Supplementary Information

Anthropogenic Land-Use Legacies Underpin Climate Change-Related Risks to Forest Ecosystems

Albert Vilà-Cabrera, Julen Astigarraga, Alistair S. Jump, Miguel A. Zavala, Francisco Seijo, Dominik Sperlich and Paloma Ruiz-Benito

The case study of Scots pine (*Pinus sylvestris* L.)

Spanish forest inventory and climate data

We used data from the second (1986–1996) and third (1997–2007) Spanish Forest Inventory (SFI), which sampled plots systematically distributed on a 1 km^2 grid across forests (i.e., forest cover greater than 5%, Villaescusa & Díaz, 1998). The SFI plots have a variable radius design with 5, 10, 15 and 25 m radius subplots where adult trees with a diameter at breast height (d.b.h.) of 7.5–12.4, 12.5–22.4, 22.5–42.4 and >42.4 cm are measured, respectively. For each adult tree (i.e. trees with d.b.h. ≥ 7.5 and height > 130 cm, Villaescusa & Díaz, 1998) tree height, d.b.h., species identity, distance to the centre of the plot, the position of the tree and tree status (alive or dead present or absent) are recorded. For saplings (i.e. trees with 2.5 ≤ d.b.h. ≤ 7.5 cm) the number of individuals per species and their mean total height are measured in the 5 m radius subplot. Further information on the Spanish Forest Inventory comparing natural and planted stands of pine species can be found in Ruiz-Benito et al (2012) and to see the evolution of forest structure, demography and productivity over consecutive censuses see Astigarraga et al (2020).

We selected all trees in permanent stands between the second and third SFI census in which *Pinus sylvestris* had ≥ 50% of basal area (142340 *Pinus sylvestris* trees in 5418 stands, representing 14.7% of all trees) to calculate tree demography and link with their natural or planted origin. We calculated tree growth as the percentage of annual increase in basal area of living trees over consecutive inventories relative to initial basal area, and tree mortality as the percentage of basal area lost annually by dead adult trees over consecutive inventories relative to initial basal area (% ha^{-1} $year^{-1}$). We used mortality of present and absent trees between consecutive SFI censuses, but if there was evidence of cuts between censuses, we selected only dead trees present. We obtained the natural or planted origin of each stand using the Spanish Regions of Provenance for Forest Species (hereafter SRP, see Alía et al. 2005). This database determines the origin of the species (i.e., natural or planted during the 20th century) based on historical information (Ceballos 1966) and available phenotypic, genetic, and environmental information (Alía et al. 2009). Thus, joining the selected *Pinus sylvestris* stands with the SRP database, we determined the natural or planted origin of the stands (Ruiz-Benito, Gómez-Aparicio, and Zavala 2012) (96241 natural individuals located in 3966 stands and 46099 planted individuals located in 1452 stands).

We calculated forest structure using information from adult trees and saplings from the second and third SFI censuses. This allowed us to have the broadest possible structural representation of the analysed stands. We calculated stand-level tree density (No. trees ha^{-1}), basal area (m^2 ha^{-1}), mean d.b.h. (mm), mean height (m), coefficient of variation of d.b.h. (adimensional), coefficient of variation of height (adimensional), the ratio height:d.b.h. (adimensional), species richness (No. of species), and tree vigour (semiquantitative variable ranging from one, meaning healthy tree, to six, meaning dead standing but non-rotten tree, calculated as the mean value from the tree level quality index, see Moreno-Fernández et al. 2019). To have more comprehensive view of forest structure we also considered sapling density (No. saplings ha^{-1}) in the cluster analysis.

Using all forest structural variables, we performed a cluster analysis considering all planted and natural stands using a K-means cluster analyses, which is the most used clustering method for splitting a dataset into a set of k groups (MacQueen 1967), with the {kmeans} function from {stats} R package. The data was standardised with mean zero and standard deviation of one to make variables comparable. All the variables used were quantitative variables (tree density, stand basal area, mean d.b.h., coefficient of variation of d.b.h. and height, ratio height:d.b.h., species richness and sapling density). Tree vigour was the only semiquantitative variable with higher numbers indicating lower tree vigour. Using the cluster analyses, we obtained a variable of the forest structural typology to which each stand belonged (clusters thereafter are referred as forest structural typologies). We performed Wilcoxon tests ($p < 0.05$) to analyse differences in the structural variables between both forest structural typologies and observed that all variables between forest structural typologies were statistically significant (see Zenodo repository for further information on statistical test, https://doi.org/10.5281/zenodo.7801925). Finally, we obtained the variable forest type by combining forest structural typology and stand origin (natural or planted origin).

