

# SCIENTIFIC DATA

OPEN

## Data Descriptor: EU-Forest, a high-resolution tree occurrence dataset for Europe

Achille Mauri<sup>1</sup>, Giovanni Strona<sup>1</sup> & Jesús San-Miguel-Ayán<sup>1</sup>

Received: 2 August 2016

Accepted: 30 November 2016

Published: 5 January 2017

We present EU-Forest, a dataset that integrates and extends by almost one order of magnitude the publicly available information on European tree species distribution. The core of our dataset (~96% of the occurrence records) came from an unpublished, large database harmonising forest plot surveys from National Forest Inventories on an INSPIRE-compliant 1 km × 1 km grid. These new data can potentially benefit several disciplines, including forestry, biodiversity conservation, palaeoecology, plant ecology, the bioeconomy, and pest management.

<b>Design Type(s)</b>	data integration objective • observation design • biodiversity assessment objective
<b>Measurement Type(s)</b>	Species
<b>Technology Type(s)</b>	digital curation
<b>Factor Type(s)</b>	
<b>Sample Characteristic(s)</b>	forest • Austria • Belarus • Belgium • Bulgaria • Croatia • Czech Republic • Estonia • Finland • French Republic • Germany • Greece • Hungary • Island of Cyprus • Italy • Kingdom of Denmark • Kingdom of Norway • Kingdom of Spain • Latvia • Lithuania • Moldova • Poland • Portuguese Republic • Republic of Ireland • Romania • Slovak Republic • Slovenia • Sweden • Switzerland • The Netherlands • United Kingdom

<sup>1</sup>European Commission, Joint Research Centre, Directorate D—Sustainable Resources, Bio-Economy Unit, Via Enrico Fermi 2749, Ispra 21027, Italy. Correspondence and requests for materials should be addressed to A.M. (email: achille.mauri@ec.europa.eu).

## Background & Summary

Forests play a key role in biodiversity conservation, by providing a habitat for the majority of world's terrestrial animal and plant species<sup>1,2</sup>. Although this is particularly evident in tropical environments, it is also true for temperate habitats<sup>3–6</sup>. In Europe, forests cover approximately 33% of total land area, and their spatial extent has even increased over the past 25 years as a result of land abandonment following growing urbanisation<sup>7,8</sup>. Furthermore, the distribution of forested areas and their species composition is rapidly changing as a consequence of climate change<sup>9</sup>, anthropogenic activities<sup>10</sup>, and increasing natural hazards such as forest fires and windstorms<sup>11</sup>. In this context, satellite data can provide valuable information by detecting forest location and extent<sup>12</sup>, but can do little to map tree species distribution with the appropriate spatial and taxonomic resolution needed to monitor changes in terms of tree community composition and structure.

To tackle this issue, the Joint Research Centre (JRC) of the European Commission supported projects that aim to harmonise European forest information, and coordinated the implementation of European Union (EU) actions regarding forest monitoring, including the Forest Focus Regulation<sup>13</sup>. Hence, the Forest Focus monitoring database (hereafter, the 'Forest Focus')<sup>14</sup> was set up to monitor the effects of atmospheric pollution on forests. In the same framework, a supplementary pilot study, the Biosoil project<sup>15</sup>, was also set up to integrate information about forest soils and forest biodiversity at the European level.

Although the tree occurrence data resulting from the Forest Focus database and the Biosoil project have proved useful in various studies dealing with forestry, climate change, biodiversity and ecosystem services<sup>16–24</sup>, they have significant limitations in terms of spatial data density, spatial resolution and geographical coverage. Furthermore, the methodological protocols for data acquisition in Forest Focus, which were established in the 1980s under Council Regulation EEC No. 3528/86 (ref. 25), lack detail, which possibly affects the reliability and spatial accuracy of the resulting occurrence data.

Another valuable source of forest data for Europe is the Atlas Florae Europaeae<sup>26,27</sup>, which provides an exhaustive inventory of vascular plants at the continental scale, and which has been extensively used to estimate large-scale tree species distribution and tree species richness<sup>28–32</sup>. However, the low spatial resolution of the dataset (nearly 50 km) has often limited its applicability, especially in conservation studies, forcing users to manipulate the data by applying downscaling techniques, which can potentially affect the reliability of results<sup>33</sup>.

