# Mapping tree density at a global scale

Crowther, T.W.[1]*, Glick, H.B.[1], Covey, K.R.[1], Bettigole, C.[1], Maynard, D.S.[1], Thomas, S.M.[2], Smith, J.R.[1], Hintler, G.[1], Duguid, M.C.[1], Amatulli, G.[3] Tuanmu, M.-N.[3], Jetz, W.[3], Salas, C.[4], Stam, C.[5], Piotto, D.[6], Tavani, R.[7], Green, S.[8,9], Bruce, G.[8], Williams, S.J.[10], Wiser, S.K.[11], Huber, M.O.[12], Hengeveld, G.M.[13], Nabuurs, G.-J.[13] Tikhonova, E.[14], Borchardt, P.[15], Li, C.-F.[16], Powrie, L.W.[17], Fischer, M.[18,19], Hemp, A.[20], Homeier, J. [21], Cho, P.[22], Vibrans, A.C.[23], Umunay, P.M.[1], Piao, S.[24], Rowe, C.W.,[1] Ashton, M.S[1], Crane, P.R.[1] & Bradford, M.A.[1]

1. Yale School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA
2. Department of Environmental Sciences, University of Helsinki, Helsinki, Finland
3. Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA
4. Departamento de Ciencias Forestales, Universidad de La Frontera, Temuco, Chile
5. RedCastle Resources, Inc., Salt Lake City, UT, USA
6. Universidade Federal do Sul da Bahia, Ferradas, Itabuna, Brazil
7. Forestry Department, Food and Agriculture Organization of the United Nations, Rome, Italy
8. Operation Wallacea, Spilbsy, Lincolnshire, UK
9. Durrell Institute of Conservation and Ecology (DICE), School of Anthropology and Conservation (SAC), University of Kent, Canterbury, UK
10. Molecular Imaging Research Center MIRCen / CEA, Orsay, France
11. Landcare Research, Lincoln, New Zealand
12. WSL, Swiss Federal Institute for Forest, Snow and Landscape Research, Birmensdorf, Switzerland

13. Environmental Science Group, Wageningen University & Research Centre, Netherlands
14. Center for Forest Ecology and Productivity RAS, Moscow, Russia
15. CEN Center for Earth System Research and Sustainability, Institute of Geography, University of Hamburg, Hamburg, Germany
16. Department of Botany and Zoology, Masaryk University, Brno, Czech Republic
17. South African National Biodiversity Institute, Kirstenbosch Research Centre, Claremont, South Africa
18. Institute of Plant Sciences, University of Bern, Bern, Switzerland
19. Senckenberg Gesellschaft für Naturforschung, Biodiversity and Climate Research Centre BIK-F, Frankfurt, Germany
20. Department of Plant Systematics, University of Bayreuth, Bayreuth, Germany
21. Albrecht von Haller Institute of Plant Sciences, Georg August University of Göttingen, Göttingen, Germany
22. Tropical Ecology Research Group, Lancaster Environment Centre, Lancaster University, Lancaster, UK
23. Universidade Regional de Blumenau, Departamento de Engenharia Florestal, Blumenau/Santa Catarina, Brazil
24. College of Urban and Environmental Sciences, Peking University, Beijing, China

Running title: Global map of tree density

Correspondence: Thomas Crowther, Yale School of Forestry and Environmental Studies, 370 Prospect St, Yale University, New Haven, CT, USA

E-mail: Thomas.Crowther@yale.edu

Phone: +1 203 668 0064

**Abstract**

**The global extent and distribution of forest trees is central to our understanding of the terrestrial biosphere. We provide the first spatially continuous map of forest tree density at a global-scale. This map reveals that the global number of trees is approximately 3.04 trillion, an order of magnitude higher than the previous estimate. Of these trees, approximately 1.39 trillion exist in tropical and subtropical regions, with 0.74, and 0.61 trillion in boreal and temperate regions, respectively. Biome-level trends in tree density demonstrate the importance of climate and topography in controlling local tree densities at finer scales, as well as the overwhelming impact of humans across most of the world. Based on our projected tree densities, we estimate that deforestation is currently responsible for removing over 15 billion trees each year, and the global number of trees has fallen by approximately 46% since the start of human civilisation.**