We quantified temporal and spatial climatic variability. We used SPEI as a proxy of temporal availability of water, and water availability index as a proxy of spatial availability of water (see e.g., Ruiz-Benito et al. 2017). Drought conditions were characterised using 12-month SPEI for the period between the inventory censuses (adimensional) with a 0.5 degrees spatial resolution from SPEIbase v2.5 (Vicente-Serrano, Beguería, and López-Moreno 2010) (Fig. S5a). Water availability index was calculated as the difference between annual precipitation and potential evapotranspiration with respect to potential evapotranspiration (i.e. (annual precipitation – PET) / PET, %) for the period 1970-2000 with a 1 km^2 spatial resolution using data from WorldClim 2 (Fick and Hijmans 2017) and Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database v2 (Trabucco and Zomer 2019) (Fig. S5b).

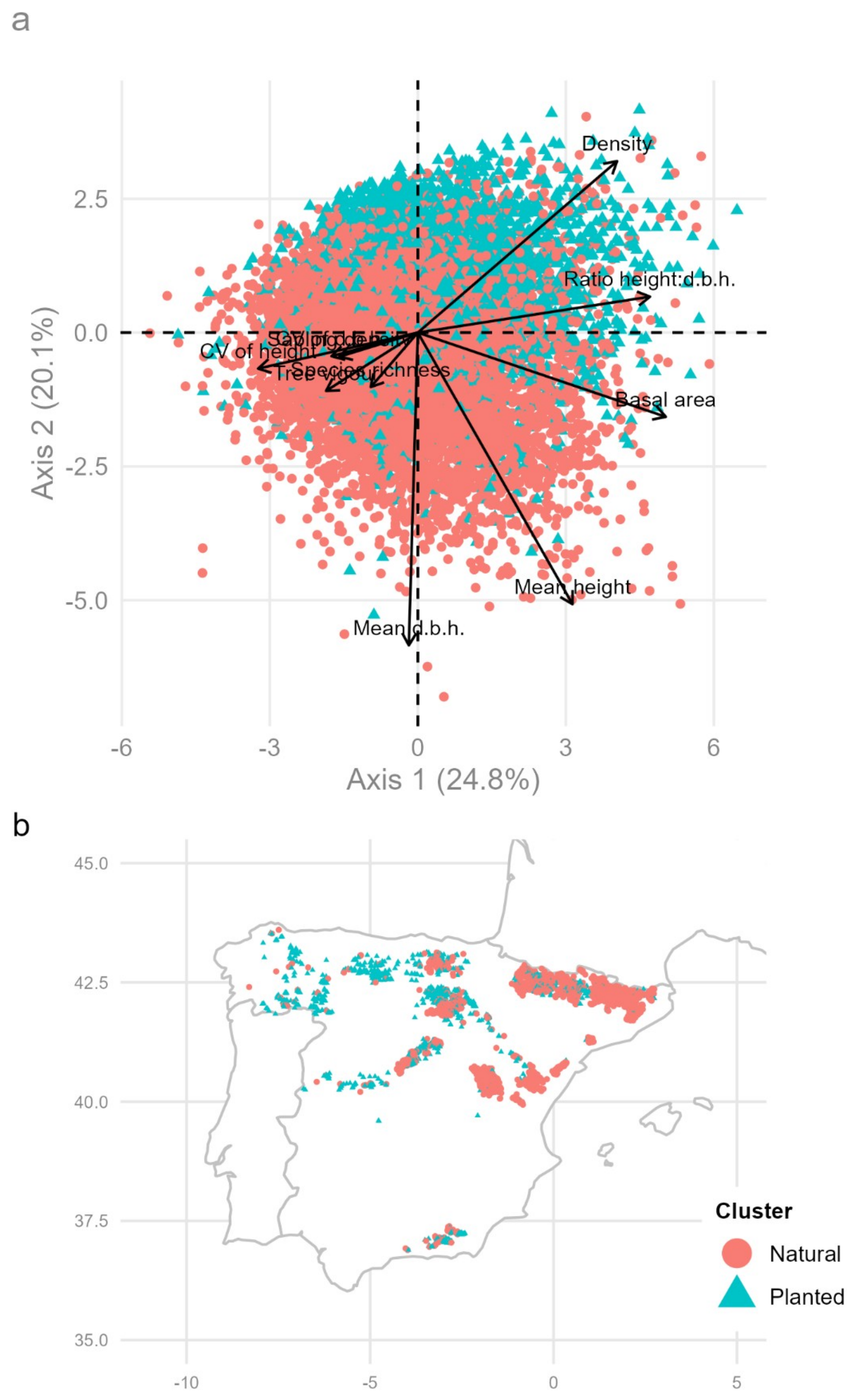

Figure S1. (a) Distribution of inventory plots (coloured according to stand origin) over the two first axes of the cluster analyses for forest structural variables (tree density, stand basal area, mean d.b.h., coefficient of variation of d.b.h. and height, ratio height:d.b.h., species richness, tree vigour and sapling density). The contribution of each variable to each axis is shown by the direction and length of the arrows. (b) Map showing the geographic distribution of stand origin.