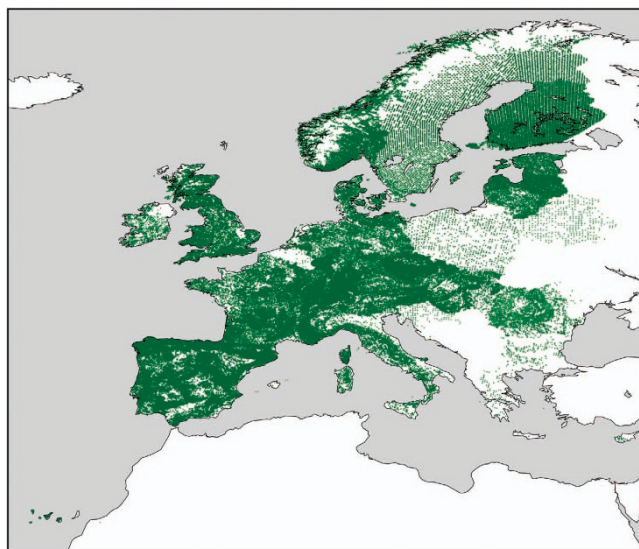
Additionally, most European countries have implemented forest monitoring and inventory programmes at the national level (the National Forest Inventories, NFIs). The data collected in the context of these programmes have huge potential, both in terms of quantity and quality. However, this important resource has long remained unexploited due to country-specific restrictions on their availability, and to fundamental differences in the national survey procedures, which complicate data harmonisation. To overcome these problems and make NFI data publicly available and readily usable, in 2008 the JRC established a framework contract with European countries (including EU Member States and neighbouring countries) to regulate the provision of forest data and services in support to the European Forest Data Centre (EFDAC)<sup>34</sup>. The motivation behind the framework contract was to provide EU decision-makers with high-quality harmonised forest data to develop and implement environmental policies at European level<sup>35,36</sup>. This eventually led to the compilation of a comprehensive dataset that includes tree occurrence data from 21 countries' NFIs.

We here make those data publicly available, by harmonising and merging them with the tree occurrence data provided by the Forest Focus and Biosoil datasets. The joint dataset, called EU-Forest, has a very high data density (including more than half a million occurrences) and a moderately high resolution of 1 square kilometre (Figs 1 and 2), and is by far the largest collection of tree species occurrences in Europe.

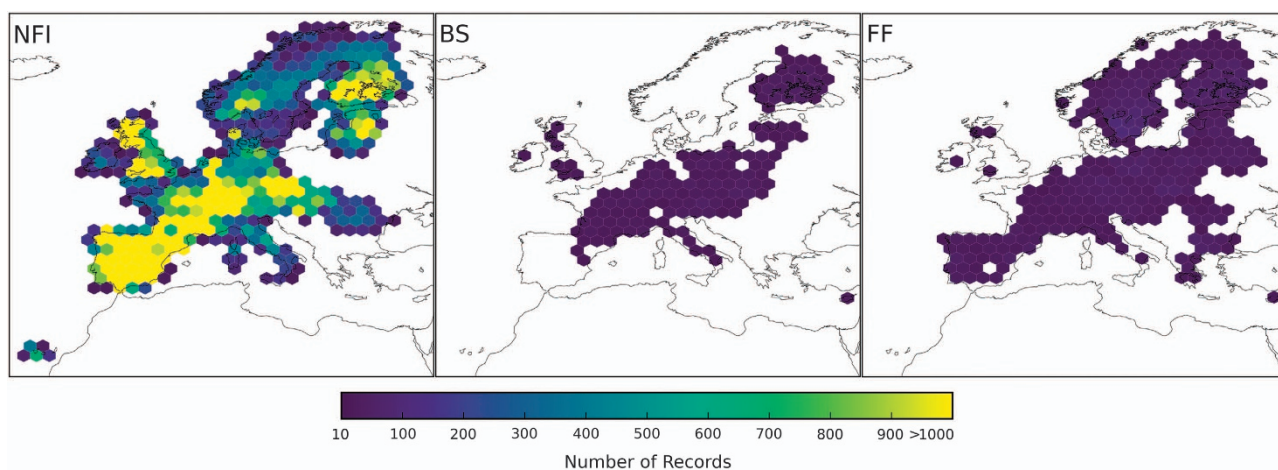
The EU-Forest dataset is an extremely valuable resource for ecological and conservation studies<sup>37</sup>. For instance, detailed data on large-scale tree species distribution may help orient conservation efforts by developing accurate biodiversity indicators<sup>38</sup>, and improve our understanding of how European forests will respond to climate change<sup>39,40</sup>. Furthermore, considering the importance of wood to the European economy<sup>18,41</sup>, the information within the dataset will no doubt have strong political and social implications, possibly improving the international transparency on the geo-political distribution of wood resources.

The EU-Forest dataset also has the potential to improve our preparedness with regard to forest pests, and to help mitigate the threats posed by emerging forest diseases<sup>42</sup>. In particular, drawing detailed maps of tree species that are capable of hosting harmful pathogens could provide an important resource in the context of pest-spread modelling and management<sup>43,44</sup>. Moreover, EU-Forest could be used in conservation management to accurately map the European distribution of tree species richness and rarity (Fig. 3) and improving our understanding of European tree biogeography (Fig. 4). For example, the dataset has already been used to show how the ecological spatial structure (in terms of nestedness) of actual tree vegetation departs from its natural potential<sup>45</sup>.