## Introduction

Forest ecosystems harbour a large proportion of global biodiversity, contribute extensively to biogeochemical cycles, and provide countless ecosystem services, including water quality control, timber stocks and carbon sequestration[1–4]. Our current understanding of the global forest extent has been generated using remote sensing approaches that provide spatially explicit values relating to forest area and canopy cover[3,6,7]. Used in a wide variety of global models, these maps have enhanced our understanding of the Earth system[3,6,7], but they do not currently address population numbers, densities or timber stocks. These variables are valuable for the modelling of

broad-scale biological and biogeochemical processes[8–10] because tree density is a prominent component of ecosystem structure, governing elemental processing and retention rates [8,10,11], as well as competitive dynamics and habitat suitability for many plant and animal species[12–14]. The number of trees in a given area can also be a meaningful currency to guide forest management practices and inform decision-making in public and non-governmental sectors[15,16]. For example, international afforestation efforts such as the 'Billion Trees Campaign', and city-wide projects including the numerous 'Million Tree' initiatives around the world have motivated civil society and political leaders to promote environmental stewardship and sustainable land management by planting vast numbers of trees[15,17,18]. Establishing targets and evaluating the proportional contribution of such projects requires a sound baseline understanding of current and potential tree population numbers at regional and global scales[17,18].

The current estimate of global tree number is approximately 400.25 billion[5]. Generated using satellite imagery and scaled based on global forest area, this estimate engaged policy makers and environmental practitioners worldwide by suggesting that the ratio of trees-to-people is 61:1. This has, however, been thrown into doubt by a recent broad-scale inventory that used 1,170 ground-truthed measurements of tree density to estimate that there are 390 billion trees in the Amazon Basin alone[19].

**Mapping tree density**

Here, we use 421,529 ground-sourced measurements of tree density from every continent on Earth except Antarctica to generate a global map forest trees. Forested areas are found

in most of Earth's biomes, even those as counterintuitive as desert, tundra, and grassland (Fig. 1a; b). We generated predictive regression models for the forested areas in each of the 14 biomes (as defined by The Nature Conservancy - www.nature.org). These models link tree density to spatially explicit remote sensing and GIS layers of climate, topography, vegetation characteristics and anthropogenic land use (see Extended data Table 1). Following almost all of the collected data sources, we define a tree as a plant with woody stems larger than 10 cm diameter at breast height (DBH)[19].

Incorporating plot-level measurements from more than 50 countries, the measured tree density values were inherently variable within and among biomes (Fig. 1, 2). However, the vast number of tree density measurements ensured that the confidence in our mean (and total) estimates is high (Fig. 3). Furthermore, the scale of these data ensures that our modelled estimates are unlikely to be influenced significantly by recent deforestation, reforestation or natural forest regeneration, which are responsible for a net global change of < 1% of the global forest area each year[3]. Biome-level validation estimates indicate that our models have high precision when predicting the mean tree densities of omitted validation plots (Fig. 3a). Although the accuracy of our models is limited at the level of an individual hectare, the precision of the mean density estimates is high (+/- 40 trees $ha^{-1}$) beyond a threshold of ~200 plots (Fig. 3b).

**Global and biome-level patterns**

Together, the biome-level models provide the first spatially continuous map of global tree densities at a 1-$km^2$ (30 arc-seconds) resolution (Fig. 4a). Based on this map, we estimate

that the global number of trees is approximately 3.04 (± 0.096, 95% CI) trillion. An order of magnitude higher than the previous global estimate[5], the scale of our projection is consistent with recent large-scale inventories in Europe, North America and the Amazon Basin[19] (Fig. 4d). With a human population of 7.2 billion, our estimate of global tree density revises the ratio of trees-per-person from 61:1 to 422:1.

At the biome-level, the highest tree densities exist in forested regions of the Boreal and Tundra zones (Fig. 1b). In these northern latitudes, limited temperature and moisture lead to the establishment of stress-tolerant coniferous tree species that can reach the highest densities on Earth (Fig. 1). Despite this, the tropical regions contain a far greater proportion of the world's forested land. As a result, 42.8% of the planet's trees exist in tropical and subtropical regions, with 24.2% and 21.8% in boreal and temperate biomes, respectively (Fig. 4a).

**Within-biome trends**

Our models also provide mechanistic insights into potential controls on tree density within biomes (Fig. 5). For example, various climatic parameters correlate with mean forest density within all ecosystem types. Tree density generally increases with temperature (mean annual temperature and thermal consistency) and moisture availability (precipitation regimes, evapotranspiration or aridity). These patterns are consistent with previous broad scale tree inventory studies and support the idea that, within ecosystem types, moist, warm conditions are generally optimal for tree growth[12,13].

