

Light regimes organize plant diversity across spatial scales

Wanben Wu

wanben.wu@bio.au.dk

Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), Department of Biology, Aarhus University <https://orcid.org/0000-0002-0581-9774>

Jens-Christian Svenning

Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), Department of Biology, Aarhus University <https://orcid.org/0000-0002-3415-0862>

Sean E. H. Pang

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

Jonathan Lenoir

Université de Picardie Jules Verne <https://orcid.org/0000-0003-0638-9582>

Jürgen Dengler

Zurich University of Applied Sciences <https://orcid.org/0000-0003-3221-660X>

Jesper Erenskjold Moeslund

Aarhus University <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

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

Koenraad Van Meerbeek

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

Idoia Biurrun

University of the Basque Country <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 <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

University of South Bohemia in České Budějovice <https://orcid.org/0000-0002-5829-4051>

Jonathan Millett

Loughborough University <https://orcid.org/0000-0003-4701-3071>

Josep Penuelas

Global Ecology Unit CREAL-CSIC-UAB <https://orcid.org/0000-0002-7215-0150>

Klaus Ecker

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

Tetiana Dziuba

National Academy of Sciences of Ukraine <https://orcid.org/0000-0001-8621-0890>

Borja Jiménez-Alfaro

Biodiversity Research Institute <https://orcid.org/0000-0001-6601-9597>

Ciara Dwyer

Lund University <https://orcid.org/0000-0002-7558-3664>

Fabio Attorre

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

Francesca Napoleone

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

Jozef Šibík

Donald Danforth Plant Science Center <https://orcid.org/0000-0002-5949-862X>

Renata Čušterevska

Faculty of Natural Sciences and Mathematics <https://orcid.org/0000-0002-3849-6983>

Romà Ogaya

Global Ecology Unit, CREAL-CSIC-Universitat Autònoma de Barcelona <https://orcid.org/0000-0003-4927-8479>

Svetlana Aćić

University of Belgrade <https://orcid.org/0000-0001-6553-3797>

Tomáš Černý

Czech University of Life Sciences, Faculty of Forestry and Wood Sciences <https://orcid.org/0000-0003-2637-808X>

Michele De Sanctis

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

Ute Jandt

German Centre for Integrative Biodiversity Research (iDiv) <https://orcid.org/0000-0002-3177-3669>

Wolfgang Schmidt

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

Željko Škvorc

University of Zagreb Faculty of Forestry and Wood Technology <https://orcid.org/0000-0002-2848-1454>

Robert Buitenwerf

Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), Department of Biology, Aarhus University <https://orcid.org/0000-0003-3356-2301>

Research Article

Keywords: Light regime, EIVE, Europe, Plant diversity

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10938320/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Light regimes organize plant diversity across spatial scales

Wanben Wu* (0000-0002-0581-9774)

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

Email: wanben.wu@bio.au.dk

Jens-Christian Svenning (0000-0002-3415-0862)

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

Email: svenning@bio.au.dk

Sean E. H. Pang (0000-0003-4574-6028)

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

Department of Biology, University of Oxford, Oxford, UK

Email: sean.pang@biology.ox.ac.uk

Jonathan Lenoir (0000-0003-0638-9582)

Affiliation: UMR CNRS 7058 Ecologie et Dynamique des Systèmes Anthropisés (EDYSAN), Université de Picardie Jules Verne, Amiens, France

Email: jonathan.lenoir@u-picardie.fr

Jürgen Dengler (0000-0003-3221-660X)

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

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

Email: dr.juergen.dengler@gmail.com

Jesper Erenskjold Moeslund (0000-0001-8591-7149)

Affiliation: Department of Ecoscience, Aarhus University, Denmark

Email: jesper@ecos.au.dk

Zvezdana Stančić (0000-0002-6124-811X)

Affiliation: Faculty of Geotechnical Engineering University of Zagreb, Croatia

Email: zvstan@gfv.hr

Grzegorz Swacha (0000-0002-6380-2954)

Affiliation: Botanical Garden, University of Wrocław, Poland

Email: grzegorz.swacha@uwr.edu.pl

Koenraad Van Meerbeek (0000-0002-9260-3815)

Affiliation: Department of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium;

KU Leuven Plant Institute, KU Leuven, Leuven, Belgium

Email: koenraad.vanmeerbeek@kuleuven.be

Idoia Biurrun (0000-0002-1454-0433)

Affiliation: University of the Basque Country (EHU), Spain

Email: idoia.biurrun@ehu.eus

Behlül Güler (0000-0003-2638-4340)

Affiliation: Biology Education, Dokuz Eylül University, Izmir, Türkiye

Email: behlul.guler2@gmail.com

Emmanuel Garbolino (0000-0002-4954-6069)

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

Email: garbolino_emmanuel@yahoo.fr

Gianmaria Bonari (0000-0002-5574-6067)

Affiliation: Department of Life Sciences, University of Siena, Siena, Italy

Email: gianmaria.bonari@unisi.it

Giuseppe Fenu (0000-0003-4762-5043)

Affiliation: Department of Life and Environmental Sciences, University of Cagliari, Italy

Email: gfenu@unica.it

Hans Henrik Bruun (0000-0003-0674-2577)

Affiliation: Department of Biology, University of Copenhagen, Denmark

Email: hbbuun@bio.ku.dk

Jiri Dolezal (0000-0002-5829-4051)

Affiliation: Institute of Botany of Czech Academy of Sciences, Pruhonice, Czechia; Department of Botany, Faculty of Science, University of South Bohemia, Ceske Budejovice, Czechia

Email: jiridolezal@gmail.com

Jonathan Millett (0000-0003-4701-3071)

Affiliation: Geography and Environment, Loughborough University, Loughborough, LE11 3TU, UK

Email: j.millett@lboro.ac.uk

Josep Penuelas (0000-0002-7215-0150)

64 Affiliation: CSIC, Global Ecology Unit CREAM-CSIC-UAB, 08193 Bellaterra, Catalonia, Spain; and CREAM, 08193
 65 Cerdanyola del Vallès, Catalonia, Spain
 66 Email: josep.penuelas@creaf.cat
 67 **Klaus Ecker** (0000-0003-3388-1353)
 68 Affiliation: Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland
 69 Email: klaus.ecker@wsl.ch
 70 **Tetiana Dziuba** (0000-0001-8621-0890)
 71 Affiliation: Department of Geobotany and Ecology, M.G. Kholodny Institute of Botany, National Academy of Sciences of
 72 Ukraine, Kyiv, Ukraine
 73 Email: tdziuba2014@gmail.com
 74 **Borja Jiménez-Alfaro** (0000-0001-6601-9597)
 75 Affiliation: Biodiversity Research Institute (CSIC/UO/PA), University of Oviedo, Mieres, Spain. Department of Biology and
 76 Organism Systems. Oviedo, Spain
 77 Email: jimenezalfaro@uniovi.es
 78 **Ciara Dwyer** (0000-0002-7558-3664)
 79 Affiliation: Department of Earth and Environmental Sciences, Lund University, Lund, Sweden and Geography and
 80 Environment, Loughborough University, Loughborough, LE11 3TU, UK
 81 Email: ciara.dwyer@mgeo.lu.se
 82 **Fabio Attorre** (0000-0002-7744-2195)
 83 Affiliation: Department of Environmental Biology, Sapienza University of Rome, 00185 Rome, Italy
 84 Email: fabio.atorre@uniroma1.it
 85 **Francesca Napoleone** (0000-0002-3807-7180)
 86 Affiliation: Department of Environmental Biology, Sapienza University of Rome, Rome, Italy.
 87 Email: francesca.napoleone@uniroma1.it
 88 **Jozef Šibík** (0000-0002-5949-862X)
 89 Affiliation: Plant Science & Biodiversity Center SAS, Bratislava, Slovakia
 90 Email: jozef.sibik@savba.sk
 91 **Renata Čušterevska** (0000-0002-3849-6983)
 92 Affiliation: Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University, Skopje, North Macedonia.
 93 Email: renatapmf@yahoo.com
 94 **Romà Ogaya** (0000-0003-4927-8479)
 95 Affiliation: Global Ecology Unit, CREAM-CSIC-Universitat Autònoma de Barcelona, Cerdanyola del Vallès (Catalonia), E-
 96 08193, Spain.
 97 Email: r.ogaya@creaf.uab.cat
 98 **Svetlana Ačić** (0000-0001-6553-3797)
 99 Affiliation: University of Belgrade, Faculty of Agriculture, Department of Botany, Belgrade, Serbia
 100 Email: acic@agrf.bg.ac.rs
 101 **Tomáš Černý** (0000-0003-2637-808X)
 102 Affiliation: Czech University of Life Sciences, Faculty of Forestry and Wood Sciences, Prague, Czechia
 103 Email: cernyt@fld.czu.cz
 104 **Michele De Sanctis** (0000-0002-7280-6199)
 105 Affiliation: Department of Environmental Biology, Sapienza University of Rome, Rome, Italy.
 106 Email: michele.desanctis@uniroma1.it
 107 **Ute Jandt** (0000-0002-3177-3669)
 108 Affiliation: Institute of Biology/Geobotany and Botanical Garden, Martin Luther University Halle-Wittenberg, Halle,
 109 Germany.
 110 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany.
 111 Email: ute.jandt@botanik.uni-halle.de
 112 **Wolfgang Schmidt** (0000-0001-5356-4684)
 113 Affiliation: Department of Silviculture and Forest Ecology of the Temperate Zones, University of Göttingen, Göttingen,
 114 Germany
 115 Email: wschmid1@gwdg.de
 116 **Željko Škvorc** (0000-0002-2848-1454)
 117 Affiliation: University of Zagreb Faculty of Forestry and Wood Technology, Zagreb, Croatia
 118 Email: zskvorc@sumfak.hr
 119 **Robert Buitenwerf** (0000-0003-3356-2301)
 120 Affiliation: Center for Ecological Dynamics in a Novel Biosphere (ECONOVO), Department of Biology, Aarhus University,
 121 Denmark
 122 Email: buitenwerf@bio.au.dk
 123

Abstract

Light regulates ecological processes from organism performance to ecosystem functioning, yet its role in shaping plant diversity across spatial scales remains unresolved. Here, we combine >650,000 vegetation plots, 14,835 species-specific light indicator values based on Ecological Indicator Values for Europe (EIVE), and Earth observation data to provide the first pan-European assessment of how light regimes near the ground shape plant diversity across scales. Incorporating light-regime metrics beyond macroclimatic and edaphic drivers improved predictions of plant diversity, with relative increases in explained variance of 3% and 27.4 - 34.5% at the local and landscape scale, respectively. Plant diversity peaked under intermediate-to-high light availability coupled with high spatial heterogeneity. Strikingly, at the landscape level, we identified a pronounced mismatch between the light regimes currently prevailing across Europe and those associated with high plant diversity. Together, these findings reveal light regime as a fundamental axis for plant biodiversity and suggest that widespread shifts towards very low light availability and reduced spatial heterogeneity may decrease plant diversity across Europe. Maintaining non-shaded or increasing spatial heterogeneity in light conditions near the ground could therefore help reduce this mismatch, expanding the extent of light regimes with high potential to support plant diversity.