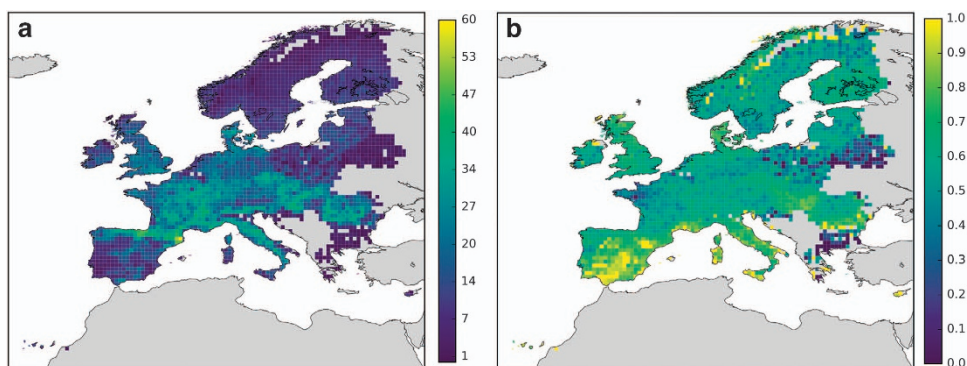
Knowing the distribution of forest tree species is also crucial to support ecosystem services and functions. A common assumption is that these aspects are closely associated to tree diversity, i.e., the more tree species a forest hosts, the higher its overall biodiversity, and the more numerous the ecosystem services it can offer<sup>46</sup>. This idea, which has clear implications for prioritisation and management, has



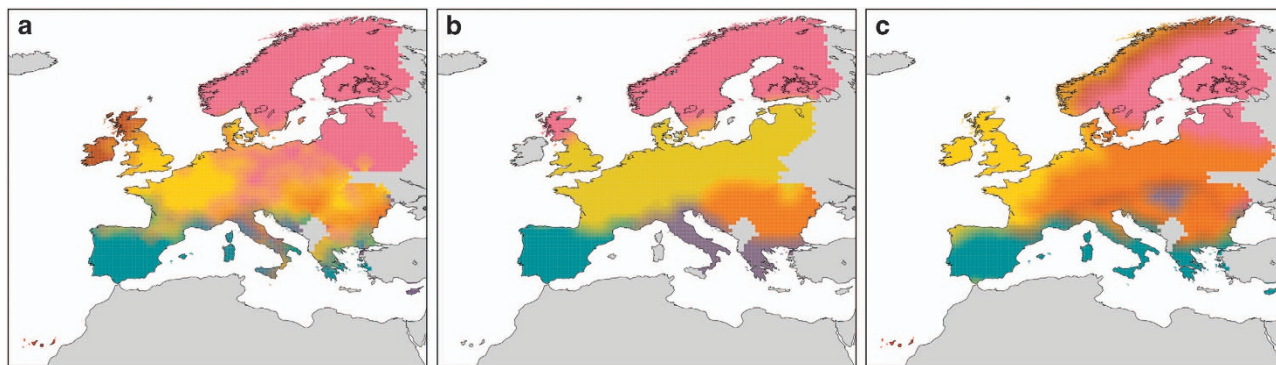
**Figure 1.** Spatial distribution of all the occurrences present in EU-Forest.



**Figure 2.** Geographical coverage and density of occurrence records for the three different sources used to compile the dataset. BS, Biosoil; FF, Forest Focus monitoring database; NFI, National Forest Inventories dataset. The colour of each hexagon indicates the corresponding number of records. Hexagons were drawn only if they overlapped with at least 10 occurrences.



**Figure 3.** Tree diversity indices. Tree species richness (a) and relative rarity (b) estimated from the merged dataset at a resolution of  $0.5 \times 0.5$  degrees. Relative rarity was computed following Leroy *et al.*<sup>68</sup>.



**Figure 4. Biogeographical regions.** Comparison between the biogeographical classification obtained from our merged dataset using the procedure by Vilhena and Antonelli<sup>59</sup> at a resolution of  $0.5 \times 0.5$  degrees (a), the classification by Rueda *et al.*<sup>61</sup> (upscaled to the same resolution to ease comparison, b), and the European classification of biogeographical regions (c).

been confirmed by large-scale studies in North America and part of China<sup>47</sup>. Nevertheless, it still lacks of convincing support in Europe, mostly due to limitations in the extension and resolution of previously available forest datasets that could not provide comprehensive coverage of diverse climatic and vegetation regions<sup>7</sup>. The high data density of EU-Forest may provide essential information to fill this knowledge gap.

Besides these general considerations, we are confident that the data we share in this paper will serve many other purposes. For instance, the EU-Forest dataset will also be extremely valuable in other disciplines such as palaeoecology and palaeoclimatology, improving reconstructions of past climate changes, and enabling interesting comparisons between the actual and palaeo-distribution of major tree species in Europe<sup>48,49</sup>.