Given the generally positive effects of moisture availability and warmth on tree density within biomes, the negative relationships observed in some regions may seem surprising (Fig. 5). This highlights the complex suite of population- and community-level selection pressures that can obscure the expected effects of climate across landscapes. For example, in colder (boreal or tundra) biomes, increasing moisture levels can cause hydric and permafrost conditions in lower lying topographies, which then limit nutrient availability for tree development[20]. In addition, current and historical anthropogenic land use decisions have the potential to drive these relationships in several regions. The negative relationships between tree density and moisture availability in flooded grasslands and tropical dry forests are, for example, likely to be driven by preferential conversion of moist, productive land to agriculture[21]. As a result, forest ecosystems are often relegated to drier regions, reversing the expected within-biome relationships between moisture availability and tree density. Such effects will vary among countries, depending on human population densities, alternative resource availability and socio-economic status[22,23].

Along with these indirect effects of human activity, the direct impact of human development (% developed and managed land)[7] on tree density represented the only common mechanism across all biomes (Fig. 5). The negative relationships between tree density and anthropogenic land use exemplify how humans contend directly with natural forest ecosystems for space. Whereas the negative impact of human activity on tree numbers are highly apparent at local scales, the present study provides a new measure of the scale of anthropogenic effects, relative to other environmental variables. Current rates

171 of global forest loss are approximately 192 thousand $km^2$ each year[3]. By combining our
172 tree density information with the most recent spatially explicit map of forest area loss
173 over the past 12 years[3], we estimate that deforestation and land use change are currently
174 responsible for removing approximately 15.3 billion trees on an annual basis. Although
175 these rates of deforestation are currently highest in tropical regions[3], the scale and
176 consistency of this negative human effect across all forested biomes highlights how
177 historical land use decisions have shaped natural ecosystems on a global scale. Indeed,
178 using the projected maps of current and historic forest cover provided by the United
179 Nations Environment Programme (http://geodata.grid.unep.ch), our map reveals that the
180 global number of trees has fallen by approximately 45.8% since the onset of human
181 civilisation.
182
183 **Discussion**
184 The global map of tree density can facilitate ongoing efforts to understand
185 biogeochemical Earth system dynamics[3,7,8,10] by incorporating ecosystem features that
186 relate to elemental cycling rates[10,11]. For example, tree abundance can explain some of
187 the variation in carbon storage and productivity within ecosystem types[8,10], but the
188 strength of these effects remain untested across biomes[9]. We assessed the relationship
189 between tree density and plant carbon storage at a global scale by regressing our plot-
190 level tree counts against modelled estimates of plant biomass carbon in those sites[24]. This
191 revealed a positive effect of tree density on plant carbon storage ($p < 0.001$). However,
192 the strength of the relationship is weak ($r^2 = 0.14$), reflecting the vast array of local
193 ecological forces that can obscure such global trends. For example, the effect of tree

density is likely to interact strongly with tree size. Larger trees contain the greatest proportion of carbon in woodlands[25], but the highest tree densities within a given ecosystem type are often associated with young or recovering forests characterised by many small trees[14,20]. A thorough understanding of total vegetative carbon storage requires information about both the size and the number of individual trees.

A dense forest environment is a fundamentally different ecosystem from a sparse one and this influences a vast array of biotic and abiotic processes[11–13]. Current remote sensing tools capture some, but not all of this information. The tree density layer that we provide can therefore augment the currently available layers by providing unique insights into ecological dynamics that are not represented by estimates of forest cover or biomass[3,6,7]. It can inform biodiversity estimates and species distribution models by capturing perceivable environmental characteristics that determine habitat suitability for a wide variety of plants and animals[12–14]. Baseline estimates of tree populations are also critical for projecting population- and community-level tree demographics under current and future climate change scenarios[27], and for guiding local, national and international reforestation/afforestation efforts[15–18]. Finally, by allowing us to comprehend the global forest extent in terms of tree numbers, this map contributes to our fundamental understanding of the Earth's terrestrial system.