Main

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.

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, and evidence for spatial heterogeneity in light conditions near the ground is similarly mixed and restricted to forest understory^{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 restricted spatial grain and extent 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. This raises the question of whether long-term canopy closure and forest expansion have shifted contemporary European light regimes away from the conditions that support high plant diversity.

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 integrate species-specific light preferences from Ecological Indicator Values for Europe (EIVE)²⁴, for 14,835 vascular plant species across more than 650,000 vegetation plots from the European Vegetation Archive (EVA)²⁵ combined with multi-source Earth observation data 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: first, light-regime predictors improve plant diversity models beyond macroclimatic and edaphic variables; second, mean light availability and spatial heterogeneity interact to shape plant diversity more strongly than either dimension alone. Finally, we asked whether a mismatch exists between contemporary European light regimes and the light-regime conditions 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 topography predictors (*Extended Data Table 1*). Models were applied to all natural and semi-natural areas, as defined by 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.

Model performance was consistent across the eight biogeographic regions of Europe, explaining 56–73% of the observed variance in light availability and with a RMSE between 0.32 and 0.8 (*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, a 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*).

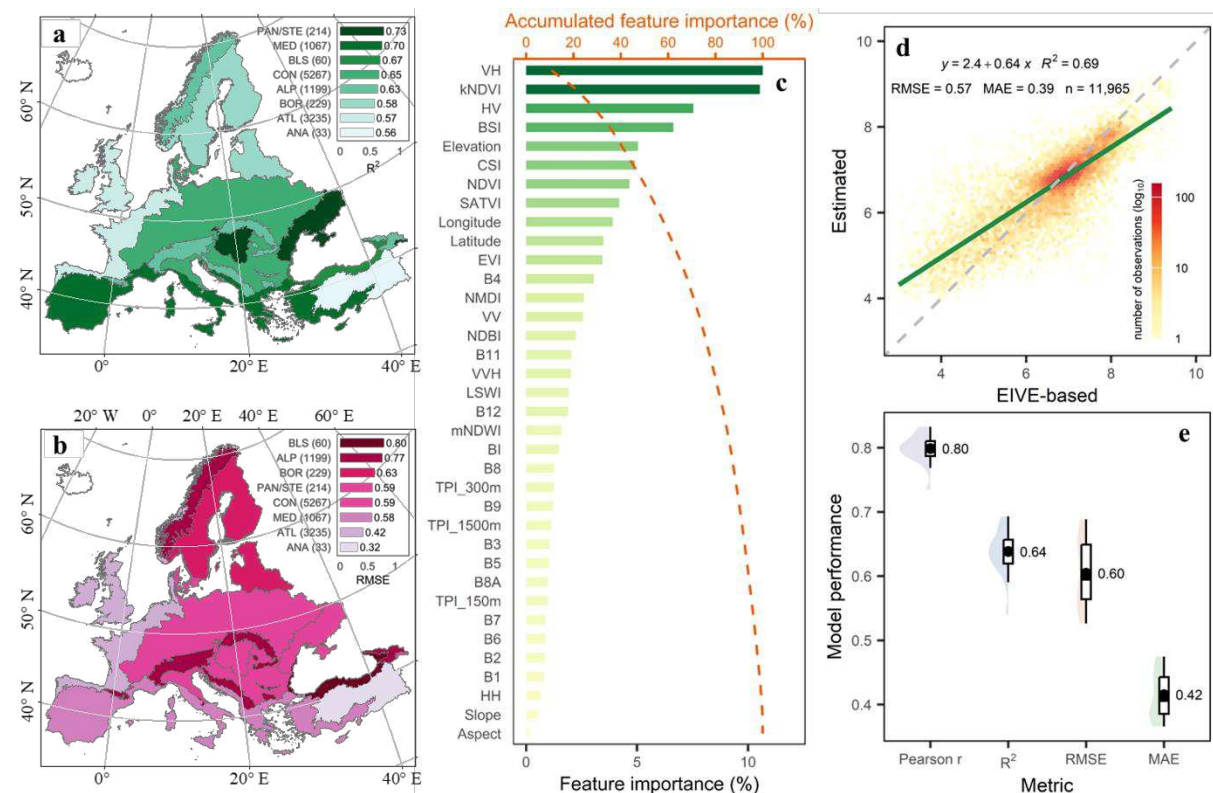


Fig.1 Validation and feature importance analysis of the continental light-availability model.

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*, Observed versus predicted light availability under tenfold cross-validation, demonstrating robust predictive performance across Europe.

Geospatial patterns of light regimes across Europe

Our map of light availability provides the first ever fine-grained (10 m) spatial distribution of near ground light regimes across Europe (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 ($\sim 40 - 55^\circ$) (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) (Extended Data Fig.2).

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 (Extended Data Fig.1 a–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 (Extended Data Fig.1 c–d). Across landscapes, light availability and its local spatial heterogeneity exhibit a pronounced nonlinear relationship (Extended Data Fig.1 e), with heterogeneity peaking at intermediate-to-high levels.

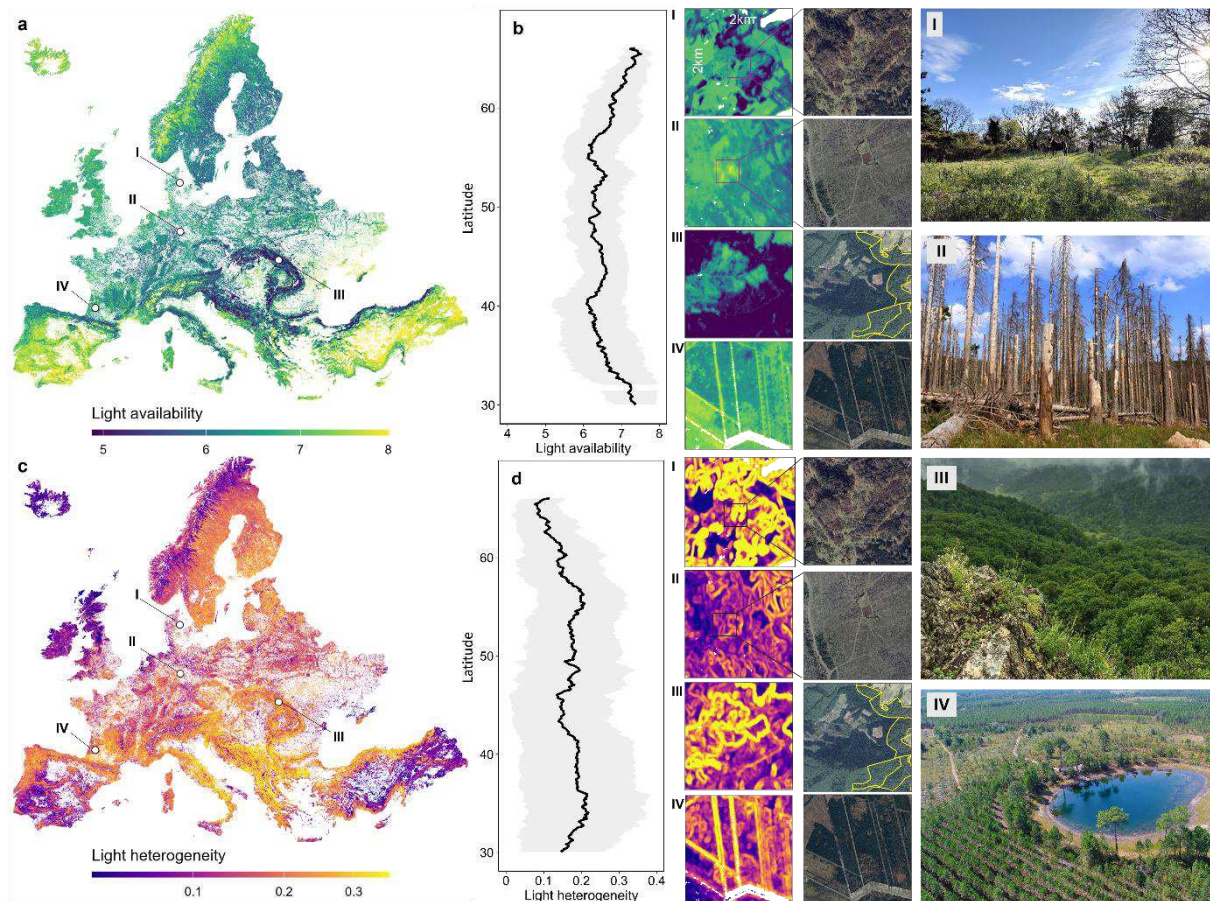


Fig.2 European light-regime maps near the ground.

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/eulr10m-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) (*Extended Data Fig. 3*), 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. The 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 (*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 ≈ 0.23 - 0.27), reflecting greater spatial variation in canopy openness.

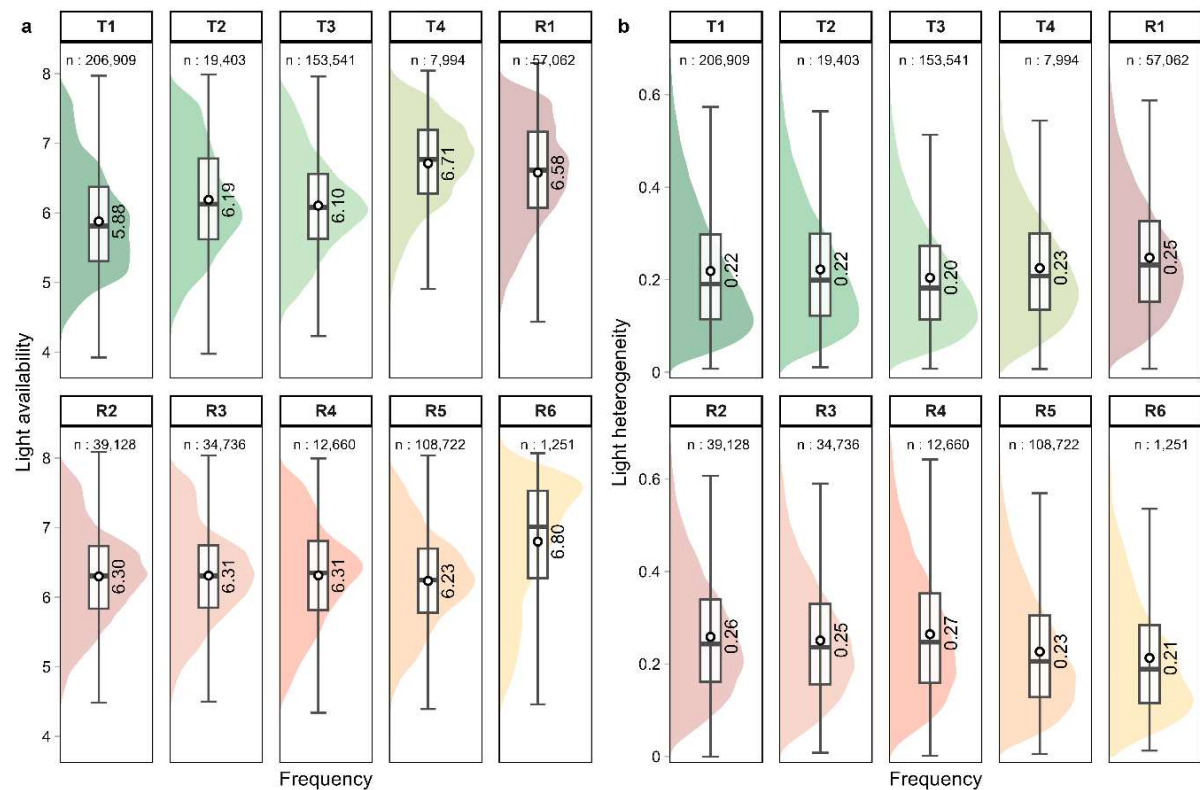


Fig.3 Distributions of modeled light regimes across EUNIS habitat types.

a, Light availability, and *b*, light heterogeneity (within-50m radius standard deviation), 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. Where 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. Extended Data Fig. 3 shows the light regime patterns across EUNIS level 3 forest (T) and grassland (R) habitat types.

