

# Unveiling African rainforest composition and vulnerability to global change

<https://doi.org/10.1038/s41586-021-03483-6>

Received: 29 October 2019

Accepted: 22 March 2021

Published online: 21 April 2021

 Check for updates

Maxime Réjou-Méchain<sup>1✉</sup>, Frédéric Mortier<sup>2,3</sup>, Jean-François Bastin<sup>1,2,3,4</sup>, Guillaume Cornu<sup>2,3</sup>, Nicolas Barbier<sup>1</sup>, Nicolas Bayol<sup>5</sup>, Fabrice Bénédet<sup>2,3</sup>, Xavier Bry<sup>6</sup>, Gilles Dauby<sup>1</sup>, Vincent Deblauwe<sup>7,8</sup>, Jean-Louis Doucet<sup>4</sup>, Charles Doumenge<sup>2,3</sup>, Adeline Fayolle<sup>4</sup>, Claude Garcia<sup>2,3,9</sup>, Jean-Paul Kibambe Lubamba<sup>10,11</sup>, Jean-Joël Loumeto<sup>12</sup>, Alfred Ngomanda<sup>13</sup>, Pierre Ploton<sup>1</sup>, Bonaventure Sonké<sup>14</sup>, Catherine Trottier<sup>6,15</sup>, Ruppert Vimal<sup>16</sup>, Olga Yongo<sup>17</sup>, Raphaël Pélissier<sup>1</sup> & Sylvie Gourlet-Fleury<sup>2,3</sup>

Africa is forecasted to experience large and rapid climate change<sup>1</sup> and population growth<sup>2</sup> during the twenty-first century, which threatens the world's second largest rainforest. Protecting and sustainably managing these African forests requires an increased understanding of their compositional heterogeneity, the environmental drivers of forest composition and their vulnerability to ongoing changes. Here, using a very large dataset of 6 million trees in more than 180,000 field plots, we jointly model the distribution in abundance of the most dominant tree taxa in central Africa, and produce continuous maps of the floristic and functional composition of central African forests. Our results show that the uncertainty in taxon-specific distributions averages out at the community level, and reveal highly deterministic assemblages. We uncover contrasting floristic and functional compositions across climates, soil types and anthropogenic gradients, with functional convergence among types of forest that are floristically dissimilar. Combining these spatial predictions with scenarios of climatic and anthropogenic global change suggests a high vulnerability of the northern and southern forest margins, the Atlantic forests and most forests in the Democratic Republic of the Congo, where both climate and anthropogenic threats are expected to increase sharply by 2085. These results constitute key quantitative benchmarks for scientists and policymakers to shape transnational conservation and management strategies that aim to provide a sustainable future for central African forests.

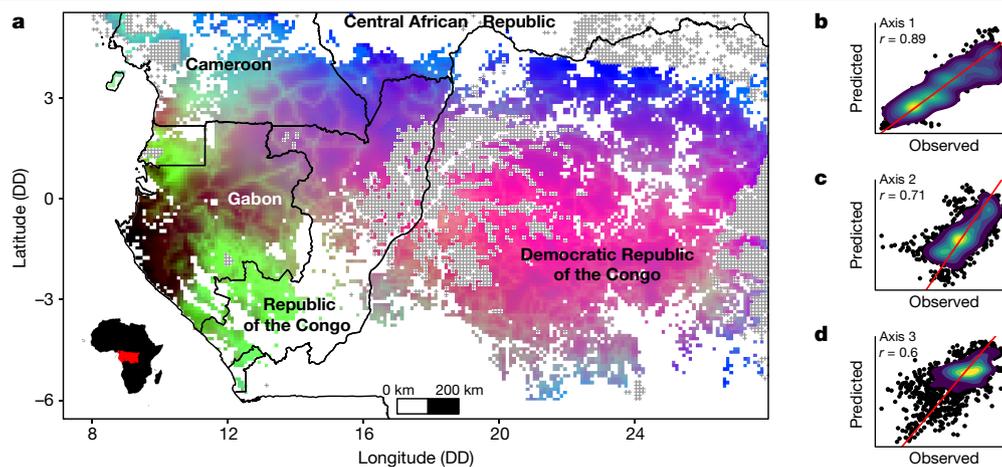
Concomitant increases in climate stress, human population needs and resource extraction in central Africa raise environmental concerns<sup>3–5</sup>. These threats may have considerable impacts on the carbon budget<sup>6</sup>, climate<sup>7</sup> and biodiversity of central African forests<sup>8</sup>, which shelter some of the world's most iconic wildlife species and which are already experiencing a much drier and seasonal climate than other tropical forests<sup>9</sup>. However, the current composition of central African forests and its determinants at regional scale are still poorly known, often being studied in limited areas<sup>10–12</sup> and datasets<sup>13</sup> or at a very coarse grain with heterogeneous occurrence records<sup>14</sup>. Vast regions of central African forests remain poorly explored scientifically<sup>15</sup>, and most space-borne systems of Earth observation provide very limited information on forest composition<sup>16</sup>. This hinders our ability to understand how the composition and functions of forests vary regionally, to forecast

how these forests will face upcoming global changes and, ultimately, to anticipate—on scientific bases—how to protect and manage them beyond national boundaries.