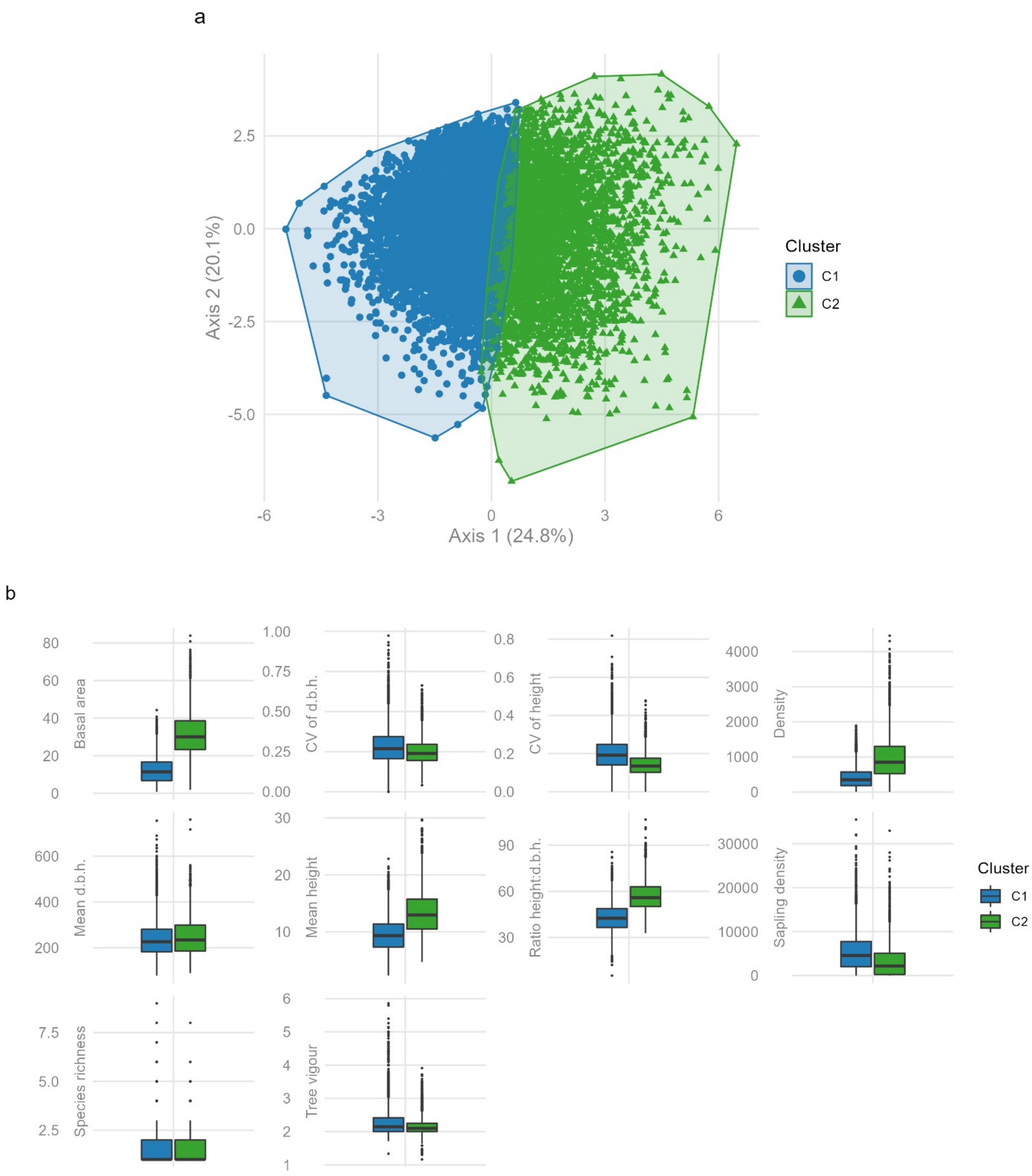

Figure S2. (a) Distribution of inventory plots (classified according to the two main forest structural typologies) over the two first dimensions of the cluster analyses for forest structural variables. (b) Boxplots showing the distribution of forest structural variables within each forest structural typology (basal area (m^2 ha^{-1}), coefficient of variation of d.b.h. (adimensional), coefficient of variation of height (adimensional), density (No. trees ha^{-1}), mean d.b.h. (mm), mean