Light regime near the ground as a predictor of plant diversity

To quantify the contribution of light regime to plant-diversity predictions, we fitted three hierarchical spatial and spatiotemporal GLMMs (Generalized Linear Mixed Effects Models) models using sdmTMB package²⁸: (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, at landscape scale, the model that included light variables increased explained variance by 27.4% for species richness (N_0 ; $\Delta R^2 = 0.104$), 33.2% for Shannon diversity (N_1 ; $\Delta R^2 = 0.107$), and 34.5% for Simpson diversity (N_2 ; $\Delta R^2 = 0.106$). Adding an interaction between mean light availability and light heterogeneity yielded only marginal further improvements, indicating that their effects on plant diversity were largely additive. 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. Five-fold spatial cross-validation further demonstrated strong agreement between observed and predicted diversity, particularly at the landscape scale (*Extended Data Fig. 5*). Moran's I tests indicated little residual spatial autocorrelation in the final interaction models (Moran's I < 0.05, P > 0.05; *Extended Data Table 2*), 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
	N_2	Environment-only model	0.307 ± 0.018	0.579 ± 0.007
		Light-added model	0.409 ± 0.014	0.535 ± 0.008
		Light interaction model	0.413 ± 0.014	0.534 ± 0.008
Local	N_0	Environment-only model	0.134 ± 0.007	0.600 ± 0.003
		Light-added model	0.137 ± 0.005	0.600 ± 0.004
		Light interaction model	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.

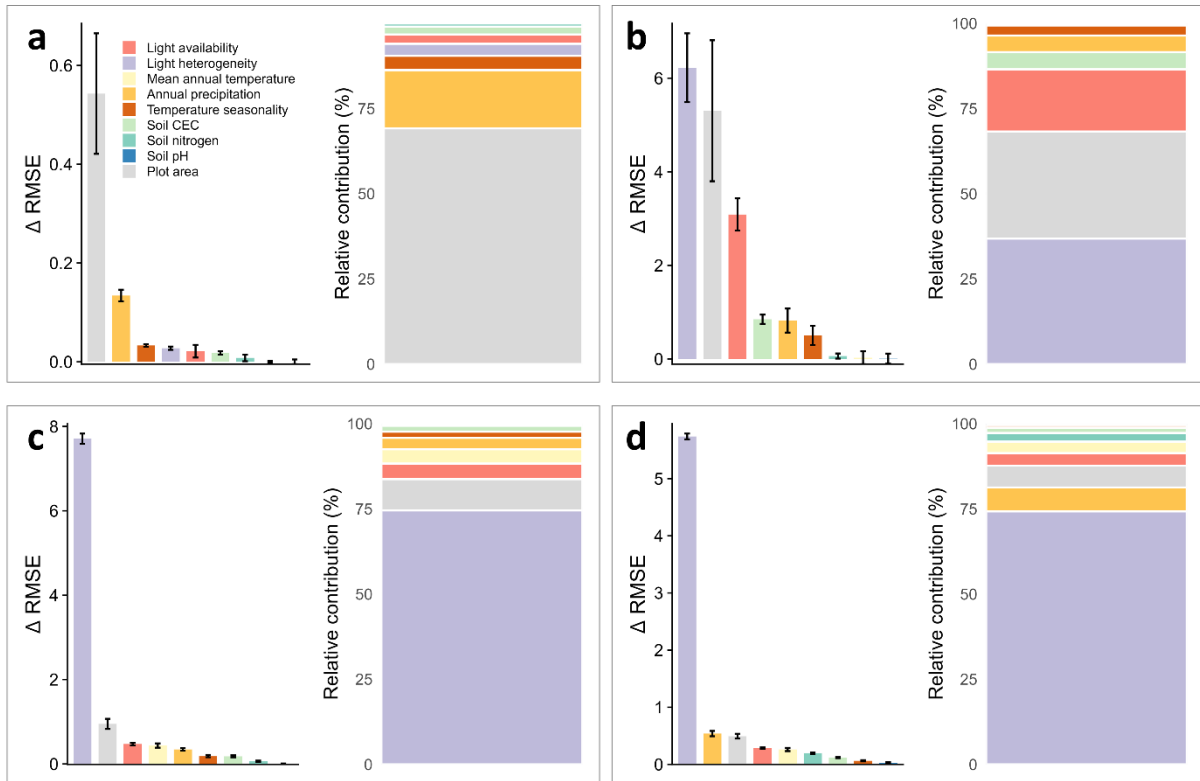


Fig. 4 Relative contributions of light-regime, climatic and edaphic predictors to plant diversity across spatial scales.

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

By modelling the joint effects (the light interaction models) of mean light availability and spatial heterogeneity, we identified a distinct light-regime space associated with elevated 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, spatial 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.

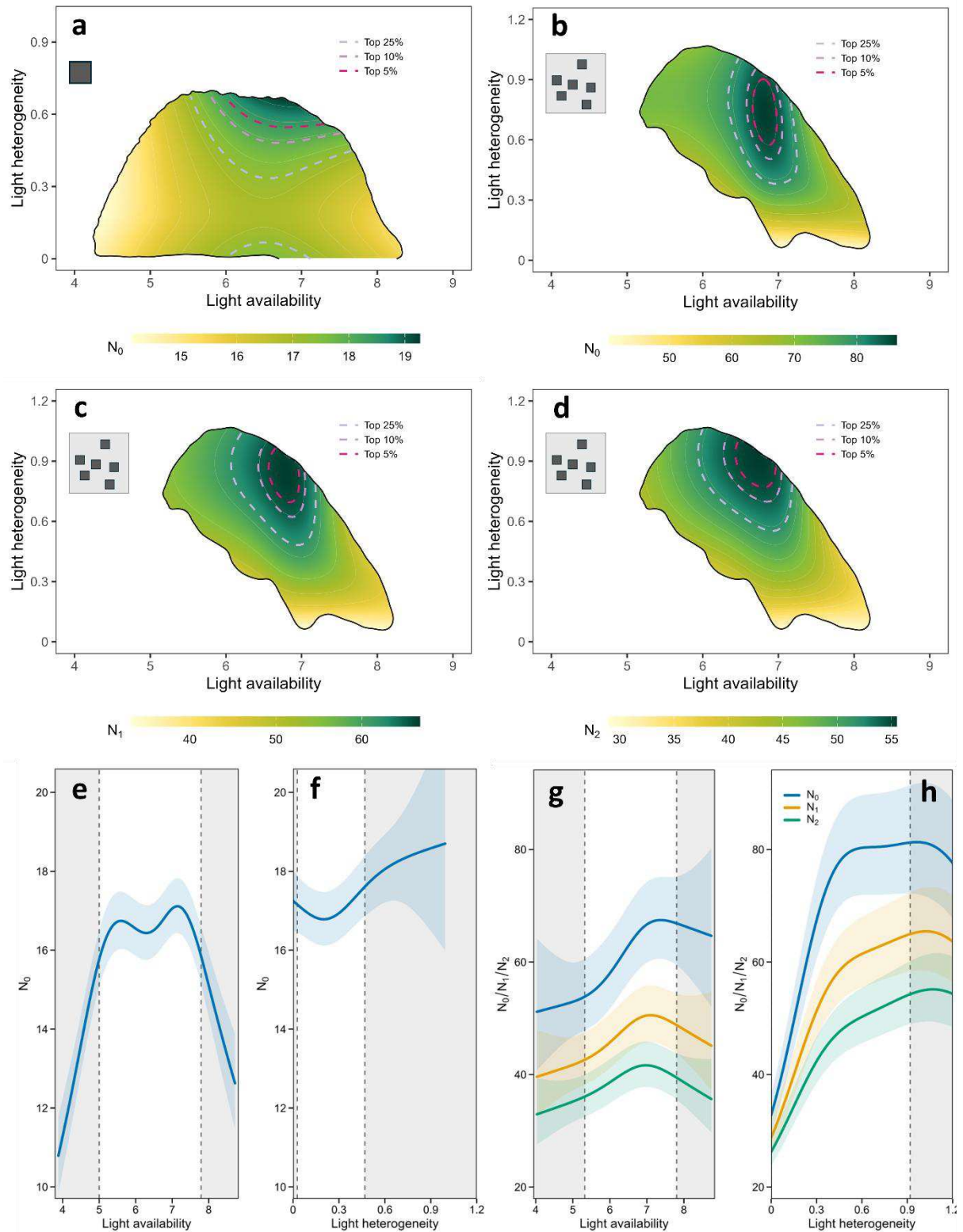


Fig. 5 | Predicted plant-diversity responses to light regimes near the ground.

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 light availability, whereas forests under “close-to-nature” management occupied the darkest portion of the observed space, and thus exhibiting the lowest potential for high local plant species diversity. 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.

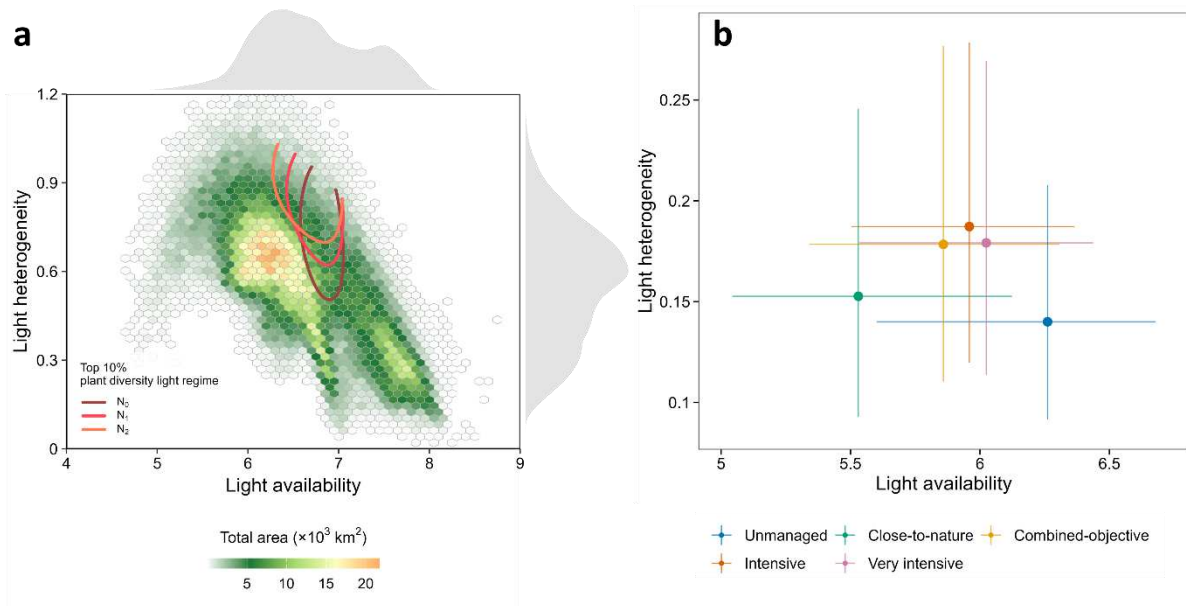


Fig. 6 Continental mismatch between realized and the high-plant-diversity light regimes and management-specific differentiation of light-regime space.