In this study, we used an extensive dataset of forest inventories to (1) model the main floristic and functional gradients over central African forests; and (2) assess their expected vulnerability under forecasted conditions of global (climatic and anthropogenic) change. We compiled the abundance distributions of 193 dominant tree taxa in 185,665 field plots (around 90,000 ha) from commercial forest inventories spread over the 5 main forested countries in central Africa (Cameroon, Central African Republic, Democratic Republic of the Congo, Gabon and Republic of the Congo) (Extended Data Fig. 1). We modelled the joint distributions of taxon abundances at 10-km resolution using supervised component generalized linear regression (SCGLR)<sup>17</sup>,

<sup>1</sup>AMAP, Univ. Montpellier, IRD, CNRS, CIRAD, INRAE, Montpellier, France. <sup>2</sup>CIRAD, Forêts et Sociétés, Montpellier, France. <sup>3</sup>Forêts et Sociétés, Univ. Montpellier, CIRAD, Montpellier, France.

<sup>4</sup>TERRA Teaching and Research Centre, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium. <sup>5</sup>FRM Ingénierie, Maugeio, France. <sup>6</sup>IMAG, CNRS, Univ. Montpellier, Montpellier, France. <sup>7</sup>Center for Tropical Research, Institute of the Environment and Sustainability, University of California, Los Angeles, Los Angeles, CA, USA. <sup>8</sup>International Institute of Tropical Agriculture, Yaoundé, Cameroon. <sup>9</sup>Forest Management and Development Team, Department of Environmental System Science, Institute of Integrative Biology, ETH Zürich, Zürich, Switzerland. <sup>10</sup>Faculté des Sciences Agronomiques, Département de Gestion des Ressources Naturelles, Université de Kinshasa, Kinshasa, Democratic Republic of the Congo. <sup>11</sup>Wildlife Conservation Society, Kinshasa, Democratic Republic of the Congo. <sup>12</sup>Faculté des Sciences et Techniques, Université Marien Ngouabi, Brazzaville, Congo. <sup>13</sup>Institut de Recherche en Écologie Tropicale (IRET/CENAREST), Libreville, Gabon. <sup>14</sup>Plant Systematic and Ecology Laboratory, Higher Teacher's Training College, University of Yaoundé I, Yaoundé, Cameroon. <sup>15</sup>AMIS, Univ. Paul Valéry Montpellier 3, Montpellier, France. <sup>16</sup>GEODE UMR 5602, CNRS, Université Jean-Jaurès, Toulouse, France. <sup>17</sup>Laboratoire de Biodiversité Végétale et Fongique, Faculté des Sciences, Université de Bangui, Bangui, Central African Republic. ✉e-mail: maxime.rejou@ird.fr



**Fig. 1 | Floristic composition of central African forests.** **a**, Spatialized RGB composition of the three first axes of a correspondence analysis (CA) performed on jointly predicted taxon abundances at 10-km resolution ( $n = 193$  taxa; axis 1, blue; axis 2, red; axis 3, green). Grey crosses represent forested areas outside the calibration domain, including permanently flooded forests,

and country boundaries are represented in black. DD, decimal degrees. **b–d**, Cross-validation results comparing the observed and predicted CA gradients for axis 1 (**b**), axis 2 (**c**) and axis 3 (**d**). The 1:1 line is displayed in red. Taxon CA planes 1–2 and 1–3 are provided in Extended Data Fig. 2.

a modelling method that extends partial least-squares regression to the multivariate generalized linear framework. SCGLR models a set of responses (here the abundances of taxa) from synthetic orthogonal explanatory components derived from 24 climatic variables (hereafter, climatic components, CCs) and additional soil type (here, sand versus clay) and anthropogenic pressure covariates. We developed for this study an index, based on population density and road networks, that is specifically designed and calibrated to predict the intensity of recent human-induced forest disturbances in central Africa (see Methods). Finally, thanks to the very large size of the dataset, the predicted floristic and functional gradients were cross-validated with spatially independent observations using Spearman correlation coefficients,  $\rho_{CV}$ .