height (m), the ratio height:d.b.h. (adimensional), sapling density (No. saplings ha^{-1}), species richness (No. of species), and tree vigour (semiquantitative, see *Spanish forest inventory and climate data*)). (c) Geographic distribution of inventory plots classified according to their corresponding forest structural typology.

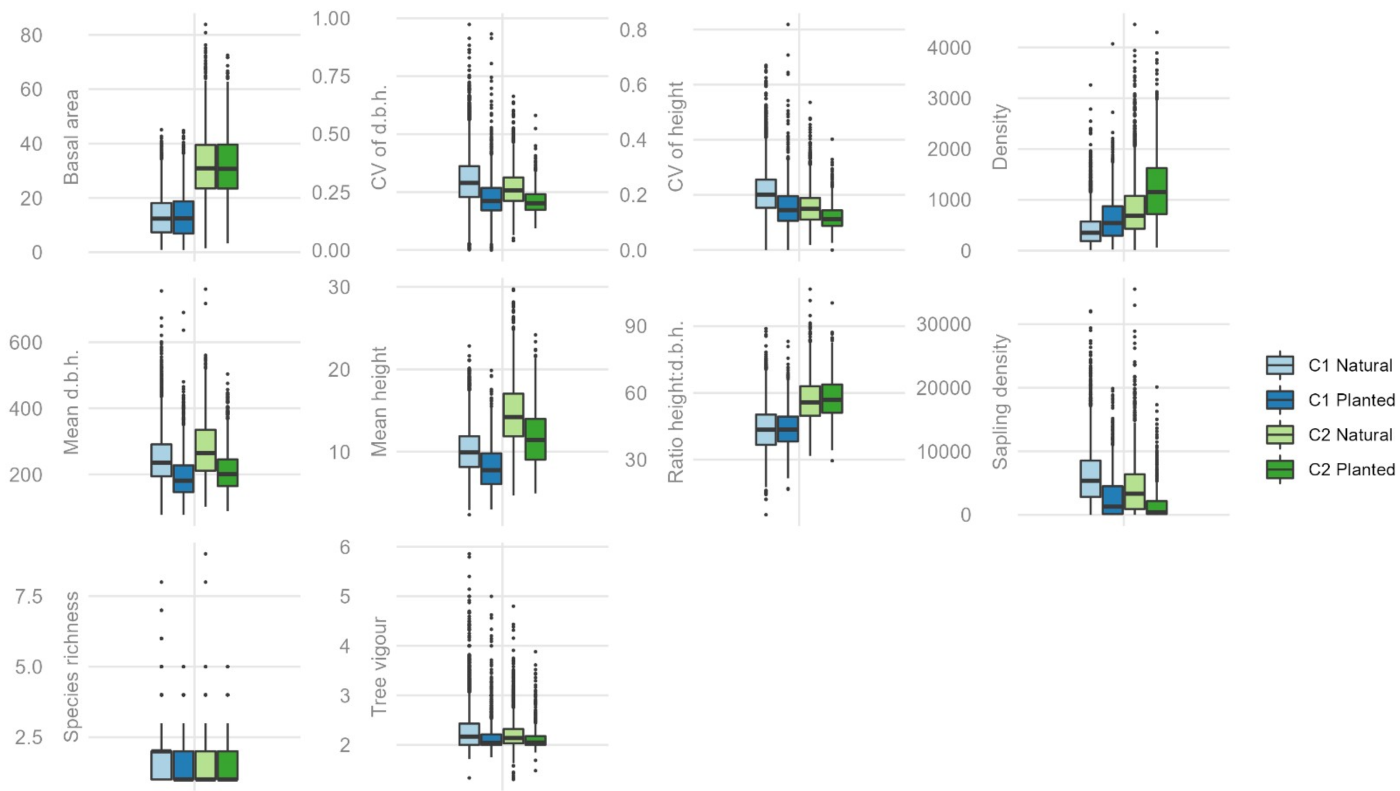

Figure S3. Boxplots showing the distribution of forest structural variables within each forest type (defined as stand origin x forest structural typology).