## Methods

We merged information from the two largest existing datasets on European tree distribution (the Forest Focus and Biosoil databases, see previous paragraph), and an unpublished, much larger dataset derived from National Forest Inventories (hereafter, the ‘NFI dataset’). The NFI dataset was generated for the provision of forest data and services to support the European Forest Data Centre (EFDAC). This is the result of a framework contract established between the JRC of the European Commission (<https://ec.europa.eu/jrc/>) and European countries. The collection of tree occurrence data involved 19 EU Member States and two neighbouring countries (Norway and Switzerland), for a total of 21 countries<sup>34</sup>. Harmonisation procedures were agreed by the JRC and participant countries. These included the establishment of a common nomenclature for forest tree species among participant countries for the most common trees (approximately 200 tree species), and the establishment of standardised procedures for the provision of data to the JRC.

Although raw data uploaded by each country were available at a higher spatial resolution, they were subsequently aggregated at a lower spatial resolution of 1 square kilometre, in line with an INSPIRE-compliant  $1 \text{ km} \times 1 \text{ km}$  grid<sup>50</sup>, specifically designed for pan-European mapping<sup>51</sup>. This upscaling procedure was necessary both for practical and legal reasons. First, it provided a way to standardise the information from different NFIs, which was heterogeneous in terms of spatial resolution and accuracy due to national differences in sampling design and in the establishment of sampling plots<sup>52,53</sup>. Second, the upscaling was a necessary condition to comply with different national rules regarding owners’ privacy protection, which often prevented the exact location of tree occurrences from being made public. For the same reason, the original data from Forest Focus and Biosoil, although available at higher than 1-km resolution, were harmonised with NFI data by attributing each occurrence record to the centroid of the corresponding cell into which the record fell.

The resulting EU-Forest dataset (Figs 1 and 2) includes a total of 1,000,525 occurrence records, 96% of which were obtained from the NFI dataset, with the remaining 4% coming from Biosoil and Forest Focus data. The NFI dataset occurrences are uniformly distributed over 248,776 plots across most of the European territory, although there are (few) geographical gaps for Poland, Croatia, Slovenia, Greece, Bulgaria, Cyprus, Belarus, Moldova and the Canary Islands (Fig. 2). The Forest Focus and Biosoil data are crucial to fill in some of these gaps with the addition of 8,564 plots and 20,634 occurrences from the former dataset, and 3,367 plots and 19,114 occurrences from the latter. Besides geographic extent, the NFI dataset is also much more complete than the other two datasets in terms of taxonomic diversity, including information for 78 tree genera and 242 tree species (c.f. Forest Focus: 23 genera and 47 species, and Biosoil: 57 genera and 187 species).



After the removal of incomplete records (i.e., where taxonomic identification was not at the species level) and of duplicate occurrences of the same tree species in the same plot (mostly arising from repeated sampling at different times), we obtained two datasets including over 249,410 plots, one at species level (588,983 occurrences), and one at genus level (589,657 occurrences). For the sake of completeness, we provide a summary table that summarizes the number of plots, the number of species/genera, and the number of species/genera occurrences per country (Table 1).

Data Records

We provide two datasets for tree occurrences at, respectively, species and genus level (EU-Forest\_species, EU-Forest\_genus). Both datasets are available from *figshare* (Data Citation 1) as compressed (.zip) files in comma separated values (.csv) format, with 10 columns indicating coordinates in a ETRS89-LAEA reference coordinate system, representing the centroid of the INSPIRE-compliant 1 km × 1 km European grid (X, Y); the country where the forest plot was sampled (COUNTRY); the source datasets (NFI, BS, FF); the name of the species/genus sampled (SPECIES/GENUS NAME); the class for the diameter at breast height (DBH-1, DBH-2) of the trunk, which is used to distinguish a new recruit of a given tree. This latter measurement has a value of 1 for trunks with a diameter of less than 120 mm, 2 for trunks with diameter greater than 120 mm, and -9999 for trunks with unknown diameters. The value of 120 mm represents the common minimum threshold adopted by the NFIs, although single NFIs have different minimum thresholds, ranging from 0 mm in Finland to 120 mm in Cyprus and Switzerland<sup>54</sup>. The choice