**Acknowledgements**

We thank Phillis Peterkins for their support throughout the study. The main project was funded by grants to Thomas Crowther from the Yale Climate and Energy Institute and the British Ecological Society. We acknowledge various sources for tree density measurements and estimates: the Canadian National Forest Inventory (https://nfi.nfis.org/index.php), the US Department of Agriculture Forest Service for their National Forest Inventory and Analysis (http://fia.fs.fed.us/), the Taiwan Forestry Bureau (which provided the National Vegetation Database of Taiwan), the DFG (German Research Foundation), BMBF (Federal Ministry of Education and Science of Germany), the Floristic and Forest Inventory of Santa Catarina (IFFSC) and the Chilean research grants FONDECYT No. 1151495, 11110270. For Europe NFI plot data were brought together with input from Jacques Rondeux and Martine Waterinckx, Belgium, Thierry Bélouard, France, Heino Polley, Germany, Wim Daamen and Henny Schoonderwoerd, Netherlands, Stein Tomter, Norway, J. Villanueva and Antoni Trasobares, Spain, Göran Kempe, Sweden. New Zealand Natural Forest plot data were collected by the LUCAS programme for the Ministry for the Environment (NZ) and sourced from the National Vegetation Survey Databank (NZ) (nvs.landcareresearch.co.nz). We also acknowledge the BCI forest dynamics research project, which was funded by National Science Foundation grants to Stephen P. Hubbell, support from the Center for Tropical Forest Science, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund, and numerous private individuals, and the hard work of hundreds of people from 51 countries over the past two decades. The plot project is part of the Center for Tropical Forest Science, a global network of large-scale demographic tree plots.

## Contributions

The study was conceived by T.W.C and G.H, and designed by T.W.C., and M.A.B. Statistical analyses and mapping were conducted by H.B.G., S.M.T., J.R.S., C.B., D.S.M. and T.W.C. The manuscript was written by T.W.C. with input from all authors. All other authors contributed substantially by providing tree density measurements or geospatial data from all over the world.

## References


1. Pan, Y. *et al.* A large and persistent carbon sink in the world's forests. *Science* **333,** 988–93 (2011).

2. Crowther, T. W. *et al.* Predicting the responsiveness of soil biodiversity to deforestation: a cross-biome study. *Glob. Chang. Biol.* **20,** 1–12 (2014).

3. Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science.* **134,** 850–854 (2013).

4. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320,** 1444–1449 (2008).

5. Nadkarni, N. *Between Earth and Sky: Our Intimate Connections with Trees*. (University of California Press, 2008).

6. Pfeifer, M., Disney, M., Quaife, T. & Marchant, R. Terrestrial ecosystems from space: A review of earth observation products for macroecology applications. *Glob. Ecol. Biogeogr.* **21,** 603–624 (2012).

7. Tuanmu, M. N. & Jetz, W. A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Glob. Ecol. Biogeogr.* **23,** 1031–1045 (2014).

8. Walker, A. P. *et al.* Predicting long-term carbon sequestration in response to $CO_2$ enrichment: How and why do current ecosystem models differ? *Global Biogeochem. Cycles* **5,** 1–20 (2015).

9. Asner, G. P. *et al.* A universal airborne LiDAR approach for tropical forest carbon mapping. *Oecologia* **168,** 1147–1160 (2012).

10. Fauset, S. *et al.* Hyperdominance in Amazonian forest carbon cycling. *Nat. Commun.* **6,** 6857 (2015).

11. Slik, J. W. F. *et al.* Environmental correlates of tree biomass, basal area, wood specific gravity and stem density gradients in Borneo's tropical forests. *Glob. Ecol. Biogeogr.* **19,** 50–60 (2010).

12. Leathwick, L. & Austin, M. Competitive interactions between tree species in New Zealand old-growth indigenous forests. *Ecology* **82,** 2560–2573 (2001).

13. Oliver, C. & Larson, B. C. *Forest stand dynamics*. (John Wiley and Sons, Inc., 1996).

14. Riginos, C. & Grace, J. B. Savanna tree density, herbivores, and the herbaceous community: Bottom-up vs. top-down effects. *Ecology* **89,** 2228–2238 (2008).

15. O'Neil-Dunne, J., MacFaden, S. & Royar, A. A versatile, production-oriented approach to high-resolution tree-canopy mapping in urban and suburban landscapes using GEOBIA and data fusion. *Remote Sens.* **6,** 12837–12865 (2014).

16. Guldin, R. W. Forest science and forest policy in the Americas: Building bridges to a sustainable future. *For. Policy Econ.* **5,** 329–337 (2003).