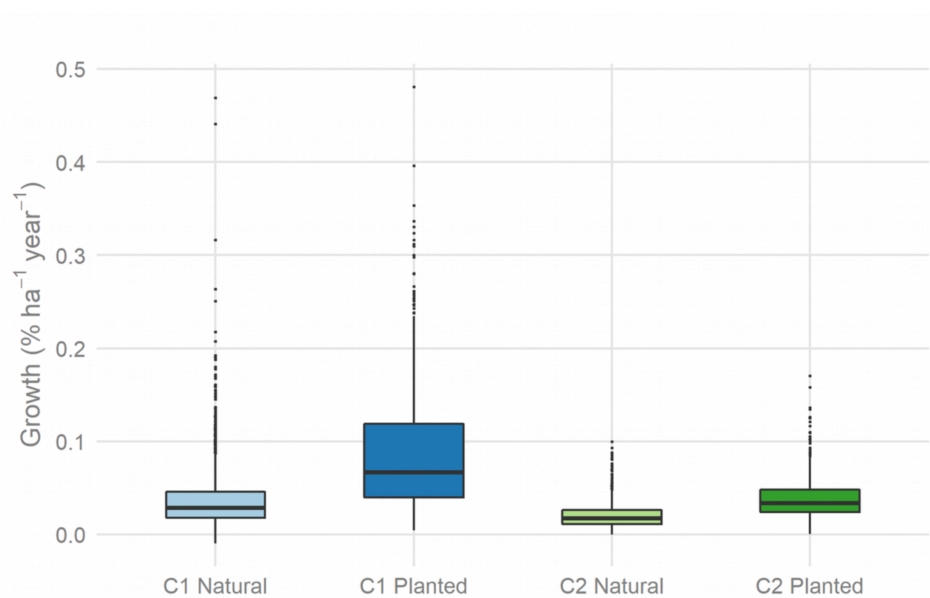

Figure S4. Growth rate for each forest type (defined as stand origin x forest structural typology) during the study period.