Country	N. Plots	N. Species	Species Occs.	N. Genus	Genus Occs.
Austria	3,525	66	10,924	34	10,816
Belarus	408	12	797	11	795
Belgium	32	24	108	20	104
Bulgaria	220	10	243	8	231
Croatia	78	25	181	16	179
Cyprus	19	14	71	9	50
Czech Republic	10,444	75	31,783	35	31,118
Denmark	2,470	47	7,208	32	9,185
Estonia	2,958	19	8,076	19	10,182
Finland	24,649	34	66,849	18	63,074
France	31,925	141	81,300	61	77,944
Germany	18,903	84	67,297	38	68,469
Greece	82	19	116	9	104
Hungary	4,428	81	15,427	39	13,076
Ireland	1,457	48	2,647	31	2,762
Italy	6,810	128	21,151	59	20,025
Latvia	4,144	28	5,701	21	8,444
Lithuania	2,093	34	7,454	18	8,786
Moldova	10	8	17	6	17
Netherlands	2,672	26	5,694	20	7,017
Norway	10,844	22	24,388	20	25,288
Poland	981	61	3,622	30	3,454
Portugal	4,646	17	6,211	17	7,261
Romania	3,291	92	13,814	46	13,138
Slovakia	1,449	64	4,311	33	4,122
Slovenia	69	64	558	32	500
Spain	74,411	130	119,360	64	111,065
Sweden	11,344	23	24,837	18	29,708
Switzerland	5,530	45	13,130	29	14,070
United Kingdom	19,518	52	42,949	32	49,793

**Table 1.** Summary table illustrating the number of plots (N. Plots), number of species and their occurrences (N. Species and Species Occs.), number of genera and their occurrences (N. Genus and Genus Occs.) for each country listed in our dataset. The values are summarized after the technical validation (i.e., excluding occurrences falling outside of the respective alpha shape EOs).

of threshold has important implications, since estimations of biomass based on different thresholds can deliver very different results<sup>55</sup>.

In addition, for the species-level dataset, we provide an additional field stating whether or not the target occurrence falls within the species' geographical range, measured as the extent of occurrence (EOO, see Technical validation paragraph for details). Finally, we provide two compressed archives (.zip) containing, respectively, the individual species occurrences as 242 point shapefiles, and the EOOs of all species having at least three occurrences as 203 polygon shapefiles.

### Technical Validation

We validated the scientific names of trees using the Taxonomic Name Resolution Service v4.0 (TNRS)<sup>56,57</sup>, replacing invalid synonyms with their valid names (e.g., *Acacia farnesiana* was replaced with *Vachellia farnesiana*). For the few records for which information was missing in the TNRS, we referred to the Integrated Taxonomic Information System (ITIS) Catalogue of Life<sup>58</sup>.

Although we obviously had no direct control of the raw tree occurrence data collected by individual countries, the fact that all surveys were conducted by trained professional staff using standardised protocols<sup>52</sup> ensures data reliability. Nevertheless, we evaluated the overall meaningfulness of the dataset by comparing its biogeographical consistency with established knowledge. For this, we used a recent method based on network analysis<sup>59</sup> to identify biogeographical regions for tree species in Europe (Fig. 4, left panel), applying the online tool Infomap Bioregions<sup>60</sup> (<http://bioregions.mapequation.org>) at a spatial resolution of 0.5 degrees. This revealed a strong consistency with the biogeographical regions presented in Rueda *et al.*<sup>61</sup> (Fig. 4, central panel) that, similarly to our approach, are based exclusively on tree species data. However, our results are also consistent with other classifications that take into account other variables such as climate, soil, and land cover<sup>62–64</sup>. The main differences between our classification and that of the European Environment Agency (EEA, latest version<sup>62</sup>) can be observed in Central-Eastern Europe and Fennoscandia (Fig. 4, right panel), where we could not distinguish between boreal and hemiboreal regions, probably due to the high presence of coniferous taxa in the latter. However, similarly to our classification, the EEA biogeographical regions show a distinct transition from West to East that is lacking from Rueda *et al.*<sup>61</sup> map. Interestingly, the EU-Forest data led to the identification of three clearly distinct biogeographical Mediterranean regions (Iberia, Italy and Cyprus), and of an additional region in the British Isles. These distinctions, although absent from the EEA classification, are consistent with the independent analyses carried out by Casalegno *et al.*<sup>16</sup>, Barbati *et al.*<sup>63</sup> and Metzger *et al.*<sup>64</sup>. The overall good fit between our biogeographical classification and previous studies provides indirect support to the soundness of our dataset, and highlights its potential for biogeographical studies and species distribution modelling.

We performed an additional validation by assessing the range (in terms of EOO) of each species that had at least three occurrences as alpha ( $\alpha$ ) shapes. This technique helps to identify the EEO of species with different degrees of restrictiveness, modulated by a single parameter ( $\alpha$ ). Large  $\alpha$  values lead to EEOs close to the convex hull defined by the target species' occurrences (in a two-dimensional space, given a set of points, the convex hull is the smallest convex polygon containing all points), while small  $\alpha$  values generate EEOs close to the original set of points. Note that an EEO identified by the  $\alpha$ -shape procedure may consist of disjointed polygons (this will be most likely the case for small  $\alpha$  values), and may or may not include all of the occurrences.

We used an alpha level of 6 degrees, as recommended by Garcia-Rosello *et al.*<sup>65</sup>, using the ModestR software<sup>66</sup>. We then identified as possible outliers the occurrences falling outside the EEOs. Estimating EEOs using concave hulls is recommended over the standard use of convex hulls, because the former can better approximate species range by excluding discontinuities<sup>67</sup>. Following this analysis, we associated a binary field with each record in the dataset, indicating whether or not the target record fell within the corresponding species' EEO. The EEOs for each species are provided as individual shapefiles (Data Citation 1).