17. Cao, S. *et al.* Greening China naturally. *Ambio* **40,** 828–831 (2011).

18. Oldfield, E. E. *et al.* Growing the urban forest: tree performance in response to biotic and abiotic land management. *Restor. Ecol.* **23,** 491–499 (2015).

19. Ter Steege, H. *et al.* Hyperdominance in the Amazonian tree flora. *Science* **342,** 1243092 (2013).

20. Bonan, G. B. & Shugart, H. H. Environmental factors and ecological processes in boreal forests. *Annu. Rev. Ecol. Syst.* **20,** 1–28 (1989).

21. Meyfroidt, P. & Lambin, E. F. Global Forest Transition: Prospects for an End to Deforestation. Annu. Rev. Environ. Resour. **36,** 343-371 (2011).

22. Rudel, T. K. The national determinants of deforestation in sub-Saharan Africa. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20120405 (2013).

23. Hengeveld, G. M. *et al.* A forest management map of European forests. *Ecol. Soc.* **17,** 53 (2012).

24. Kindermann, G. E., McCallum, I., Fritz, S. & Obersteiner, M. A global forest growing stock, biomass and carbon map based on FAO statistics. *Silva Fenn.* **42,** 387–396 (2008).

25. Stephenson, N. L. *et al.* Rate of tree carbon accumulation increases continuously with tree size. *Nature* **507,** 90–3 (2014).

26. Lewis, S. L. *et al.* Above-ground biomass and structure of 260 African tropical forests. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20120295 (2013).

27. Zhu, K., Woodall, C. W., Ghosh, S., Gelfand, A. E. & Clark, J. S. Dual impacts of climate change: Forest migration and turnover through life history. *Glob. Chang. Biol.* **20,** 251–264 (2014).

## Figure Legends

**Fig. 1: Map of data points and raw biome-level forest density data.** Panel (a) highlights the Ecoregions (shape files provided by the Nature Conservancy - www.nature.org) from which ground-sourced measurements of tree density were collected. Shading indicates the total number of plot measurements collected in each Ecoregion. A global forest map was overlaid in green to highlight that collected data span the majority of forest ecosystems on a global-scale. Panel (b) shows the median and interquartile range of tree density values collected in the forested areas of each biome.

**Fig. 2: Heat plots showing the relationships between predicted and measured tree density data.** Predictions were generated using Generalised Linear Models (n = 421,529). Dotted lines indicate 1:1 lines (perfect correspondence) between predicted and observed points, scaled to the km-level. Colours indicate the proportion of data points

from that biome that fall within each pixel. Biomes with a greater number of plot measurements have greater variability, but higher confidence in the mean estimates, highlighting the trade-off between broad-scale precision and fine-scale accuracy. Note that axes are log-transformed to account for exceptionally high variability in tree density.

**Fig. 3. Validation plots for biome-level predictions.** Panel (a) shows how biome-level regression models predict the mean values of the omitted validation plot measurements in 12 biomes. Overall, the models underestimated mean tree density by ~3% (slope=0.97) but this difference was not statistically significant ($p$=0.51). Bars show +/- one standard deviation for the predicted mean. The values plotted here represent mean densities for the plot measurements (i.e. for forested ecosystems), rather than those predicted for each entire biome. Panel (b) shows the standard deviation of the predicted mean values as a function of sample size. As sample size increases, the variability of the predicted mean tree density reaches a threshold, beyond which an increase in sample size results in a minimal increase in precision. Standard deviations were calculated using a bootstrapping approach (see Methods), and smooth curves were modelled using standard linear regression with a log-log transformation.

**Fig. 4: The global map of tree density at the 1-km$^2$ pixel (30 arc-seconds) scale.** The scale refers to the number of trees in each pixel (a). We highlight the map predictions for two areas (South American Andes (b) and Sardinia (c)) and include the corresponding images for visual comparison. All maps and images were generated using ESRI basemap imagry. We also include a scatterplot as validation for our broad-scale estimates of total

tree number (d). This shows the relationship between our predicted tree estimates and reported totals for regions with previous broad-scale tree inventories (see Methods for details). The straight line and the dotted line are the predicted best fit line and the 1:1 line, respectively.