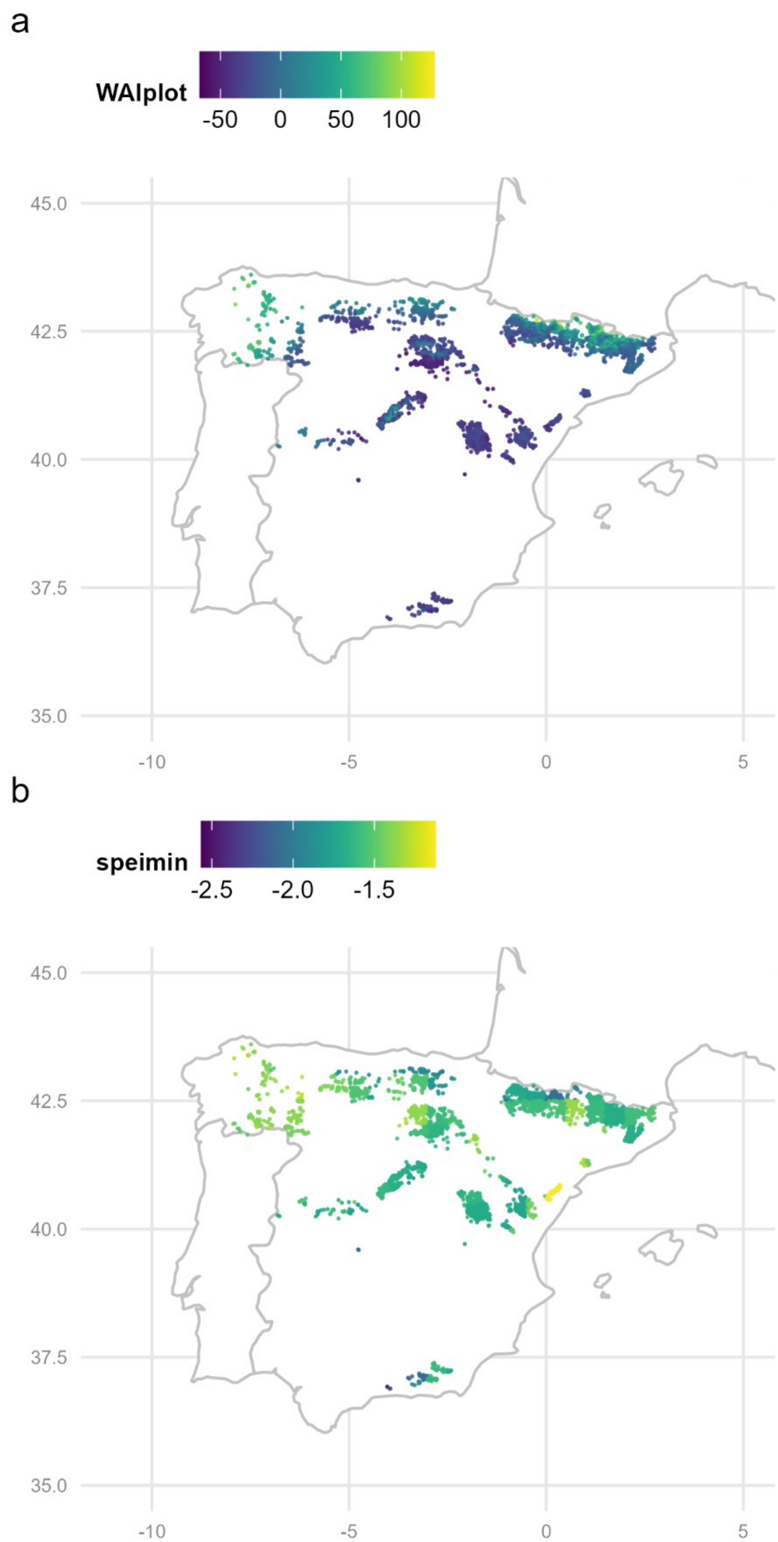

Figure S5. Geographic distribution of climatic variables: (a) water availability index (WAIplot, %), (b) drought (speimin).

Statistical models of tree mortality depending on anthropogenic land-use legacies and climate

We fitted a hurdle-gamma model using the glmmTMB R package (Brooks et al. 2022). The model was composed by (i) a conditional model with estimates as contrasts with a ziGamma conditional error distribution and a log link and (ii) a zero-inflation model with a logit link. The zero-inflation model quantifies the occurrence of mortality between two censuses by estimating the probability of an extra zero. The conditional model estimates the abundance of mortality in plots where mortality occurs (Brooks et al. 2017). The conditional and zero-inflation models included as predictors forest type (defined as stand origin x forest structural typology), water availability index, SPEI and their interactions following an Information Theoretical Approach (Zuur et al. 2013) an a priori knowledge of the effects of forest structure, aridity and drought on tree mortality (e.g. Ruiz-Benito et al. 2013; see Zenodo repository to see the coefficients and standard errors, https://doi.org/10.5281/zenodo.7801925). The predictor variables were not strongly collinear ($r = -0.07$) and model residual were diagnosed using DHARMa R package (Hartig 2022) (Fig. S6).

We used average predictive comparisons to quantify the impact of decreasing water availability and increasing drought severity on tree mortality as a function of forest type (Gelman and Pardoe 2007). Predictive comparisons gives the expected, or average, difference in $\Pr(y = 1)$ corresponding to a unit difference in each of the predictors (for further information on average predictive comparisons see Regression and Other Stories by Andrew Gelman, Jennifer Hill, and Aki Vehtari, https://avehtari.github.io/ROS-Examples/). Based on the IPCC predictions for the Mediterranean basin, we applied a 20% reduction in water availability and SPEI when considering all current stand conditions. Thus, the comparison between model-predicted tree mortality and model-predicted tree mortality under a climate change scenario (assuming a 20% reduction of both water availability and SPEI conditions) allowed us to quantify the impact of climate change and forest type on tree mortality (Fig. I (Box 1) and Zenodo repository, https://doi.org/10.5281/zenodo.7801925).

All analyses were performed in R Statistical Software (v4.2.0; R Core Team 2022) and in addition to the above-mentioned R packages we also used tidyverse (Wickham 2021), here (Müller 2020), patchwork (Pedersen 2020), sf (Pebesma 2022), viridis (Garnier 2021), MASS (Ripley 2022), ggdist (Kay 2022), RColorBrewer (Neuwirth 2022), testthat (Wickham 2022), factoextra (Kassambara and Mundt 2020), ggh4x (van den Brand 2021), janitor (Firke 2021) and gt (Iannone, Cheng, and Schloerke 2022) R packages for data processing, analyses and visualisation.

