 240) and Google Earth Engine for analytical support.

Author contributions

W.W., J.-C.S. and R.B. conceived the study. W.W. designed the methodology and conducted the data preparation and analyses, with assistance from S.P. in preprocessing the EVA plot data. J.L., J.D., J.E.M., Z.S., G.S., K.V.M., I.B., B.G., E.G., G.B., G.F., H.H.B., J.D., J.M., J.P., K.E., T.D., B.J.-A., C.D., F.A., F.N., J.Š., R.Ć., R.O., S.A., T.Č., M.D.S., U.J., W.S. and Ž.Š. contributed data. W.W. visualized the results and prepared all figures and tables with input from all authors. W.W. wrote the manuscript with substantial contributions from R.B. and J.-C.S. and additional input from all co-authors.

739 **Funding**

740 This study was supported by the Horizon Europe project wildE (grant agreement No. 101081251). We also
741 consider this work a contribution to the Center for Ecological Dynamics in a Novel Biosphere (ECONOVO),
742 funded by the Danish National Research Foundation (grant DNRF173).

743 **Competing interests**

744 The authors declare that they have no conflict of interest.

745

- 747 1 Ruban, A. V. Plants in light. *Communicative & Integrative Biology* **2**, 50–55 (2009).
- 748 2 Grunwald, Y., Yaaran, A. & Moshelion, M. Illuminating plant water dynamics: the role of
749 light in leaf hydraulic regulation. *New Phytologist* **241**, 1404–1414 (2024).
- 750 3 Austin, A. T. & Vivanco, L. Plant litter decomposition in a semi-arid ecosystem controlled by
751 photodegradation. *Nature* **442**, 555–558 (2006).
- 752 4 Helbach, J., Frey, J., Messier, C., Mörsdorf, M. & Scherer-Lorenzen, M. Light heterogeneity
753 affects understory plant species richness in temperate forests supporting the heterogeneity–
754 diversity hypothesis. *Ecology and Evolution* **12**, e8534 (2022).
- 755 5 Nicotra, A. B., Chazdon, R. L. & Iriarte, S. V. Spatial heterogeneity of light and woody
756 seedling regeneration in tropical wet forests. *Ecology* **80**, 1908–1926 (1999).
- 757 6 Silvertown, J. Plant coexistence and the niche. *Trends in Ecology & evolution* **19**, 605–611
758 (2004).
- 759 7 Storch, D., Bohdalková, E. & Okie, J. The more-individuals hypothesis revisited: the role of
760 community abundance in species richness regulation and the productivity–diversity
761 relationship. *Ecology Letters* **21**, 920–937 (2018).
- 762 8 Chesson, P. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and*
763 *Systematics* **31**, 343–366 (2000).
- 764 9 Dormann, C. F. *et al.* Plant species richness increases with light availability, but not
765 variability, in temperate forests understorey. *BMC Ecology* **20**, 43 (2020).
- 766 10 Márialigeti, S., Tinya, F., Bidló, A. & Ódor, P. Environmental drivers of the composition and
767 diversity of the herb layer in mixed temperate forests in Hungary. *Plant Ecology* **217**, 549–
768 563 (2016).
- 769 11 Bartels, S. F. & Chen, H. Y. Interactions between overstorey and understorey vegetation
770 along an overstorey compositional gradient. *Journal of Vegetation Science* **24**, 543–552
771 (2013).
- 772 12 Fuxai, X., Fousseni, F., Chungang, P., Huaijiang, H. & Xiuhai, Z. Effect of overstory on the
773 seasonal variability of understory herbs in primary broad-leaved Korean pine forest of
774 Changbai Mountain. *African Journal of Biotechnology* **13** (2014).
- 775 13 Vega-Álvarez, J., García-Rodríguez, J. A. & Cayuela, L. Facilitation beyond species richness.
776 *Journal of Ecology* **107**, 722–734 (2019).
- 777 14 Nieto-Lugilde, D. *et al.* Tree cover at fine and coarse spatial grains interacts with shade
778 tolerance to shape plant species distributions across the Alps. *Ecography* **38**, 578–589 (2015).
- 779 15 Czyżewski, S. & Svenning, J.-C. Temperate forest plants are associated with heterogeneous
780 semi-open canopy conditions shaped by large herbivores. *Nature Plants*, 1–16 (2025).
- 781 16 Pearce, E. A. *et al.* Substantial light woodland and open vegetation characterized the
782 temperate forest biome before *Homo sapiens*. *Science Advances* **9**, eadi9135 (2023).
- 783 17 Davoli, M. *et al.* Megafauna diversity and functional declines in Europe from the Last
784 Interglacial to the present. *Global Ecology and Biogeography* **33**, 34–47 (2024).
- 785 18 Svenning, J.-C. *et al.* The late-Quaternary megafauna extinctions: Patterns, causes, ecological
786 consequences and implications for ecosystem management in the Anthropocene. *Cambridge*
787 *Prisms: Extinction* **2**, e5 (2024).
- 788 19 McGrath, M. J. *et al.* Reconstructing European forest management from 1600 to 2010.
789 *Biogeosciences* **12**, 4291–4316 (2015).
- 790 20 Ustaoglu, E. & Collier, M. J. Farmland abandonment in Europe: An overview of drivers,
791 consequences, and assessment of the sustainability implications. *Environmental reviews* **26**,
792 396–416 (2018).