**Fig. 5: Standardised coefficients for the variables included in final biome-level regression models**. Coefficients represent relative percent change in tree density for 1 standard deviation increase in the variable. Red and blue circles refer to positive and negative effects on tree density, respectively. Circle size refers to the magnitude of effects. All layers are available at the global scale. Certain variables have shortened names: 'Human Development' = 'Percent Developed and Managed Land', 'LAI' = 'Leaf Area Index', 'EVI' = 'Enhanced Vegetation Index', 'EVI: ASM' = 'Angular Second Moment of EVI', 'EVI: contrast' = 'Contrast of EVI' and 'EVI: dissimilarity' = 'dissimilarity of EVI' (See Extended data Table 1).

## Methods

*Data collection and standardisation*

Plot-level data were collected from international forestry databases, including the Global Index of Vegetation-Plot Database (GIVD – www.givd.info), the Smithsonian Tropical Research Institute (www.stri.si.edu), ICP-Level-I plot data which covered most of Europe (www.icp-forests.org), and National Forest Inventory (NFI) analyses from 21 countries, including the U.S.A (http://fia.fs.fed.us/) and Canada (https://nfi.nfis.org/index.php). This information was supplemented with data from peer-reviewed studies reporting large

international inventories published in the last 10 years (collected using ISI Web of Knowledge, Google Scholar and secondary references)[19,26,28].

We only included density estimates where individual trees met the criterion of ≥ 10 cm diameter at breast height (DBH). Although NFI databases can vary slightly in their definition of a mature tree (e.g. the U.S. Forest Service Forest Inventory and Analysis (FIA)[29] defines a tree as a plant with woody stems larger than 12.7 DBH) the vast majority of sources use 10 cm as the DBH cut off. Indeed, this was the only size class provided by all broad-scale inventories (including the FIA), so density estimates at other DBH values were excluded. This provided a total of 421,529 measurements of forest tree density (each generated at the hectare scale), that were then linked to spatially-explicit remote sensing data and geographic information systems (GIS) variables to explore the patterns in forest tree density at a global-scale. The scale of our plot data (in terms of number and distribution of plots) ensured that any plot location uncertainty or minor changes in global forest area are unlikely to alter mean values or modelled estimates.

*Acquisition and preprocessing of spatial data*

For predictive model development, we selected 20 geospatial covariates from a larger pool of potential covariates based on uniqueness, spatial resolution and ecological relevance (Extended Data Table 1). Covariates were derived through satellite-based remote sensing and ground-based weather stations, and can be loosely grouped into one of four categories: topographic, climatic, vegetative, or anthropogenic. Topographic covariates included elevation, slope, aspect (as northness and eastness), latitude (as

absolute value of latitude) and a terrain roughness index (TRI). Climatic covariates included annual mean temperature, temperature annual range, annual precipitation, precipitation of driest month, precipitation seasonality (coefficient of variation), precipitation of driest quarter, potential evapotranspiration per hectare per year, and indexed annual aridity. Vegetative covariates included, enhanced vegetation index (EVI), leaf area index (LAI), dissimilarity, contrast, and angular second moment. We also included a single anthropogenic covariate: proportion of urban and/or developed land cover (see Extended Data Table 1).

Several covariates bear special mention. Moving window analyses were applied to an Enhanced Vegetation Index (EVI) derived from a multi-year composite of Moderate Resolution Imaging Spectroradiometer (MODIS) imagery. From the result, we extracted four second-order textural covariates that reflect the heterogeneity of vegetation, intended to capture difference in vegetative structure. These include TRI (i.e. the mean of absolute differences between a cell and its adjacent neighbours), angular second moment (i.e. the orderliness of EVI among adjacent pixels), contrast (i.e. the exponentially weighted difference in EVI between adjacent pixels), and dissimilarity (i.e. difference in EVI between adjacent pixels). Terrain Roughness Index was computed using the eight neighbouring pixels, while the others were computed using the four neighbouring pixels located at 0°, 45°, 90°, 135°.

We preprocessed all spatial covariates using ArcMap 10.1 (ESRI, Redlands, CA, 2012) and RStudio 0.97.551 (RStudio, Inc., 2012). All covariates were reprojected to the interrupted Goode Homolosine equal-area coordinate system (which maximises spatial