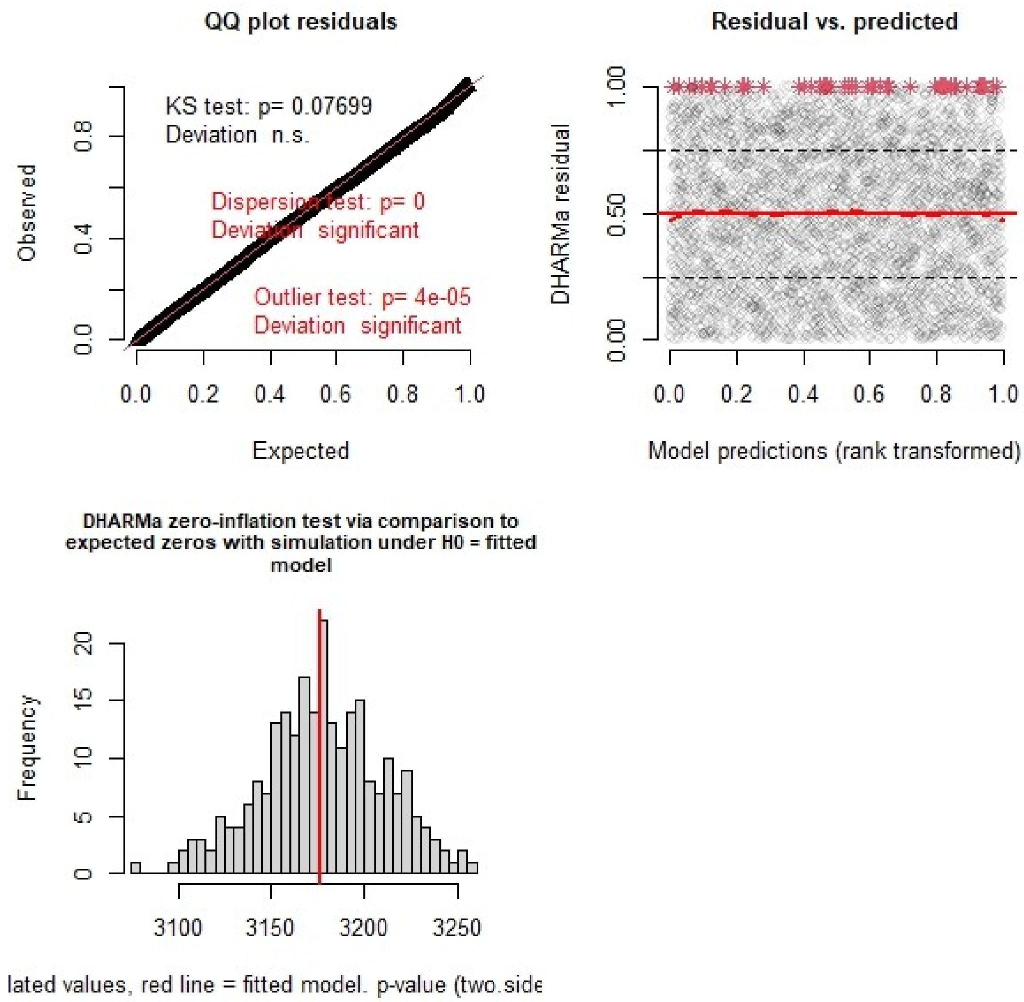

Figure S6. Model evaluation of the tree mortality model when regressed on forest type (defined as stand origin x forest structural typology) and climate variables. QQplot of the residuals, residuals versus predicted and a zero-inflation test are shown.

References

Alía, R., N. Alba, D. Agúndez, S. Iglesias, and (coord.). 2005. *Manual Para La Comercialización y Producción de Semillas y Plantas Forestales. Materiales de Base y de Reproducción*. Madrid: Serie Forestal. DGB.

Alía, R., J. M. García del Barrio, S. Iglesias, J. A. Mancha, J. de Miguel, J. L. Nicolás, F. Pérez-Martín, and D. Sánchez de Ron. 2009. *Regiones de Procedencia de Especies Forestales En España*. Madrid: DGB, Ministerio de Medio Ambiente, Medio Rural y Marino.

Astigarraga et al. (2020). Evidence of non-stationary relationships between climate and forest responses: Increased sensitivity to climate change in Iberian forests. Global Change Biology, 26:5063-5076.

Brooks et al. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. The R Journal, 9:2, 378-400.