### References

1. Ozanne, C. M. P. *et al.* Biodiversity meets the atmosphere: A global view of forest canopies. *Science* **301**, 183–186 (2003).
2. CBD. *The Convention on Biological Diversity* (Secretariat of the Convention on Biological Diversity, 2008).
3. Bengtsson, J., Nilsson, S. G., Franc, A. & Menozzi, P. Biodiversity, disturbances, ecosystem function and management of European forests. *Forest Ecology and Management* **132**, 39–50 (2000).
4. Kerley, G. I. H., Kowalczyk, R. & Crooms, J. P. G. M. Conservation implications of the refugee species concept and the European bison: king of the forest or refugee in a marginal habitat? *Ecography* **35**, 519–529 (2012).
5. Kowalczyk, R. *et al.* Influence of management practices on large herbivore diet—Case of European bison in Białowieża primeval forest (Poland). *Forest Ecology and Management* **261**, 821–828 (2011).
6. Lang, C. & Polle, A. Ectomycorrhizal fungal diversity, tree diversity and root nutrient relations in a mixed central European forest. *Tree Physiology* **31**, 531–538 (2011).
7. Alberdi Asensio, I. *et al.* *State of Europe's forests 2015* (Ministerial Conference on the Protection of Forests in Europe (FOREST EUROPE, Madrid, Spain, 2015).
8. Fares, S., Mugnozza, G. S., Corona, P. & Palahí, M. Sustainability: five steps for managing Europe's forests. *Nature* **519**, 407–409 (2015).
9. Ruiz-Labourdette, D., Schmitz, M. F. & Pineda, F. D. Changes in tree species composition in Mediterranean mountains under climate change: Indicators for conservation planning. *Ecological Indicators* **24**, 310–323 (2013).
10. Lindbladh, M., Fraver, S., Edvardsson, J. & Felton, A. Past forest composition, structures and processes—how paleoecology can contribute to forest conservation. *Biological Conservation* **168**, 116–127 (2013).