419 precision by amalgamating numerous region-specific equal-area projections) to optimise
420 the areal accuracy of our final figures[30]. These were then resampled to match the coarsest
421 resolution used during analysis (nominal 1 $km^2$ pixels), and spatially coregistered using
422 nearest neighbour resampling where necessary.
423
424 To account for broad-scale differences in vegetation types, we developed spatial models
425 at the biome scale. Individual predictive models were generated within each of 14 broad
426 ecosystem types (delineated by the Nature Conservancy - www.nature.org) to improve
427 the accuracy of estimates. These were projected to WGS84 before modelling.
428
429 *Statistical modelling*
430
431 We used generalized linear models to generate predictive maps of tree numbers within
432 forested ecosystems for each biome (Extended data Table 2). This approach also enabled
433 us to explore the mechanisms potentially governing patterns in forest tree density within
434 regions (Fig. 5). Due to the inherently interactive nature of climate, soil and human
435 impact factors across the globe, we predicted that there would be pronounced non-
436 independence within the full suite of biophysical variables extracted from the compiled
437 GIS layers. To account for this colinearity, we performed ascendant hierarchical
438 clustering using the *hclustvar* function in R's *ClustOfVar* package[31] in each biome-level
439 model. This analysis splits the variables into different clusters (similar to principal
440 components) in which all variables correlate with one another. A single best 'indicator'
441 variable is then selected from each cluster, based on squared loading values representing
442 the correlation with the central synthetic variable of each cluster (i.e. the first principal

443 component of a PCAmix analysis). This set of 'best' indicator variables for each biome
444 was then included in all subsequent models used to estimate controls on forest tree
445 density.
446
447 Using the resulting set of variables, we constructed generalized linear models with a
448 negative binomial error structure (to account for count data that could not extend below
449 zero) for each biome (Extended data Figs 1; 2; 3) and performed a multi-model dredging
450 function using the *dredge* function in R's *MuMIn* package[32]. This function constructs all
451 possible candidate sub-models nested within the global model, identifies the most
452 plausible subset of models for each data set, and then ranks them according to corrected
453 Akaike Information Criterion (AICc) values and AIC likelihood weights (AICcw). We
454 derived covariates, coefficients, and variance-covariance matrices for biome-level models
455 through weighted model averaging the dredged model results with cumulative AIC
456 weights at least equal to 0.95[33]. Given the inherent sampling bias present in our plot data
457 (tree density estimates were only collected in forested ecosystems and non-forested
458 regions are under-represented), our modelling approach was used to generate predictive
459 estimates of forest tree density, and these estimates was subsequently scaled based on the
460 total area of forested land in each pixel (see *Spatial modelling* for details).
461
462 *Model validation and testing*
463
464 We assessed the model fit by investigating the bias and precision present when predicting
465 mean tree density across an aggregate number of plots. This approach allowed us to test
466 how many plots are required to ensure that the predicted mean (or total) forest density has

467 reasonable bias and precision. 20% of the plots within each biome were randomly
468 omitted before model fitting to serve as an independent dataset for model testing. Initial
469 model validation was conducted using the biome-specific regression models (obtained
470 from the remaining 80% of the data) to predict the tree density for each omitted plot. The
471 mean predicted tree density of the omitted data was then regressed against the mean
472 observed tree density of the omitted data for each biome (Fig 2). In addition, a
473 bootstrapping algorithm was used to quantify the standard deviation of the mean
474 prediction as a function of sample size following[34]. For each biome, we generated
475 empirical bootstrap estimates of the standard deviation of the predicted mean using
476 random samples drawn from the withheld validation plots. Specifically, for each biome a
477 bootstrap sample of size N was selected, with replacement, from the omitted data in that
478 biome. The fitted regression model for that biome (based on the 80% retained data) was
479 used to predict the tree density of each point, and the mean of the *n* samples was
480 calculated. This process was repeated 10,000 times for each sample size (*n*=10, 20, …,
481 500) and in each case the empirical standard deviation of the 10,000 sample mean was
482 calculated and plotted (Fig. 2). Where the number of plot records in a biome fell below
483 the sample size threshold identified through bootstrapping, we used models from the
484 most similar biome available (in terms of phylogenetic relatedness of the dominant tree
485 species and mean tree density from the few plot values collected). This was the case for
486 the two smallest biomes: 'Mangroves' (0.23% of land surface) and 'Tropical Coniferous'
487 (0.46% of land surface) forests, which used models from 'Tropical Moist' and
488 'Temperate Coniferous', respectively.
489
490 *Spatial modelling*

Following model averaging and bootstrapping, we applied the final negative binomial regression equations used in bootstrapping to pixel-level spatial data at the biome level. Regressions were run in a map algebra framework wherein equation intercepts and coefficients were applied independently to each pixel of our coregistered global covariates to produce a single map of forest tree density on a per-hectare scale. We then scaled our per-hectare forest density estimates to the 1 $km^2$ scale based on the total area of forested land within each pixel, as estimated by the Global 1-km Consensus Land Cover dataset for 2014[7]. This process was then validated using an older (2013) dataset that used fine-scale (30 m) forest cover information[3], which revealed equivalent total tree counts. By multiplying our predicted forest density by the area of forest, we ensured that we did not overestimate tree densities in non-forested sites. From the resulting maps, summary statistics (mean tree density, total tree number) were derived for each polygonal area of interest. The variances of the global and biome-specific totals were calculated using a Taylor series approximation to account for the log-link negative binomial regression function and correlation among the regression-based predicted values[35].