Brooks, Mollie, Ben Bolker, Kasper Kristensen, Martin Maechler, Arni Magnusson, Hans Skaug, Anders Nielsen, Casper Berg, and Koen van Bentham. 2022. *glmmTMB: Generalized Linear Mixed Models Using Template Model Builder*. https://github.com/glmmTMB/glmmTMB.

Ceballos, L. 1966. *Forestal de España Escala 1:400000*. Madrid: Ministerio de Agricultura.

Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37 (12): 43024315. https://doi.org/10.1002/joc.5086.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Garnier, Simon. 2021. *Viridis: Colorblind-Friendly Color Maps for r*. https://CRAN.R-project.org/package=viridis.

Gelman, Andrew, and Iain Pardoe. 2007. "Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components." *Sociological Methodology* 37 (1): 23–51. https://doi.org/10.1111/j.1467-9531.2007.00181.x.

Gelman et al. (2020). Regression and Other Stories. Cambridge University Press.

Hartig, Florian. 2022. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. http://florianhartig.github.io/DHARMa/.

Iannone, Richard, Joe Cheng, and Barret Schloerke. 2022. *Gt: Easily Create Presentation-Ready Display Tables*. https://CRAN.R-project.org/package=gt.

IPCC Sixth Assessment Report (2022). Climate Change 2022: Impacts, Adaptation and Vulnerability

Kassambara, Alboukadel, and Fabian Mundt. 2020. *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. http://www.sthda.com/english/rpkgs/factoextra.

Kay, Matthew. 2022. *Ggdist: Visualizations of Distributions and Uncertainty*. https://CRAN.R-project.org/package=ggdist.

MacQueen (1967). Some methods for classification and analysis of multivariate observations. Editor(s) LM Le Cam, J Neyman. Berkeley Symp. on Math. Statist. and Prob., 1967: 281-297 (1967).

Moreno-Fernández, Daniel, Alicia Ledo, Darío Martín-Benito, Isabel Cañellas, and Guillermo Gea-Izquierdo. 2019. "Negative Synergistic Effects of Land-Use Legacies and Climate Drive Widespread Oak Decline in Evergreen Mediterranean Open Woodlands." *Forest Ecology and Management* 432 (January): 884–94. https://doi.org/10.1016/j.foreco.2018.10.023.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Neuwirth, Erich. 2022. *RColorBrewer: ColorBrewer Palettes*. https://CRAN.R-project.org/package=RColorBrewer.

Pebesma, Edzer. 2022. *Sf: Simple Features for r*. https://CRAN.R-project.org/package=sf.

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.<https://www.R-project.org/>.

Ripley, Brian. 2022. *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. http://www.stats.ox.ac.uk/pub/MASS4/.

Ruiz-Benito, Paloma, Lorena Gómez-Aparicio, and Miguel A. Zavala. 2012. "Large-Scale Assessment of Regeneration and Diversity in Mediterranean Planted Pine Forests Along Ecological Gradients." *Diversity and Distributions* 18 (11): 1092–1106. https://doi.org/10.1111/j.1472-4642.2012.00901.x.

Ruiz-Benito et al. (2013). Patterns and Drivers of Tree Mortality in Iberian Forests: Climatic Effects Are Modified by Competition. PLOS ONE, 8(2): e56843.

Ruiz-Benito et al. (2017). Climate- and successional-related changes in functional composition of European forests are strongly driven by tree mortality. Global Change Biology, 23:4162-4176.

Trabucco, Antonio, and Robert Zomer. 2019. "Global Aridity Index and Potential Evapotranspiration (Et0) Climate Database V2." *Figshare. Fileset*, January. https://doi.org/10.6084/m9.figshare.7504448.v3.

van den Brand, Teun. 2021. *Ggh4x: Hacks for Ggplot2*. https://github.com/teunbrand/ggh4x.

Vicente-Serrano, Sergio M., Santiago Beguería, and Juan I. López-Moreno. 2010. "A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index." *Journal of Climate* 23 (7): 1696–1718. https://doi.org/10.1175/2009JCLI2909.1.

Villaescusa, R., & Díaz, R. (1998). *Segundo Inventario Forestal Nacional (1986–1996)*. Madrid, Spain: Ministerio de Medio Ambiente, ICONA.

Wickham, Hadley. 2021. *Tidyverse: Easily Install and Load the Tidyverse*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley. 2022. *Testthat: Unit Testing for r*. https://CRAN.R-project.org/package=testthat.

Zuur (2013). Book: Beginner's Guide to GLM and GLMM with R.