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

Light availability has long been recognized as a driver of plant growth, community assembly and vegetation succession²⁹, yet it has rarely been treated as an integrative axis for explaining plant biodiversity patterns over large areas and across spatial scales. Here, we developed a continental-scale, 10-m product of ecologically inferred light availability near the ground and used it to quantify how light regime structure plant diversity patterns from individual vegetation plots to entire landscapes. We show that high plant diversity was associated with a distinct light regime near the ground, characterized by intermediate-to-high light availability and high spatial heterogeneity. These relationships were weak at the local scale but much stronger at the landscape scale, where contemporary European light regimes diverged from those associated with the highest plant diversity.

This pattern suggests that plant diversity was shaped by the joint effects of light availability near the ground and its spatial heterogeneity, rather than by either dimension alone (Fig. 5). 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, diversity did not peak under the highest light availability, indicating that resource supply alone is unlikely to explain the observed optimum. Instead, the highest diversity occurred where relatively open conditions coincided with strong spatial heterogeneity in light conditions near the ground. Such heterogeneity implies a broader range of light niches, reduce the dominance of any single competitive strategy, and allow species with contrasting light requirements to coexist across landscapes³⁰. This interpretation is consistent with resource-based theory³¹, which predicts that energy availability can constrain richness, but also with coexistence theory, which emphasizes that spatial variation is needed for resources to support multiple species rather than a few dominant competitors³².

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. The majority of European plant species evolved under landscapes containing mosaics of open habitats, semi-open woodlands and closed-canopy patches, maintained by herbivory and other disturbance^{16,33}. If a large fraction of the regional flora is adapted to such conditions, contemporary diversity should be highest where light regimes still approximate these historically heterogeneous environments. Our results are therefore consistent with a combined ecological and historical mechanism: intermediate-to-high light availability may relax local light limitation, whereas spatial heterogeneity and landscape history determine whether that resource space is partitioned among many species.

The scale dependence of light-regime effects provides important insight into how light near the ground structures plant diversity. Light availability is clearly a microsite-scale driver of plant performance and community composition, but our results show that light regimes also regulate diversity as spatially integrated landscape properties. At the local scale, light availability and light heterogeneity contributed relatively little additional explanatory power beyond macroclimate, soil conditions and plot area. In contrast, at the landscape scale, spatial heterogeneity emerged as one of the strongest predictors of plant diversity. Heterogeneous landscapes may support more species simply because they contain a wider range of habitat types, successional stages and competitive environments. In addition, spatially structured variation in light conditions may promote species sorting, mass effects and spatial storage effects, thereby allowing species with different ecological strategies to persist within the same landscape. This interpretation links our findings to heterogeneity-diversity theory^{4,34} and metacommunity perspectives³⁵. Finally, landscapes with heterogeneous 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^{36,37}, in line with the species pool hypothesis³⁸.

Our analyses further suggest that contemporary European landscapes occupy only a limited portion of the light-regime space associated with high plant diversity (*Fig. 6a*). This mismatch may reflect a long-term shift away from historically more open and structurally heterogeneous conditions, linked to woody densification³⁹, land abandonment⁴⁰, and the prior loss and persisting low levels large herbivores⁴¹. As vegetation becomes increasingly dense and homogeneous, landscapes may move further away from the light-regime conditions that can support a large proportion of regional species pool: combining intermediate levels of light availability with a spatially heterogeneous light distribution. Consequently, biodiversity-oriented management should consider not only forest extent or canopy cover, but also the spatial configuration of light conditions across landscapes. Processes that maintaining mosaics of open, semi-open and closed-canopy habitats may therefore be critical for sustaining plant diversity at broader spatial scales.

These findings also caution against assuming that conservation-oriented management automatically maintains light regimes conducive to high biodiversity. In our analysis, “close-to-nature” forestry was associated with darker and more homogeneous light conditions (*Fig. 6b*), suggesting that management approaches emphasizing canopy continuity, retention forestry or low-intensity intervention do not necessarily generate the structural heterogeneity required to support diverse plant communities. 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. In such systems, active restoration measures, such as naturalistic grazing⁴², trophic rewilding⁴³, coppicing⁴⁴ or prescribed burning⁴⁵, may be used to maintain or restore a mosaic of open and high spatial heterogeneity in light conditions associated with high plant diversity.

More broadly, our results suggest that light regimes constitute a previously underappreciated axis of plant diversity organization. By integrating mean light availability and spatial heterogeneity, the light-regime framework links vegetation structure, disturbance dynamics and biodiversity across spatial scales, while substantially improving predictions of plant diversity beyond traditional climate and soil predictors. This framework provides a more general basis for understanding and forecasting biodiversity responses to global change drivers. Although developed here using European plant communities, this framework is likely applicable to a wide range of ecosystems in which vegetation structure, land use and disturbance regulate access to light. Monitoring shifts in light regimes may therefore provide a useful way to assess whether landscapes are moving toward or away from environmental conditions that support high plant diversity under ongoing global change.

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 the interaction between mean light availability and its spatial heterogeneity as a key regulator of plant diversity. Plant diversity was maximized under light regimes combining relatively intermediate-to-high light availability with pronounced spatial heterogeneity. This pattern is likely attributable to the joint effects of resource availability, niche differentiation and the historical adaptation of European floras to a mosaic of open habitats, semi-open woodlands and closed-canopy patches, rendering the distribution of light near the ground more heterogeneous. We also observed strong scale dependence, with light-regime effects being modest at the local scale but emerging as major predictors of diversity 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 strongly contributed to this variation, with “close-to-nature” forestry producing particularly dark light regimes, whereas more intensive strategies mainly increased light heterogeneity. Our findings establish the light regime as an underappreciated dimension of biodiversity regulation and highlight the need to manage landscapes not in terms of canopy cover continuity exclusively, but for the spatial configuration of heterogeneous light environments that sustain plant diversity.

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 reaching the near 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.

Sentinel-1 backscatter data

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

Sentinel-2 spectral data

Sentinel-2 is a wide-swath, multi-spectral (13 spectral channels), high spatial resolution (10 m for visible and near infrared, 20 m for other near infrared and short-wave infrared band), and high temporal resolution (revisit period 5 days) imaging mission supporting Copernicus Land Monitoring studies. A total of 597,956 Sentinel-2 Level-2A scenes on surface reflectance from January 2019 to December 2021 were processed across Europe. To remove low-quality observations, we masked clouds and cirrus using the Sentinel-2 QA60 band and the Sentinel-2 Cloud Probability product, excluding pixels with an estimated cloud probability exceeding 25%. The distribution of total Sentinel-2 observations used in the final analysis is shown in *Extended Data Fig. 6b*.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

PALSAR data

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

Topographic predictors

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

EUNIS Habitat Maps

We used the EUNIS habitat maps, which provides spatially explicit predictions of 260 terrestrial habitat types (Level 3) across Europe at 100 m resolution²⁷. The dataset was produced using ensemble machine-learning 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. The dataset offers wall-to-wall maps with uncertainty layers and is publicly available via Zenodo (zenodo.org/records/11108226).

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).

Preparations 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 weighed 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.

669 **Plant diversity measures**

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

$$681 \quad N_q = \left(\sum_{i=1}^S p_i^q \right)^{\frac{1}{1-q}},$$

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

684 For $q = 0$, $N_0 = S$, corresponds to species richness.

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

$$686 \quad N_1 = \exp \left(- \sum_{i=1}^S p_i \ln p_i \right)$$

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

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

$$689 \quad N_2 = \left(\sum_{i=1}^S p_i^2 \right)^{-1}$$

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

691 Hill numbers integrate species richness (N_0), abundance-weighted diversity (N_1) and diversity among
692 dominant species (N_2), enabling consistent comparisons of plant diversity responses to mean light
693 availability and spatial light heterogeneity across landscapes.

Modeling light availability

We modeled the light regime as a function of the spectral, SAR backscatter, and topography factors, location, with a total of 328 predictors using the Random Forest (RF)⁷¹ machine learning algorithm. The final modeling dataset included 59,733 plots for model training and validation. The RF model was first trained using 80% of the plots and validated on the remaining 20% and was then re-trained using all 59,733 plots to produce the final model.

To better achieve more robust and high-accuracy estimation, we systematically tested a range of hyperparameter combinations for 768 random forest models, including: number of trees (300, 500, 800, 1200), number of variables considered at each split (mtry: default, 55, 164), minimum leaf population (5, 10, 20, 50), bag fraction (0.6, 0.7, 0.8, 0.9), and maximum number of terminal nodes (300, 500, 800, 1200). The final model adopted the hyperparameter combination that minimized the Root Mean Square Error (RMSE). The optimized settings included 500 trees, 164 variables considered at each split, a minimum leaf population of 20, a bag fraction of 0.9, and a maximum of 1,200 terminal nodes. This optimized model was subsequently used to generate spatial predictions across the pan-European region. The relative importance of each variable was evaluated using the Mean Decrease in Gini (MDG) method⁷¹. We scaled the variable importance values per covariate to a 0 - 1 scale for comparison purposes. All calculations were implemented in GEE, a cloud-based platform for remote sensing data processing.

To evaluate the model's performance, several commonly used metrics are employed, including the coefficient of determination (R^2), RMSE, and Mean Absolute Error (MAE), which respectively quantify the explained variance, total prediction error, and average absolute deviation. Model evaluation was conducted in two ways. First, we used a randomly selected 20% subset of the data as an independent validation set. Second, we applied a five-fold spatial cross-validation approach, in which models were evaluated using spatially separated folds to account for potential spatial autocorrelation and to provide a more robust assessment of predictive performance.

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

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

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

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 ($|r| < 0.6$; *Extended Data Fig. 4*), 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 R package sdmTMB²⁸. 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 five-fold spatial block cross-validation implemented with blockCV⁷². 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 96% 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 five-fold 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/eulr10m-explorer>. EVA plot data are available under restricted access, as they are owned by the contributors of each vegetation database. Data can be requested from the European Vegetation Archive (EVA) 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>. ESA WorldCover 10 m 2020 data is available at GEE (assets id: ESA/WorldCover/v100).

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>.