By generating models at the biome-level, we were able to account for broad-scale differences in vegetation types between biomes, whilst maintaining high precision of our mean (and total) estimates at the global scale (due to the high number of plot measurements within biomes). However, biome-level models are limited in their accuracy when predicting tree density at fine-scales, which might ultimately have the potential to alter final numbers. We therefore constructed models within each of 813 global ecoregions (delineated by TNC - www.nature.org) as a validation for the first

biome-level approach. We generated models and estimated tree numbers using exactly the same approach as for the biome-level models. Total, and biome-level, tree estimates did not differ significantly ($p < 0.05$) from those generated using the biome-level models (Extended data Fig. 4).

## References


28. Brus, D. J. *et al.* Statistical mapping of tree species over Europe. *Eur. J. For. Res.* **131,** 145–157 (2012).

29. FAO. in *Food Agric. Organ. United Nations* 375 (2010).

30. Steinwand, R. S., Hutchinson, J. A. & Snyder, J. P. Map projections for global and continental data sets and an analysis of pixel distortion caused by reprojection. *Photogramm. Eng. Remote Sensing* **61,** 1487–1499 (1995).

31. Chavent, M., Kuentz, V., Liquet, B. & Saracco, J. ClustOfVar: An R Package for the Clustering of Variables. *J. Stat. Softw.* **50,** 1–16 (2012).

32. Bartoń, K. MuMIn: Model selection and model averaging based on information criteria (AICc and alike). (2015).

33. MacKenzie, D. I. *et al.* Occupancy estimation and modeling. (2006).

34. Maclean, M. G. *et al.* Requirements for labelling forest polygons in an object-based image analysis classification. *Int. J. Remote Sens.* **34,** 2531–2547 (2013).

35. Stahl, G. *et al.* Model-based inference for biomass estimation in a LiDAR sample survey in Hedmark County, Norway. *Can. J. For. Res.* **41,** 96–107 (2011).


**Extended Data Table 1:** List of Geographic Information System layers collected and used to predict tree density. Each layer is available at the global level. We also indicate

whether each later was included (Inclusion) or not in final predictive models following clustering analysis and model selection/averaging.

**Extended Data Table 2:** Summary table showing the number of plot estimates and total tree numbers (with 95% confidence interval) at the biome and global scale.

**Extended Data Fig. 1:** Histogram of the collected measurements of forest tree density in each biome. The red line and the blue dotted lines indicate the mean and median for the collected data respectively. Data in each biome fitted a negative binomial error structure.

**Extended Data Fig. 2:** Histogram of the predicted forest tree density values for the locations that density measurements were collected in each biome. The red line and the blue dotted lines indicate the mean and median for the collected data respectively. As our models were based on mean values, the majority of points fall on or close to the mean values in each biome.

**Extended Data Fig. 3:** Histogram of the total predicted forest tree density values for the each pixel within each biome. This illustrates the spread of pixels throughout each biome, and highlights that our map accounts for the sampling bias in tree density plots (e.g. although we had no zero values in our Desert plots, the vast majority of desert pixels contain no trees).

**Extended Data Fig. 4:** Comparison between approaches to generate the global tree density map. The initial map was generated using 14 biome-level models (biomes delineated by the Nature conservancy (TNC) - www.nature.org) to account for broad-scale variations in terrestrial vegetation types. With several thousand plot-level density measurements in most biomes, this approach provided highly accurate estimates at the global scale. However, to improve precision at the local scale, we also generated a map using ecoregion-scale models. Separate models were generated within each of 813 global ecoregions (also delineated by TNC to reflect smaller-scale vegetation types) using exactly the same statistical approach (see Methods). Biome-level and ecoregion-level maps provide total tree estimates of 3.041 and 3.184 trillion trees, respectively.