793 21 Montgomery, R. & Chazdon, R. Light gradient partitioning by tropical tree seedlings in the
794 absence of canopy gaps. *Oecologia* **131**, 165–174 (2002).

795 22 Webster, C., Mazzotti, G., Essery, R. & Jonas, T. Enhancing airborne LiDAR data for
796 improved forest structure representation in shortwave transmission models. *Remote Sensing of*
797 *Environment* **249**, 112017 (2020).

798 23 Alexander, C., Moeslund, J. E., Bøcher, P. K., Arge, L. & Svenning, J.-C. Airborne laser
799 scanner (LiDAR) proxies for understory light conditions. *Remote Sensing of Environment*
800 **134**, 152–161 (2013).

801 24 Dengler, J. *et al.* Ecological Indicator Values for Europe (EIVE) 1.0. *Vegetation*
802 *Classification and Survey* **4**, 7–29 (2023).

803 25 Chytrý, M. *et al.* European Vegetation Archive (EVA): an integrated database of European
804 vegetation plots. *Applied vegetation science* **19**, 173–180 (2016).

805 26 Zanaga, D. *et al.* ESA WorldCover 10 m 2021 v200. (2022).

806 27 Si-Moussi, S. *et al.* EUNIS habitat maps: enhancing thematic and spatial resolution for
807 Europe through machine learning. *Scientific Data* **12**, 1976 (2025).

808 28 Anderson, S. C., Ward, E. J., English, P. A., Barnett, L. A. & Thorson, J. T. sdmTMB: an R
809 package for fast, flexible, and user-friendly generalized linear mixed effects models with
810 spatial and spatiotemporal random fields. *Journal of Statistical Software* **115**, 1–46 (2025).

811 29 Buitenwerf, R., Sandel, B., Normand, S., Mimet, A. & Svenning, J. C. Land surface greening
812 suggests vigorous woody regrowth throughout European semi-natural vegetation. *Global*
813 *Change Biology* **24**, 5789–5801 (2018).

814 30 Estel, S. *et al.* Mapping farmland abandonment and recultivation across Europe using MODIS
815 NDVI time series. *Remote Sensing of Environment* **163**, 312–325 (2015).

816 31 Bakker, E. S. *et al.* Combining paleo-data and modern exclosure experiments to assess the
817 impact of megafauna extinctions on woody vegetation. *Proceedings of the National Academy*
818 *of Sciences* **113**, 847–855 (2016).

819 32 Valladares, F. & Niinemets, Ü. Shade tolerance, a key plant feature of complex nature and
820 consequences. *Annual Review of Ecology, Evolution, and Systematics* **39**, 237–257 (2008).

821 33 Stein, A., Gerstner, K. & Kreft, H. Environmental heterogeneity as a universal driver of
822 species richness across taxa, biomes and spatial scales. *Ecology Letters* **17**, 866–880 (2014).

823 34 Wright, D. H. Species-energy theory: an extension of species-area theory. *Oikos*, 496–506
824 (1983).

825 35 Levine, J. M. & HilleRisLambers, J. The importance of niches for the maintenance of species
826 diversity. *Nature* **461**, 254–257 (2009).

827 36 Sandom, C. J., Ejrnæs, R., Hansen, M. D. & Svenning, J.-C. High herbivore density
828 associated with vegetation diversity in interglacial ecosystems. *Proceedings of the National*
829 *Academy of Sciences* **111**, 4162–4167 (2014).

830 37 Ricklefs, R. E. Environmental heterogeneity and plant species diversity: a hypothesis. *The*
831 *American Naturalist* **111**, 376–381 (1977).

832 38 Leibold, M. A. *et al.* The metacommunity concept: a framework for multi-scale community
833 ecology. *Ecology Letters* **7**, 601–613 (2004).

834 39 Czyżewski, S. *et al.* Revisiting Europe's temperate forests: Palaeoecological evidence for an
835 herbivory-driven woodland-grassland mosaic biome. *Biological Conservation* **316**, 111749
836 (2026).

837 40 Czyżewski, S. & Svenning, J. (2025).

838 41 Zobel, M. The species pool concept as a framework for studying patterns of plant diversity.
839 *Journal of Vegetation Science* **27**, 8–18 (2016).

840 42 Tälle, M. *et al.* Grazing vs. mowing: A meta-analysis of biodiversity benefits for grassland
841 management. *Agriculture, Ecosystems & Environment* **222**, 200–212 (2016).

842 43 Perino, A. *et al.* Rewilding complex ecosystems. *Science* **364**, eaav5570 (2019).