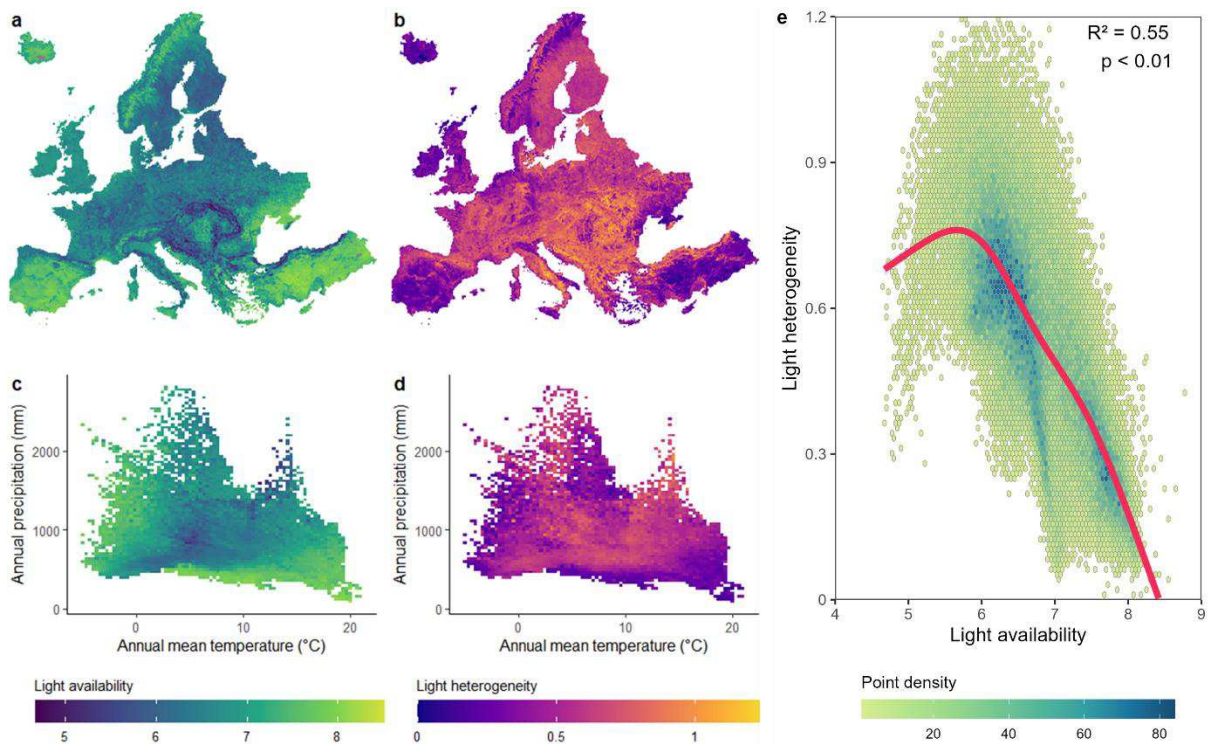
822 **Acknowledgements**

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

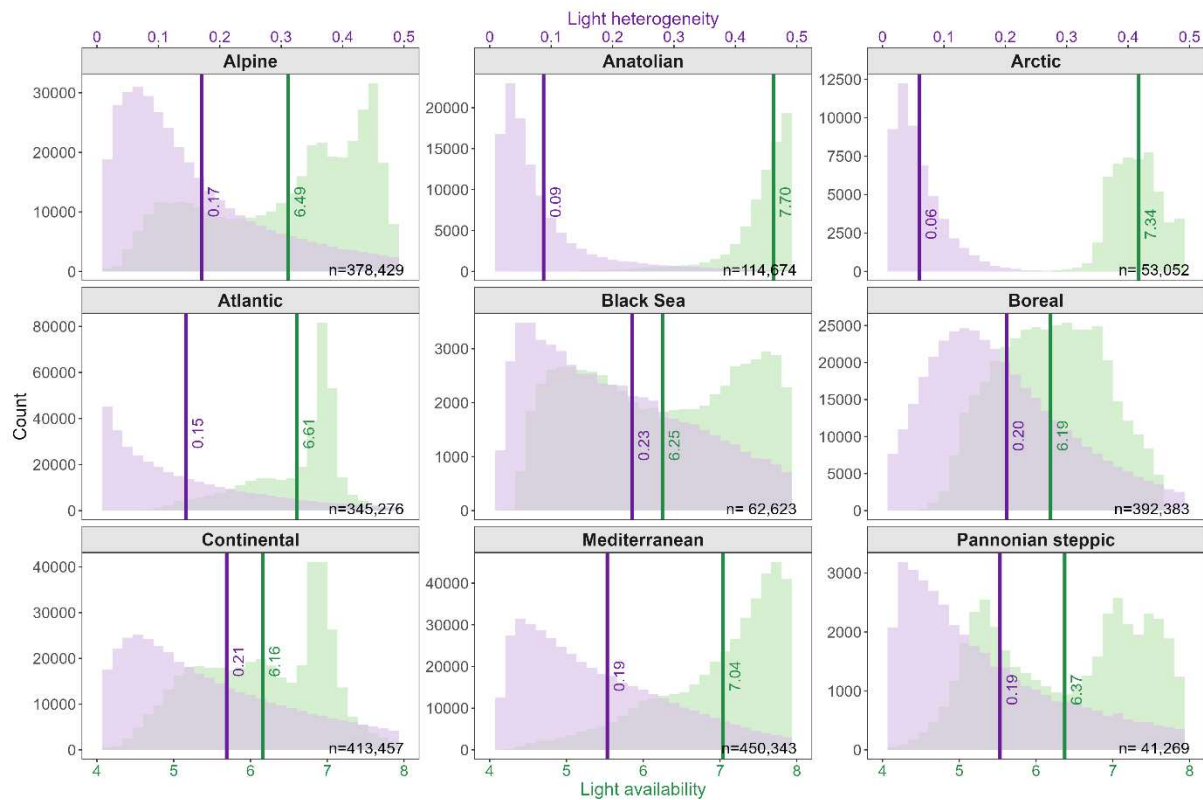
828 **Conflict of interests**

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

830

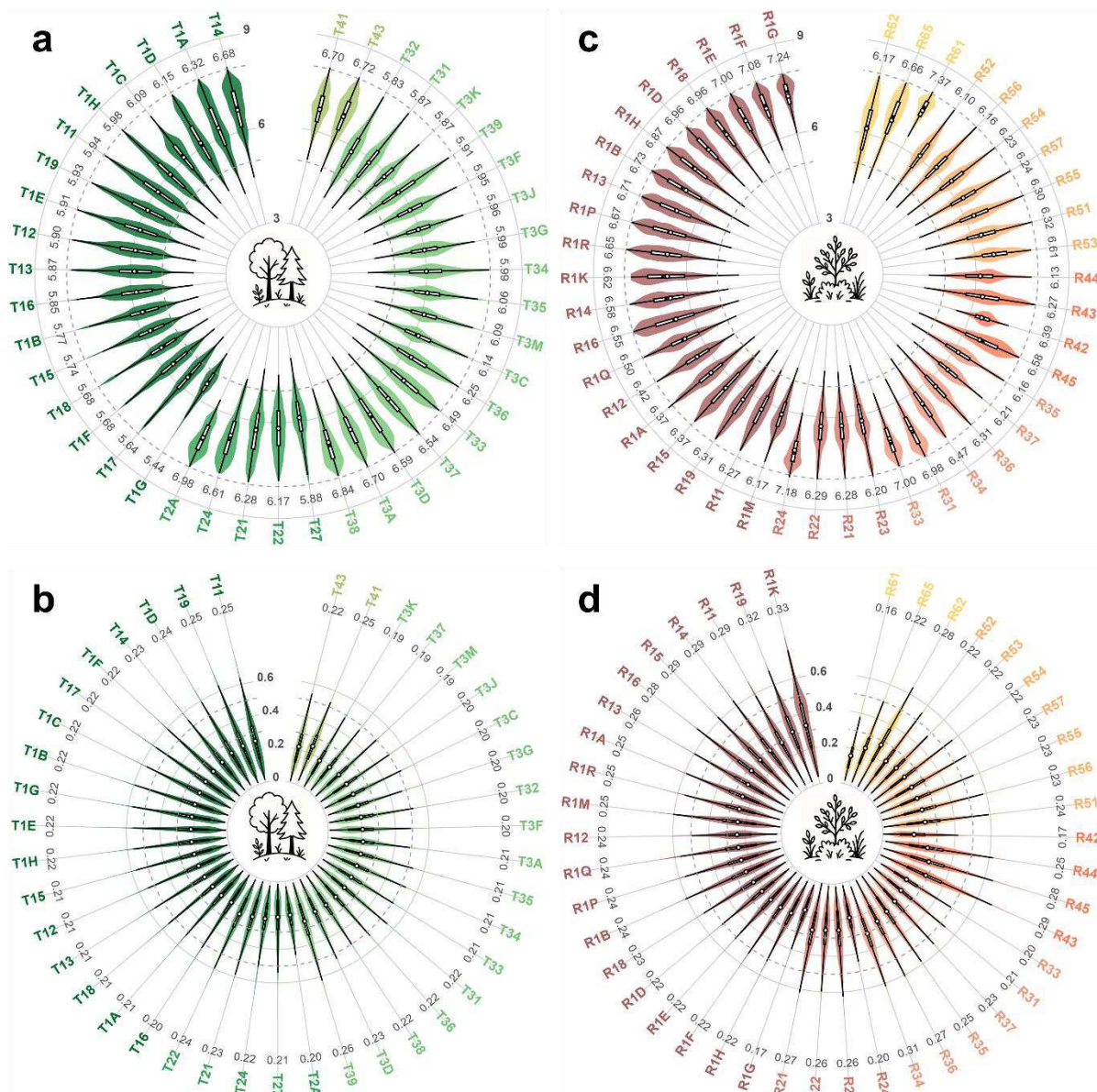


Extended Data Fig. 1 | Spatial patterns and landscape-scale relationships of light regimes across Europe. *a*, Spatial distribution of mean light availability aggregated to 10 km grid cells. *b*, Spatial distribution of light heterogeneity, computed as the standard deviation of 10 m light availability within each 10 km grid cell. *c–d*, Mean light availability (*c*) and heterogeneity (*d*) across climatological gradients of annual mean temperature and precipitation. Relationship between mean light availability and light heterogeneity across European landscapes. Hexagonal binning shows the density of observations in the two-dimensional light-regime space, with darker colours indicating higher point density. The solid line represents the fitted Generalized Additive Model (GAM) describing the smooth relationship between mean light availability and light heterogeneity. Shaded ribbons indicate the 95% confidence interval of the GAM smooth, reflecting uncertainty in the estimated mean relationship. Model fit statistics (R^2 and p value) are shown in the upper-right corner of the panel. Both metrics were derived from the 10 m light-availability map after masking non-natural land-cover types using the ESA WorldCover 2020 dataset²⁶ (excluding cropland [40], built-up areas [50], bare/sparse vegetation [60], snow and ice [70], and permanent water bodies [80]).