11. Seidl, R., Schelhaas, M.-J., Rammer, W. & Verkerk, P. J. Increasing forest disturbances in Europe and their impact on carbon storage. *Nature Climate Change* **4**, 806–810 (2014).
12. Glick, H. B. *et al.* Spatially-explicit models of global tree density. *Scientific Data* **3**, 160069 (2016).
13. Council of the European Union. Regulation (EC) no 2152/2003 of the European parliament and of the council of 17 November 2003 concerning monitoring of forests and environmental interactions in the community (Forest Focus). *Official Journal of the European Union* **1**, 1–8 (2003).
14. Hiederer, R., Houston Durrant, T. & Micheli, E. *Evaluation of BioSoil demonstration project - Soil data analysis*, vol. 24729 of EUR—Scientific and Technical Research (Publications Office of the European Union, 2011).
15. Houston Durrant, T., San-Miguel-Ayán, J., Schulte, E. & Suarez Meyer, A. *Evaluation of BioSoil demonstration project: forest biodiversity—Analysis of biodiversity module*, vol. 24777 of EUR—Scientific and Technical Research (Publications Office of the European Union, 2011).
16. Casalegno, S., Amatulli, G., Bastrup-Birk, A., Durrant, T. & Pekkari, A. Modelling and mapping the suitability of European forest formations at 1-km resolution. *European Journal of Forest Research* **130**, 971–981 (2011).
17. Ordóñez, A. & Svenning, J.-C. Strong paleoclimatic legacies in current plant functional diversity patterns across Europe. *Ecology and Evolution* **6**, 3405–3416 (2016).
18. Hanewinkel, M., Cullmann, D. A., Schelhaas, M.-J., Nabuurs, G.-J. & Zimmermann, N. E. Climate change may cause severe loss in the economic value of European forest land. *Nature Climate Change* **3**, 203–207 (2012).
19. Hanewinkel, M., Cullmann, D. A., Michiels, H.-G. & Kändler, G. Converting probabilistic tree species range shift projections into meaningful classes for management. *Journal of Environmental Management* **134**, 153–165 (2014).
20. Brus, D. J. *et al.* Statistical mapping of tree species over Europe. *European Journal of Forest Research* **131**, 145–157 (2012).
21. Falk, W. & Hempelmann, N. Species favourability shift in Europe due to climate change: A case study for *fagus sylvatica* L. and *picea abies* (L.) karst. based on an ensemble of climate models. *Journal of Climatology* **2013**, 1–18 (2013).
22. Lindner, M. *et al.* Climate change and European forests: What do we know, what are the uncertainties, and what are the implications for forest management? *Journal of Environmental Management* **146**, 69–83 (2014).
23. Warton, D. I. *et al.* So many variables: Joint modeling in community ecology. *Trends in Ecology & Evolution* **30**, 766–779 (2015).
24. Meessenburg, H. *et al.* Long-term changes of ecosystem services at Solling, Germany: Recovery from acidification, but increasing nitrogen saturation? *Ecological Indicators* **65**, 103–112 (2016).
25. Council of the European Union. Council regulation (EEC) no 3528/86 of 17 November 1986 on the protection of the community's forests against atmospheric pollution. *Official Journal of the European Union* **29**, 2–4 (1986).
26. Jalas, J. & Suominen, J. *Atlas Florae Europaeae: distribution of vascular plants in Europe Vol. 2 Gymnospermae (Pinaceae to Ephedraceae)* (Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, 1973).
27. Jalas, J. & Suominen, J. *Atlas Florae Europaeae: distribution of vascular plants in Europe Vol. 3 Salicaceae to Balanophoraceae* (Committee for Mapping the Flora of Europe and Societas Biologica Fennica Vanamo, 1976).
28. Thuiller, W., Lavorel, S., Araújo, M. B., Sykes, M. T. & Prentice, I. C. Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8245–8250 (2005).
29. Thuiller, W., Lavorel, S., Sykes, M. T. & Araújo, M. B. Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. *Diversity and Distributions* **12**, 49–60 (2006).
30. Svenning, J.-C. & Skov, F. The relative roles of environment and history as controls of tree species composition and richness in Europe. *Journal of Biogeography* **32**, 1019–1033 (2005).
31. Svenning, J.-C. & Skov, F. Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology letters* **10**, 453–460 (2007).
32. Svenning, J.-C. *et al.* Geography, topography, and history affect realized-to-potential tree species richness patterns in Europe. *Ecography* **33**, 1070–1080 (2010).
33. Araújo, M. B., Thuiller, W., Williams, P. H. & Reginster, I. Downscaling European species atlas distributions to a finer resolution: implications for conservation planning. *Global Ecology and Biogeography* **14**, 17–30 (2005).
34. European commission, Joint Research Centre. Framework service contract number 384104 for the provision of forest data and services in support to the European Forest Data Centre (2008).
35. European Commission. *Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions—Our Life Insurance, Our Natural Capital: An EU Biodiversity Strategy To 2020*, COM(2011) 244 final (Communication from the Commission to the Council and the European Parliament, 2011).
36. European Commission. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions—A new EU forest strategy: for forests and the forest-based sector*, COM (2013) 659 final (Communication from the Commission to the Council and the European Parliament, 2013).
37. Mori, A. S., Lertzman, K. P. & Gustafsson, L. Biodiversity and ecosystem services in forest ecosystems: a research agenda for applied forest ecology. *J Appl Ecol* **48**, 280–292 (2016).
38. Gao, T., Nielsen, A. B. & Hedblom, M. Reviewing the strength of evidence of biodiversity indicators for forest ecosystems in Europe. *Ecological Indicators* **57**, 420–434 (2015).
39. Pacifici, M. *et al.* Assessing species vulnerability to climate change. *Nature Climate Change* **5**, 215–224 (2015).
40. Willis, K. J., Bailey, R. M., Bhagwat, S. A. & Birks, H. J. Biodiversity baselines, thresholds and resilience: testing predictions and assumptions using palaeoecological data. *Trends in ecology & evolution* **25**, 583–591 (2010).
41. Reyer, C., Bugmann, H., Nabuurs, G.-J. & Hanewinkel, M. Models for adaptive forest management. *Reg Environ Change* **15**, 1483–1487 (2015).
42. Santini, A. *et al.* Biogeographical patterns and determinants of invasion by forest pathogens in Europe. *New Phytology* **197**, 238–250 (2013).
43. Pautasso, M., Schlegel, M. & Holdenrieder, O. Forest health in a changing world. *Microbial Ecology* **69**, 826–842 (2015).
44. Barredo, J. I. *et al.* Assessing the potential distribution of insect pests: case studies on large pine weevil (*Hyllobius abietis* L.) and horse-chestnut leaf miner (*Cameraria ohridella*) under present and future climate conditions in European forests. *EPPO Bulletin* **45**, 273–281 (2015).
45. Strona, G., Mauri, A. & San-Miguel-Ayán, J. Far from naturalness: how much does spatial ecological structure of European tree assemblages depart from Potential Natural Vegetation? *PLOS ONE* doi:10.1371/journal.pone.0165178 (in the press).
46. Gamfeldt, L. *et al.* Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Communications* **4**, 1340 (2013).
47. Watson, J. *et al.* Large-scale forest inventories of the United States and China reveal positive effects of biodiversity on productivity. *For. Ecosyst.* **2**, 1–16 (2015).
48. Lisitsyna, O. V., Giesecke, T. & Hicks, S. Exploring pollen percentage threshold values as an indication for the regional presence of major European trees. *Review of Palaeobotany and Palynology* **166**, 311–324 (2011).
49. Brewer, S. *et al.* Late-glacial and Holocene European pollen data. *Journal of Maps* **0**, 1–8 (2016).