843 44 Hédli, R., Šipoš, J., Chudomelová, M. & Utinek, D. Dynamics of herbaceous vegetation
844 during four years of experimental coppice introduction. *Folia Geobotanica* **52**, 83–99 (2017).

845 45 Bassett, T. J., Landis, D. A. & Brudvig, L. A. Effects of experimental prescribed fire and tree
846 thinning on oak savanna understory plant communities and ecosystem structure. *Forest*
847 *Ecology and Management* **464**, 118047 (2020).

848 46 Euro+ Med, E. Euro+ Med PlantBase—the information resource for Euro-Mediterranean plant
849 diversity. Available online at: <http://www.europlusmed.org> (2026).

850 47 Ostrowski, G. *et al.* Mean ecological indicator values: use EIVE but no cover-weighting.
851 *Vegetation Classification and Survey* **6**, 57–67 (2025).

852 48 Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone.
853 *Remote Sensing of Environment* **202**, 18–27 (2017).

854 49 Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation.
855 *Remote Sensing of Environment* **8**, 127–150 (1979).

856 50 Huete, A. *et al.* Overview of the radiometric and biophysical performance of the MODIS
857 vegetation indices. *Remote Sensing of Environment* **83**, 195–213 (2002).

858 51 Camps-Valls, G. *et al.* A unified vegetation index for quantifying the terrestrial biosphere.
859 *Science Advances* **7**, eabc7447 (2021).

860 52 Xiao, X. *et al.* Satellite-based modeling of gross primary production in an evergreen
861 needleleaf forest. *Remote Sensing of Environment* **89**, 519–534 (2004).

862 53 Xu, H. Modification of normalised difference water index (NDWI) to enhance open water
863 features in remotely sensed imagery. *International Journal of Remote Sensing* **27**, 3025–3033
864 (2006).

865 54 Vincini, M., Frazzi, E. & D'Alessio, P. A broad-band leaf chlorophyll vegetation index at the
866 canopy scale. *Precision Agriculture* **9**, 303–319 (2008).

867 55 Wang, L. & Qu, J. J. NMDI: A normalized multi-band drought index for monitoring soil and
868 vegetation moisture with satellite remote sensing. *Geophysical Research Letters* **34** (2007).

869 56 Sun, G. *et al.* Combinational shadow index for building shadow extraction in urban areas
870 from Sentinel-2A MSI imagery. *International Journal of Applied Earth Observation and*
871 *Geoinformation* **78**, 53–65 (2019).

872 57 Roy, P., Sharma, K. & Jain, A. Stratification of density in dry deciduous forest using satellite
873 remote sensing digital data—An approach based on spectral indices. *Journal of Biosciences*
874 **21**, 723–734 (1996).

875 58 Huete, A. R. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* **25**,
876 295–309 (1988).

877 59 Rosenqvist, A., Shimada, M., Ito, N. & Watanabe, M. ALOS PALSAR: A pathfinder mission
878 for global-scale monitoring of the environment. *IEEE Transactions on Geoscience and*
879 *Remote Sensing* **45**, 3307–3316 (2007).

880 60 Simard, M., Denbina, M., Marshak, C. & Neumann, M. A global evaluation of radar-derived
881 digital elevation models: SRTM, NASADEM, and GLO-30. *Journal of Geophysical*
882 *Research: Biogeosciences* **129**, e2023JG007672 (2024).

883 61 De Reu, J. *et al.* Application of the topographic position index to heterogeneous landscapes.
884 *Geomorphology* **186**, 39–49 (2013).

885 62 Scherpenhuijzen, N. *et al.* Mapping forest management regimes in Europe. *Forest Ecology*
886 *and Management* **594**, 122940 (2025).

887 63 Sabatini, F. M. *et al.* European primary forest database v2. 0. *Scientific Data* **8**, 220 (2021).

888 64 State of Europe's Forests 2020. (Forest Europe, 2020).
 889 65 Wu, W.-B. *et al.* A first Chinese building height estimate at 10 m resolution (CNBH-10 m)
 890 using multi-source earth observations and machine learning. *Remote Sensing of Environment*
 891 **291**, 113578 (2023).
 892 66 Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for
 893 global land areas. *International Journal of Climatology* **37**, 4302–4315 (2017).
 894 67 Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning.
 895 *PLoS One* **12**, e0169748 (2017).
 896 68 Hill, M. O. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**,
 897 427–432 (1973).
 898 69 Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: a framework for sampling
 899 and estimation in species diversity studies. *Ecological Monographs* **84**, 45–67 (2014).
 900 70 Colwell, R. K. *et al.* Models and estimators linking individual-based and sample-based
 901 rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology* **5**, 3–21
 902 (2012).
 903 71 Breiman, L. Random forests. *Machine Learning* **45**, 5–32 (2001).
 904 72 Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillera-Arroita, G. blockCV: An r package for
 905 generating spatially or environmentally separated folds for k-fold cross-validation of species
 906 distribution models. *Biorxiv*, 357798 (2018).
 907

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterials.pdf](#)