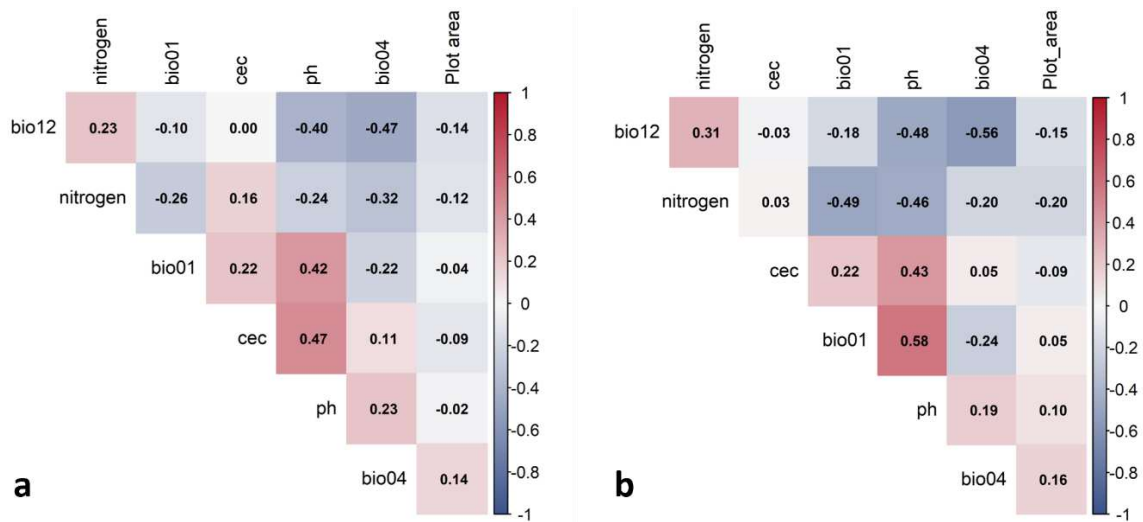
Extended Data Fig. 2 | Light regime patterns across European biogeographic regions.

Distributions of light availability (green) and light heterogeneity (purple) across nine biogeographic regions, with vertical lines marking regional means. Light-regime metrics were derived from 10 m light-availability maps; heterogeneity represents the standard deviation within a 50 m radius moving window. The y-axis (Count) represents the number of 10 m pixels falling within each bin. Only natural and semi-natural areas were included. Sample sizes (n) denote the number of 10 m pixels randomly selected within each region based on the European biogeographic map⁷³.



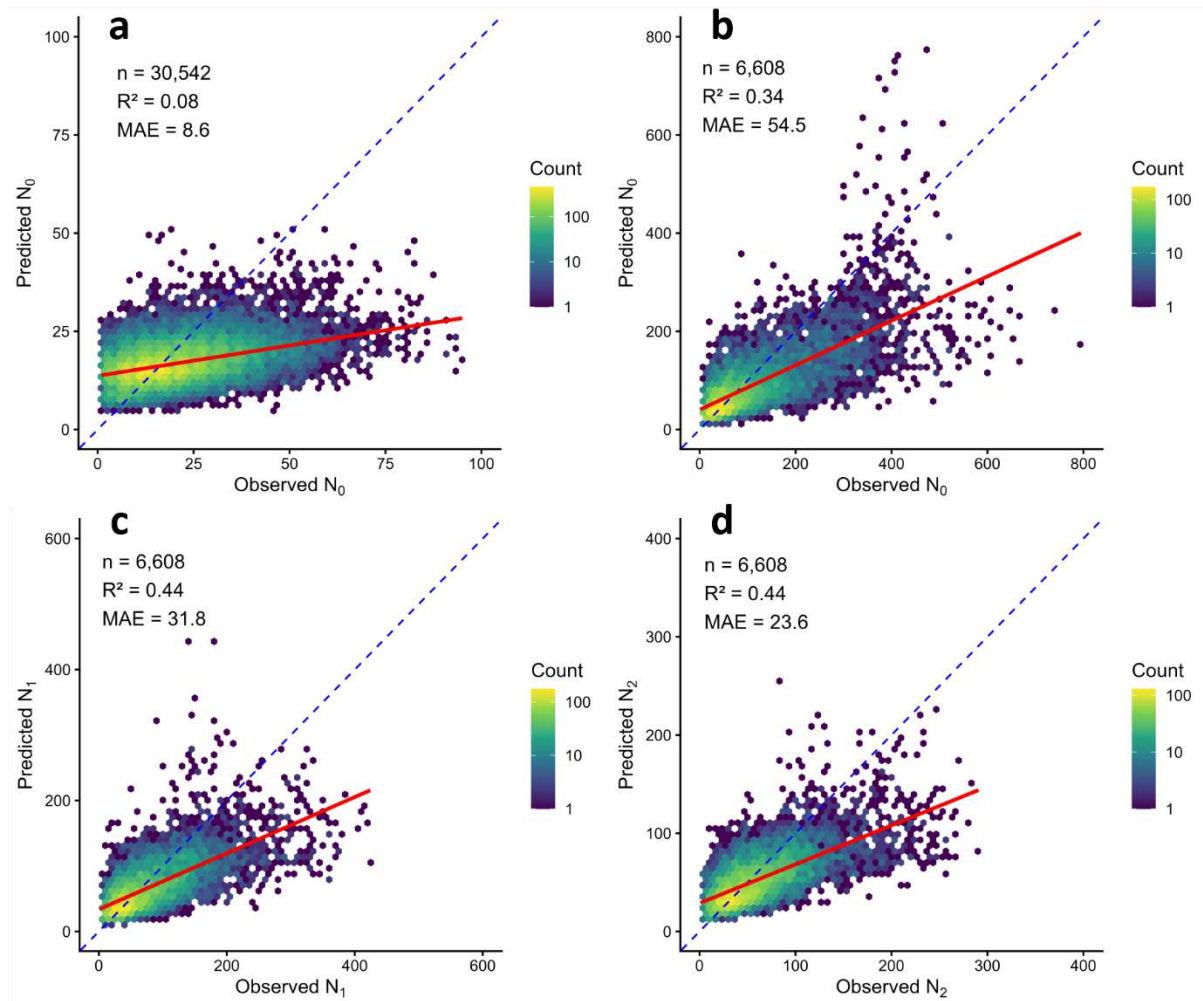
Extended Data Fig. 3 | Light regime patterns across forest (T) and grassland (R) habitat types (EUNIS level 3). Panels a–b show, respectively, light availability and light heterogeneity across forest habitat types, while c–d present the corresponding patterns for grassland habitats. White points indicate habitat-level averages. Light-regime variables were derived from 10 m resolution light-availability maps; heterogeneity was quantified as the standard deviation within a 50 m radius moving window. Analyses were restricted to natural and semi-natural land-cover classes. Random samples were drawn within each habitat type using the European EUNIS habitat map. Forest habitats (T) include **T1 – Broadleaved deciduous forests**, comprising T11 – Temperate *Salix* and *Populus* riparian forest, T12 – *Alnus glutinosa*–*Alnus incana* forest on riparian and mineral soils, T13 – Temperate hardwood riparian forest, T14 – Mediterranean and Macaronesian riparian forest, T15 – Broadleaved swamp forest on non-acid peat, T16 – Broadleaved mire forest on acid peat, T17 – *Fagus* forest on non-acid soils, T18 – *Fagus* forest on acid soils, T19 – Temperate and submediterranean thermophilous deciduous forest, T1A – Mediterranean thermophilous deciduous forest, T1B – Acidophilous *Quercus* forest, T1C – Temperate and boreal mountain *Betula* and *Populus tremula* forest on mineral soils, T1D – Southern European mountain *Betula* and *Populus tremula* forest on mineral soils, T1E – *Carpinus* and *Quercus* mesic deciduous forest, T1F – Ravine forest,

869 *T1G* – *Alnus cordata* forest, and *T1H* – Broadleaved deciduous plantation of non site-native trees. Forest habitats
 870 further include **T2 – Broadleaved evergreen forests**, consisting of *T21* – Mediterranean evergreen *Quercus* forest,
 871 *T22* – Mainland laurophyllous forest, *T24* – *Olea europaea*–*Ceratonia siliqua* forest, *T27* – *Ilex aquifolium* forest,
 872 and *T2A* – Broadleaved evergreen plantation of site-native trees; **T3 – Coniferous forests**, including *T31* –
 873 Temperate mountain *Picea* forest, *T32* – Temperate mountain *Abies* forest, *T33* – Mediterranean mountain *Abies*
 874 forest, *T34* – Temperate subalpine *Larix*, *Pinus cembra* and *Pinus uncinata* forest, *T35* – Temperate continental
 875 *Pinus sylvestris* forest, *T36* – Temperate and submediterranean montane *Pinus sylvestris*–*Pinus nigra* forest, *T37*
 876 – Mediterranean montane *Pinus sylvestris*–*Pinus nigra* forest, *T38* – Mediterranean montane *Cedrus* forest, *T39*
 877 – Mediterranean and Balkan subalpine *Pinus heldreichii*–*Pinus peuce* forest, *T3A* – Mediterranean lowland to
 878 submontane *Pinus* forest, *T3C* – *Taxus baccata* forest, *T3D* – Mediterranean Cupressaceae forest, *T3F* – Dark
 879 taiga, *T3G* – *Pinus sylvestris* light taiga, *T3J* – *Pinus* and *Larix* mire forest, and *T3K* – *Picea* mire forest; and **T4**
 880 – **Lines of trees, small anthropogenic forests, recently felled forest, early-stage forest and coppice**, including
 881 *T41* – Early-stage natural and semi-natural forest and regrowth and *T43* – Recently felled area. Grassland and
 882 open habitats (R) include **R1 – Dry grasslands**, comprising *R11* – Pannonian and Pontic sandy steppe, *R12* –
 883 Cryptogam- and annual-dominated vegetation on siliceous rock outcrops, *R13* – Cryptogam- and annual-
 884 dominated vegetation on calcareous and ultramafic rock outcrops, *R14* – Perennial rocky grassland of the Italian
 885 Peninsula, *R15* – Continental dry rocky steppic grassland and dwarf scrub on chalk outcrops, *R16* – Perennial
 886 rocky grassland of Central and South-Eastern Europe, *R18* – Perennial rocky calcareous grassland of
 887 subatlantic-submediterranean Europe, *R19* – Dry steppic submediterranean pasture of the Amphi-Adriatic region,
 888 *R1A* – Semi-dry perennial calcareous grassland (meadow steppe), *R1B* – Continental dry grassland (true steppe),
 889 *R1D* – Mediterranean closely grazed dry grassland, *R1E* – Mediterranean tall perennial dry grassland, *R1F* –
 890 Mediterranean annual-rich dry grassland, *R1G* – Iberian oromediterranean siliceous dry grassland, *R1H* –
 891 Iberian oromediterranean basiphilous dry grassland, *R1K* – Balkan and Anatolian oromediterranean dry
 892 grassland, *R1M* – Lowland to montane, dry to mesic grassland usually dominated by *Nardus stricta*, *R1P* –
 893 Oceanic to subcontinental inland sand grassland on dry acid and neutral soils, *R1Q* – Inland sand drift and dune
 894 with siliceous grassland, and *R1R* – Mediterranean to Atlantic open, dry, acid and neutral grassland. Additional
 895 open habitats include **R2 – Mesic grasslands**, consisting of *R21* – Mesic permanent pasture of lowlands and
 896 mountains, *R22* – Low and medium altitude hay meadow, *R23* – Mountain hay meadow, and *R24* – Iberian summer
 897 pasture (vallicar); **R3 – Seasonally wet and wet grasslands**, including *R31* – Mediterranean tall humid inland
 898 grassland, *R33* – Mediterranean short moist grassland of mountains, *R34* – Submediterranean moist meadow,
 899 *R35* – Moist or wet mesotrophic to eutrophic hay meadow, *R36* – Moist or wet mesotrophic to eutrophic pasture,
 900 and *R37* – Temperate and boreal moist or wet oligotrophic grassland; **R4 – Alpine and subalpine grasslands**,
 901 including *R42* – Boreal and Arctic acidophilous alpine grassland, *R43* – Temperate acidophilous alpine grassland,
 902 *R44* – Arctic-alpine calcareous grassland, and *R45* – Alpine and subalpine calcareous grassland of the Balkans
 903 and Apennines; **R5 – Woodland fringes and clearings and tall forb stands**, including *R51* – Thermophilous forest
 904 fringe of base-rich soils, *R52* – Forest fringe of acidic nutrient-poor soils, *R53* – Macaronesian thermophilous
 905 forest fringe, *R54* – *Pteridium aquilinum* vegetation, *R55* – Lowland moist or wet tall-herb and fern fringe, *R56*
 906 – Montane to subalpine moist or wet tall-herb and fern fringe, and *R57* – Herbaceous forest clearing vegetation;
 907 and **R6 – Inland salt steppes**, including *R61* – Mediterranean inland salt steppe, *R62* – Continental inland salt
 908 steppe, and *R65* – Continental subsaline alluvial pasture and meadow.