50. European Parliament. Directive 2007/2/EC of the European parliament and of the council of 14 march 2007 establishing an infrastructure for spatial information in the European community (INSPIRE). *Official Journal of the European Union* **50**, 1–14 (2007).
51. San-Miguel-Ayanz, J. *et al.* *European atlas of forest tree species* (Publications Office of the European Union, 2016).
52. Tomppo, E. *et al.* in *National Forest Inventories—Pathways for Common Reporting* (eds Tomppo, E., Gschwantner, T., Lawrence, M. & McRoberts, R. E.) 1–18 (Springer, Netherlands, 2010).
53. Chirici, G., Bertini, R., Travaglini, D., Puletti, N., Chiavetta, U. in *National Forest Inventories: Contributions to Forest Biodiversity Assessments, vol. 20 of Managing Forest Ecosystems* (eds Chirici, G., Winter, S. & McRoberts, R. E.) 99–119 (Springer, Netherlands, 2011).
54. Chirici, G. *et al.* in *National Forest Inventories: Contributions to Forest Biodiversity Assessments, vol. 20 of Managing Forest Ecosystems* (eds Chirici, G., Winter, S., McRoberts, R. E.) 121–190 (Springer, Netherlands, 2011).
55. McRoberts, R. E., Tomppo, E. O. & Næsset, E. Advances and emerging issues in national forest inventories. *Scandinavian Journal of Forest Research* **25**, 368–381 (2010).
56. Boyle, B. *et al.* The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics* **14**, 16+ (2013).
57. Whitfield, J. Species spellchecker fixes plant glitches. *Nature* **474**, 263 (2011).
58. Bisby, F. A., Froese, R., Ruggiero, M. A. & Wilson, K. L. Species 2000 and ITIS catalogue of life, annual checklist 2004: indexing the world's known species. Digital resource at [www.catalogueoflife.org/col/](http://www.catalogueoflife.org/col/) (2004).
59. Vilhena, D. A. & Antonelli, A. A network approach for identifying and delimiting biogeographical regions. *Nature Communications* **6**, 6848+ (2015).
60. Edler, D., Guedes, T., Zizka, A., Rosvall, M. & Antonelli, A. Infomap bioregions: Interactive mapping of biogeographical regions from species distributions. *Systematic Biology* **0**, 1–8 (2016).
61. Rueda, M., Rodríguez, M. Á. & Hawkins, B. A. Towards a biogeographic regionalization of the European biota. *Journal of Biogeography* **37**, 2067–2076 (2010).
62. European Environmental Agency, EEA. Biogeographical regions. <http://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3> (2016).
63. Barbati, A., Corona, P. & Marchetti, M. A forest typology for monitoring sustainable forest management: The case of European forest types. *Plant Biosystems—An International Journal Dealing with all Aspects of Plant Biology* **141**, 93–103 (2007).
64. Metzger, M. J. *et al.* A high-resolution bioclimate map of the world: a unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography* **22**, 630–638 (2013).
65. García-Roselló, E. *et al.* ModestR: a software tool for managing and analyzing species distribution map databases. *Ecography* **36**, 1202–1207 (2013).
66. García-Roselló, E. *et al.* Can we derive macroecological patterns from primary global biodiversity information facility data? *Global Ecology and Biogeography* **24**, 335–347 (2015).
67. Burgman, M. A. & Fox, J. C. Bias in species range estimates from minimum convex polygons: implications for conservation and options for improved planning. *Animal Conservation* **6**, 19–28 (2003).
68. Leroy, B., Canard, A. & Ysnel, F. Integrating multiple scales in rarity assessments of invertebrate taxa. *Diversity Distribution* **19**, 794–803 (2013).

## Data Citation

1. Mauri, A., Strona, G. & San-Miguel-Ayanz, J. *Figshare* <http://dx.doi.org/10.6084/m9.figshare.c.3288407> (2016).

## Acknowledgements

The views expressed are purely those of the writers and may not under any circumstances be regarded as stating an official position of the European Commission. The authors would like to thank all of the National Forest Inventories that provided the forest data on which our analyses are based. We also thank María del Rosario Rueda for providing the data needed for the tree biogeographical comparison and Gráinne Mulhern for proofreading the manuscript.

## Author Contributions

All the authors collaborated to design the study and wrote the paper. A.M. and G.S. performed the analysis and data validation.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Mauri, A. *et al.* EU-Forest, a high-resolution tree occurrence dataset for Europe. *Sci. Data* **4**:160123 doi: 10.1038/sdata.2016.123 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0>

Metadata associated with this Data Descriptor is available at <http://www.nature.com/sdata/> and is released under the CC0 waiver to maximize reuse.

© The Author(s) 2017