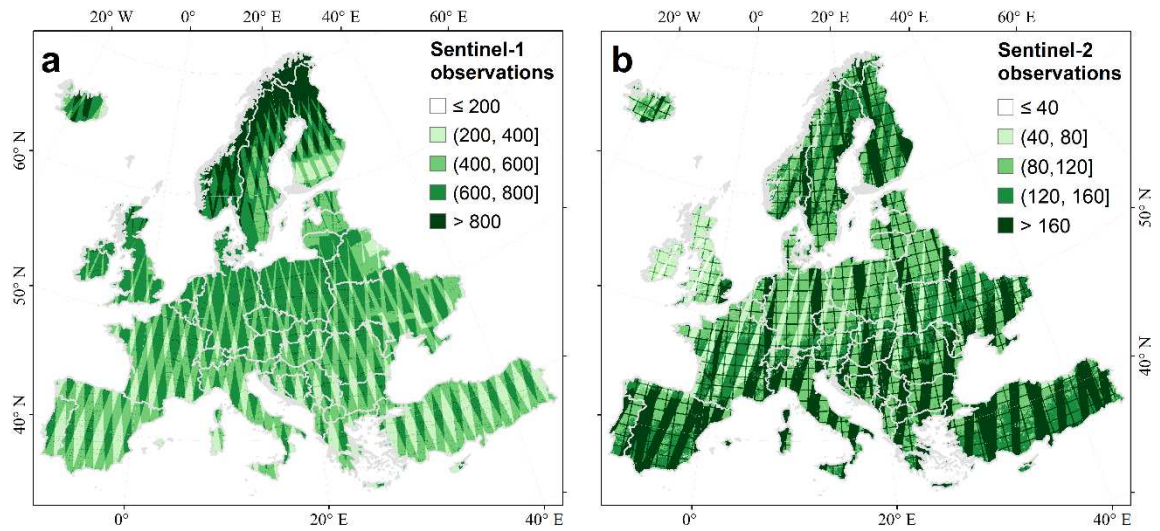
Extended Data Fig. 4 | Correlation matrix of predictor variables included in the plant diversity models.

Pearson correlation coefficients among environmental predictors used in the final plant diversity models at the local scale (a) and landscape scale (b). Numbers within cells indicate pairwise correlation coefficients (r), with red and blue colours representing positive and negative correlations, respectively. Overall, correlations among predictors were moderate ($|r| < 0.6$), indicating low risk of severe multicollinearity. Variable abbreviations are as follows: bio01, annual mean temperature; bio04, temperature seasonality; bio12, annual precipitation; CEC, soil cation exchange capacity; nitrogen, soil nitrogen content; pH, soil pH; and Plot area, vegetation plot area.



Extended Data Fig. 5 | Cross-validated observed versus predicted plant diversity across spatial scales.

Hexbin plots show the relationship between observed and predicted diversity values from five-fold spatial cross-validation. Colours indicate the number of observations per hexagonal bin on a logarithmic scale. The blue dashed line represents the 1:1 relationship, and the red solid line shows the fitted linear regression. Panel **a** shows local-scale species richness (N_0), whereas panels **b–d** show landscape-scale diversity metrics, including species richness (N_0 ; **b**), the exponential of Shannon entropy (N_1 ; **c**), and the inverse Simpson index (N_2 ; **d**). Model performance statistics (sample size, coefficient of determination, and mean absolute error) are reported in each panel.



Extended Data Fig. 6 | Spatial coverage of satellite data used. a, Number of Sentinel-1 observations acquired between 2019 and 2021 and used as inputs in the light-availability modelling framework. **b,** Number of clear-sky Sentinel-2 observations from 2019 to 2021 contributing spectral information for light-availability estimation.

Extended Data Table 1 | Earth observation predictors used to reconstruct continental-scale light availability across Europe.

Predictor variables included radar backscatter metrics from Sentinel-1 and PALSAR, spectral bands and vegetation indices derived from Sentinel-2 imagery, topographic variables, and geographic coordinates. Sentinel-1 and Sentinel-2 metrics were derived from observations acquired between 2019 and 2021. Abbreviations: VV and VH, vertical transmit–vertical receive and vertical transmit–horizontal receive polarization; HH and HV, horizontal transmit–horizontal receive and horizontal transmit–vertical receive polarization; NDVI, Normalized Difference Vegetation Index; EVI, Enhanced Vegetation Index; kNDVI, kernel Normalized Difference Vegetation Index; LSWI, Land Surface Water Index; MNDWI, Modified Normalized Difference Water Index; GCVI, Green Chlorophyll Vegetation Index; NMDI, Normalized Multi-band Drought Index; CSI, Combinational Shadow Index; BI, Brightness Index; BSI, Bare Soil Index; SATVI, Soil-Adjusted Total Vegetation Index; TPI, Topographic Position Index. Resolution refers to the native spatial resolution of each dataset.

Code	Products	Variables	Acquisition time	Resolution	Data Source
1-3	Sentinel-1	VV; VH; VVH	2019-2021	10 m	COPERNICUS/S1_GRD
4-5	PALSAR	HH; HV	2019-2021	25 m	JAXA/ALOS/PALSAR/YEARLY/SAR
6-27	Sentinel-2	Aerosols; Blue; Green; Red; NIR; Red Edge 1–4; SWIR 1; SWIR 2; NDVI; EVI; kNDVI; LSWI; MNDWI; GCVI; NMDI; CSI; BI; BSI; SATVI	2019-2021	10/20/60 m	COPERNICUS/S2_SR_HARMONIZED
28-33	Topography	Elevation; Slope; Aspect; TPI _{150m} ; TPI _{300m} ; TPI _{1500m}	/	30 m	projects/sat-io/open-datasets/GLO-30
33-35	Location	Latitude; Longitude	/	10 m	/

Extended Data Table 2 | Moran's I statistics for residual spatial autocorrelation in the final interaction models.

Residual spatial autocorrelation was evaluated using Moran's I calculated from Pearson residuals of the final interaction models for local-scale species richness (N_0) and landscape-scale diversity metrics (N_0 , N_1 and N_2). Spatial neighbourhoods were defined using an 8-nearest-neighbour graph with row-standardized spatial weights. Moran's I values close to zero indicate that most spatial dependence was accounted for by the fitted spatial random field.

Scale	Diversity metric	Moran's I	P-value
Local	N_0	-0.04	1
	N_0	0.009	0.052
Landscape	N_1	0.0079	0.084
	N_2	0.0095	0.0504

954 **References**

- 955 1 Ruban, A. V. Plants in light. *Communicative & Integrative Biology* **2**, 50–55 (2009).
- 956 2 Grunwald, Y., Yaaran, A. & Moshelion, M. Illuminating plant water dynamics: the role of
957 light in leaf hydraulic regulation. *New Phytologist* **241**, 1404–1414 (2024).
- 958 3 Austin, A. T. & Vivanco, L. Plant litter decomposition in a semi-arid ecosystem controlled by
959 photodegradation. *Nature* **442**, 555–558 (2006).
- 960 4 Helbach, J., Frey, J., Messier, C., Mörsdorf, M. & Scherer-Lorenzen, M. Light heterogeneity
961 affects understory plant species richness in temperate forests supporting the heterogeneity–
962 diversity hypothesis. *Ecology and Evolution* **12**, e8534 (2022).
- 963 5 Nicotra, A. B., Chazdon, R. L. & Iriarte, S. V. Spatial heterogeneity of light and woody
964 seedling regeneration in tropical wet forests. *Ecology* **80**, 1908–1926 (1999).
- 965 6 Silvertown, J. Plant coexistence and the niche. *Trends in Ecology & evolution* **19**, 605–611
966 (2004).
- 967 7 Storch, D., Bohdalková, E. & Okie, J. The more-individuals hypothesis revisited: the role of
968 community abundance in species richness regulation and the productivity–diversity
969 relationship. *Ecology letters* **21**, 920–937 (2018).
- 970 8 Chesson, P. Mechanisms of maintenance of species diversity. *Annual Review of Ecology and*
971 *Systematics* **31**, 343–366 (2000).
- 972 9 Dormann, C. F. *et al.* Plant species richness increases with light availability, but not
973 variability, in temperate forests understorey. *BMC Ecology* **20**, 43 (2020).
- 974 10 Márialigeti, S., Tinya, F., Bidló, A. & Ódor, P. Environmental drivers of the composition and
975 diversity of the herb layer in mixed temperate forests in Hungary. *Plant Ecology* **217**, 549–
976 563 (2016).
- 977 11 Bartels, S. F. & Chen, H. Y. Interactions between overstorey and understorey vegetation
978 along an overstorey compositional gradient. *Journal of Vegetation Science* **24**, 543–552
979 (2013).
- 980 12 Fuxai, X., Fousseni, F., Chungang, P., Huaijiang, H. & Xiuhai, Z. Effect of overstorey on the
981 seasonal variability of understory herbs in primary broad-leaved Korean pine forest of
982 Changbai Mountain. *African Journal of Biotechnology* **13** (2014).
- 983 13 Vega-Álvarez, J., García-Rodríguez, J. A. & Cayuela, L. Facilitation beyond species richness.
984 *Journal of Ecology* **107**, 722–734 (2019).
- 985 14 Nieto-Lugilde, D. *et al.* Tree cover at fine and coarse spatial grains interacts with shade
986 tolerance to shape plant species distributions across the Alps. *Ecography* **38**, 578–589 (2015).
- 987 15 Czyżewski, S. & Svenning, J.-C. Temperate forest plants are associated with heterogeneous
988 semi-open canopy conditions shaped by large herbivores. *Nature Plants*, 1–16 (2025).
- 989 16 Pearce, E. A. *et al.* Substantial light woodland and open vegetation characterized the
990 temperate forest biome before *Homo sapiens*. *Science Advances* **9**, eadi9135 (2023).
- 991 17 Davoli, M. *et al.* Megafauna diversity and functional declines in Europe from the Last
992 Interglacial to the present. *Global Ecology and Biogeography* **33**, 34–47 (2024).
- 993 18 Svenning, J.-C. *et al.* The late-Quaternary megafauna extinctions: Patterns, causes, ecological
994 consequences and implications for ecosystem management in the Anthropocene. *Cambridge*
995 *Prisms: Extinction* **2**, e5 (2024).
- 996 19 McGrath, M. J. *et al.* Reconstructing European forest management from 1600 to 2010.
997 *Biogeosciences* **12**, 4291–4316 (2015).
- 998 20 Ustaoglu, E. & Collier, M. J. Farmland abandonment in Europe: An overview of drivers,
999 consequences, and assessment of the sustainability implications. *Environmental Reviews* **26**,
1000 396–416 (2018).
- 1001 21 Montgomery, R. & Chazdon, R. Light gradient partitioning by tropical tree seedlings in the
1002 absence of canopy gaps. *Oecologia* **131**, 165–174 (2002).
- 1003 22 Webster, C., Mazzotti, G., Essery, R. & Jonas, T. Enhancing airborne LiDAR data for
1004 improved forest structure representation in shortwave transmission models. *Remote Sensing of*
1005 *Environment* **249**, 112017 (2020).

1006 23 Alexander, C., Moeslund, J. E., Bøcher, P. K., Arge, L. & Svenning, J.-C. Airborne laser
1007 scanner (LiDAR) proxies for understory light conditions. *Remote Sensing of Environment*
1008 **134**, 152–161 (2013).

1009 24 Dengler, J. *et al.* Ecological Indicator Values for Europe (EIVE) 1.0. *Vegetation*
1010 *Classification and Survey* **4**, 7–29 (2023).

1011 25 Chytrý, M. *et al.* European Vegetation Archive (EVA): an integrated database of European
1012 vegetation plots. *Applied Vegetation Science* **19**, 173–180 (2016).

1013 26 Zanaga, D. *et al.* ESA WorldCover 10 m 2021 v200. (2022).

1014 27 Si-Moussi, S. *et al.* EUNIS habitat maps: enhancing thematic and spatial resolution for
1015 Europe through machine learning. *Scientific Data* **12**, 1976 (2025).

1016 28 Anderson, S. C., Ward, E. J., English, P. A., Barnett, L. A. & Thorson, J. T. sdmTMB: an R
1017 package for fast, flexible, and user-friendly generalized linear mixed effects models with
1018 spatial and spatiotemporal random fields. *Journal of Statistical Software* **115**, 1–46 (2025).

1019 29 Valladares, F. & Niinemets, Ü. Shade tolerance, a key plant feature of complex nature and
1020 consequences. *Annual Review of Ecology, Evolution, and Systematics* **39**, 237–257 (2008).

1021 30 Stein, A., Gerstner, K. & Kreft, H. Environmental heterogeneity as a universal driver of
1022 species richness across taxa, biomes and spatial scales. *Ecology letters* **17**, 866–880 (2014).

1023 31 Wright, D. H. Species-energy theory: an extension of species-area theory. *Oikos*, 496–506
1024 (1983).

1025 32 Levine, J. M. & HilleRisLambers, J. The importance of niches for the maintenance of species
1026 diversity. *Nature* **461**, 254–257 (2009).

1027 33 Sandom, C. J., Ejrnæs, R., Hansen, M. D. & Svenning, J.-C. High herbivore density
1028 associated with vegetation diversity in interglacial ecosystems. *Proceedings of the National*
1029 *Academy of Sciences* **111**, 4162–4167 (2014).

1030 34 Ricklefs, R. E. Environmental heterogeneity and plant species diversity: a hypothesis. *The*
1031 *American Naturalist* **111**, 376–381 (1977).

1032 35 Leibold, M. A. *et al.* The metacommunity concept: a framework for multi-scale community
1033 ecology. *Ecology letters* **7**, 601–613 (2004).

1034 36 Czyżewski, S. *et al.* Revisiting Europe's temperate forests: Palaeoecological evidence for an
1035 herbivory-driven woodland-grassland mosaic biome. *Biological Conservation* **316**, 111749
1036 (2026).

1037 37 Czyżewski, S. & Svenning, J. (2025).

1038 38 Zobel, M. The species pool concept as a framework for studying patterns of plant diversity.
1039 *Journal of Vegetation Science* **27**, 8–18 (2016).

1040 39 Buitenwerf, R., Sandel, B., Normand, S., Mimet, A. & Svenning, J. C. Land surface greening
1041 suggests vigorous woody regrowth throughout European semi-natural vegetation. *Global*
1042 *Change Biology* **24**, 5789–5801 (2018).

1043 40 Estel, S. *et al.* Mapping farmland abandonment and recultivation across Europe using MODIS
1044 NDVI time series. *Remote Sensing of Environment* **163**, 312–325 (2015).

1045 41 Bakker, E. S. *et al.* Combining paleo-data and modern exclosure experiments to assess the
1046 impact of megafauna extinctions on woody vegetation. *Proceedings of the National Academy*
1047 *of Sciences* **113**, 847–855 (2016).

1048 42 Tälle, M. *et al.* Grazing vs. mowing: A meta-analysis of biodiversity benefits for grassland
1049 management. *Agriculture, Ecosystems & Environment* **222**, 200–212 (2016).

1050 43 Perino, A. *et al.* Rewilding complex ecosystems. *Science* **364**, eaav5570 (2019).

1051 44 Hédli, R., Šipoš, J., Chudomelová, M. & Utinek, D. Dynamics of herbaceous vegetation
1052 during four years of experimental coppice introduction. *Folia Geobotanica* **52**, 83–99 (2017).

1053 45 Bassett, T. J., Landis, D. A. & Brudvig, L. A. Effects of experimental prescribed fire and tree
1054 thinning on oak savanna understory plant communities and ecosystem structure. *Forest*
1055 *Ecology and Management* **464**, 118047 (2020).

1056 46 Euro+ Med, E. Euro+ Med PlantBase—the information resource for Euro-Mediterranean plant
1057 diversity. Available online at: <http://www.europlusmed.org> (2026).

1058 47 Ostrowski, G. *et al.* Mean ecological indicator values: use EIVE but no cover-weighting.
1059 *Vegetation Classification and Survey* **6**, 57–67 (2025).

1060 48 Gorelick, N. *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone.
1061 *Remote Sensing of Environment* **202**, 18–27 (2017).

1062 49 Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation.
1063 *Remote Sensing of Environment* **8**, 127–150 (1979).

1064 50 Huete, A. *et al.* Overview of the radiometric and biophysical performance of the MODIS
1065 vegetation indices. *Remote Sensing of Environment* **83**, 195–213 (2002).

1066 51 Camps-Valls, G. *et al.* A unified vegetation index for quantifying the terrestrial biosphere.
1067 *Science Advances* **7**, eabc7447 (2021).

1068 52 Xiao, X. *et al.* Satellite-based modeling of gross primary production in an evergreen
1069 needleleaf forest. *Remote Sensing of Environment* **89**, 519–534 (2004).

1070 53 Xu, H. Modification of normalised difference water index (NDWI) to enhance open water
1071 features in remotely sensed imagery. *International Journal of Remote Sensing* **27**, 3025–3033
1072 (2006).

1073 54 Vincini, M., Frazzi, E. & D’Alessio, P. A broad-band leaf chlorophyll vegetation index at the
1074 canopy scale. *Precision Agriculture* **9**, 303–319 (2008).

1075 55 Wang, L. & Qu, J. J. NMDI: A normalized multi-band drought index for monitoring soil and
1076 vegetation moisture with satellite remote sensing. *Geophysical Research Letters* **34** (2007).

1077 56 Sun, G. *et al.* Combinational shadow index for building shadow extraction in urban areas
1078 from Sentinel-2A MSI imagery. *International Journal of Applied Earth Observation and*
1079 *Geoinformation* **78**, 53–65 (2019).

1080 57 Roy, P., Sharma, K. & Jain, A. Stratification of density in dry deciduous forest using satellite
1081 remote sensing digital data—An approach based on spectral indices. *Journal of Biosciences*
1082 **21**, 723–734 (1996).

1083 58 Huete, A. R. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* **25**,
1084 295–309 (1988).

1085 59 Rosenqvist, A., Shimada, M., Ito, N. & Watanabe, M. ALOS PALSAR: A pathfinder mission
1086 for global-scale monitoring of the environment. *IEEE Transactions on Geoscience and*
1087 *Remote Sensing* **45**, 3307–3316 (2007).

1088 60 Simard, M., Denbina, M., Marshak, C. & Neumann, M. A global evaluation of radar-derived
1089 digital elevation models: SRTM, NASADEM, and GLO-30. *Journal of Geophysical*
1090 *Research: Biogeosciences* **129**, e2023JG007672 (2024).

1091 61 De Reu, J. *et al.* Application of the topographic position index to heterogeneous landscapes.
1092 *Geomorphology* **186**, 39–49 (2013).

1093 62 Scherpenhuijzen, N. *et al.* Mapping forest management regimes in Europe. *Forest Ecology*
1094 *and Management* **594**, 122940 (2025).

1095 63 Sabatini, F. M. *et al.* European primary forest database v2. 0. *Scientific Data* **8**, 220 (2021).

1096 64 State of Europe’s forests 2020. (Forest Europe, 2020).

1097 65 Wu, W.-B. *et al.* A first Chinese building height estimate at 10 m resolution (CNBH-10 m)
1098 using multi-source earth observations and machine learning. *Remote Sensing of Environment*
1099 **291**, 113578 (2023).

1100 66 Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for
1101 global land areas. *International Journal of Climatology* **37**, 4302–4315 (2017).

1102 67 Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning.
1103 *PLoS One* **12**, e0169748 (2017).

1104 68 Hill, M. O. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**,
1105 427–432 (1973).

1106 69 Chao, A. *et al.* Rarefaction and extrapolation with Hill numbers: a framework for sampling
1107 and estimation in species diversity studies. *Ecological Monographs* **84**, 45–67 (2014).

1108 70 Colwell, R. K. *et al.* Models and estimators linking individual-based and sample-based
1109 rarefaction, extrapolation and comparison of assemblages. *Journal of Plant Ecology* **5**, 3–21
1110 (2012).

1111 71 Breiman, L. Random forests. *Machine Learning* **45**, 5–32 (2001).

1112 72 Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillera-Aroita, G. blockCV: An r package for
1113 generating spatially or environmentally separated folds for k-fold cross-validation of species
1114 distribution models. *Biorxiv*, 357798 (2018).

1115 73 Roekaerts, M. The biogeographical regions map of Europe. *Basic Principles of its Creation*
1116 *and Overview of its Development* **17** (EEA, 2002).
1117

Supplementary Files

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

- [ExtendedData.docx](#)