### Floristic composition in central Africa

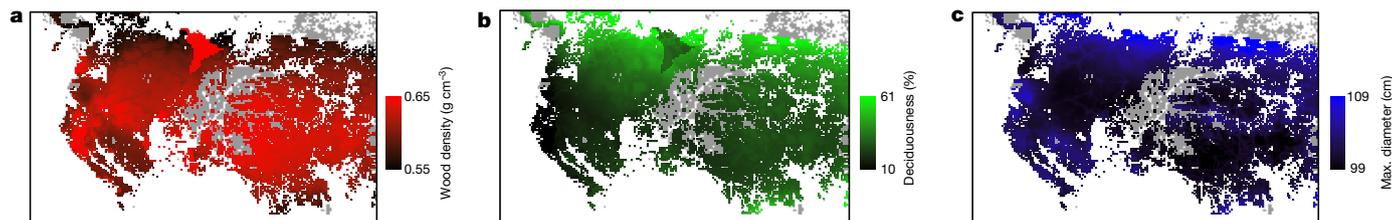
Our model predicted individual taxon abundances with an overall median correlation  $\rho_{CV}$  of 0.48 (range of  $-0.11$  to  $0.83$ ). This median value was still as high as 0.45 when unoccupied sites were removed, showing that, beyond presence and absence, our model also captured variations in abundance within a taxon's distributional range. A correspondence analysis (CA) performed on the predicted taxon abundances revealed major regional floristic gradients (Fig. 1, Extended Data Figs. 2, 3) that were highly correlated with the observed gradients ( $\rho_{CV} = 0.89$ ,  $0.71$  and  $0.6$  for CA axes 1, 2 and 3, respectively; Fig. 1b–d). Contrary to Amazonian and Southeast Asian forests, in which soil was shown to be the primary large-scale driver of tree community composition<sup>18,19</sup>, the most prominent floristic gradient predicted here (CA axis 1) was highly related to climate, and in particular to the first predictive CC (Pearson's  $r = 0.98$ ), contrasting areas with a cool and light-deficient<sup>20</sup> dry season (coastal Gabon) and areas with high evapotranspiration rates (northern limit of the central African forests; Extended Data Fig. 4). The second predicted floristic gradient (CA axis 2) was highly correlated with the two other CCs ( $r = -0.86$  and  $-0.72$  for CC2 and CC3, respectively) related to seasonality and maximum temperature, thus contrasting equatorial areas with a low water deficit and areas with a high water deficit towards the limits of the tropics. Climate seasonality was also found to be a major driver of tree community composition in Amazonia<sup>18</sup>, and maximum temperature has recently been identified as the most important pantropical driver of forest biomass, affecting woody productivity<sup>21</sup>. The third predicted floristic gradient (CA axis 3) revealed floristic variations that are more local and that highlighted

human-disturbed forests ( $r = 0.67$  with our index of human-induced forest-disturbance intensity).

As already shown in previous studies<sup>22,23</sup>, the association between taxon distributions and climate patterns may appear by chance because both are spatially autocorrelated at the regional scale. We thus used a spatially explicit null model that randomized the predictive CCs while preserving their spatial (co)structures. When keeping the soil type and human impact on forests unchanged, the null model predicted abundances with a similar predictive power to the model based on the original CCs for 67% of the taxa ( $P > 0.1$ ). This suggests that variation in taxon abundances directly depends on climate for a minimum of only one-third of the taxa, whereas for most of them, abundance may correlate with climate by chance only. By contrast, the association between climate and the main gradients of floristic assemblages was robust to autocorrelation artefacts ( $P = 0.028$ ,  $0.006$  and  $0.06$  for CA1, CA2 and CA3, respectively). This result confirms that extrapolating assemblages from climate variables is more reliable than extrapolating individual taxon abundances<sup>24</sup>. Indeed, individual taxon abundances are likely to be less predictable on the basis of only current drivers as they are also affected by unknown past human disturbances<sup>25</sup>, biotic interactions and biogeographical history<sup>26</sup>, the idiosyncratic effects of which tend to average out at the community level.

### Functional composition in central Africa

From the predicted floristic assemblages, we computed the community weighted mean<sup>27</sup> of three functional traits that are known to have a central role in ecosystem processes: wood density, deciduousness and maximum diameter (Fig. 2). The predicted functional composition was consistent with the observations ( $\rho_{CV} = 0.47$ ,  $0.75$  and  $0.45$  for the three traits, respectively; Extended Data Fig. 5). As in Amazonia<sup>18</sup>, community wood density varied with soil type, with the highest values found for sandy soils that are located at the boundaries of Cameroon, the Republic of the Congo and the Central African Republic, and where tree species with conservative resource-use strategies predominate<sup>11</sup>. However, larger-scale variations in wood density were primarily driven by human-induced forest disturbances; community wood density was lower in human-disturbed forests, indicating that they are mostly composed of fast-growing taxa<sup>28</sup>. Notably, these areas also contain a high proportion of trees that can potentially reach a large diameter. These two results indicate that human-disturbed forests tend to be



**Fig. 2 | Predicted functional composition of central African forests.** a–c, Predicted community weighted functional trait values (wood density (a), deciduousness (b) and maximum diameter (c)) at 10-km resolution.

dominated by long-lived pioneer taxa, which are characterized by a low wood density but a large potential stature and thus offer a fast and relatively long-lasting carbon sink potential in the absence of disturbances<sup>29</sup>. Finally, a marked deciduousness gradient ran from the highly evergreen forests of coastal Gabon to the northern limit of the central African forests with, again, a well-known exception on the northern sandy soil plateau<sup>11,30</sup>.

### A reference map of forest types

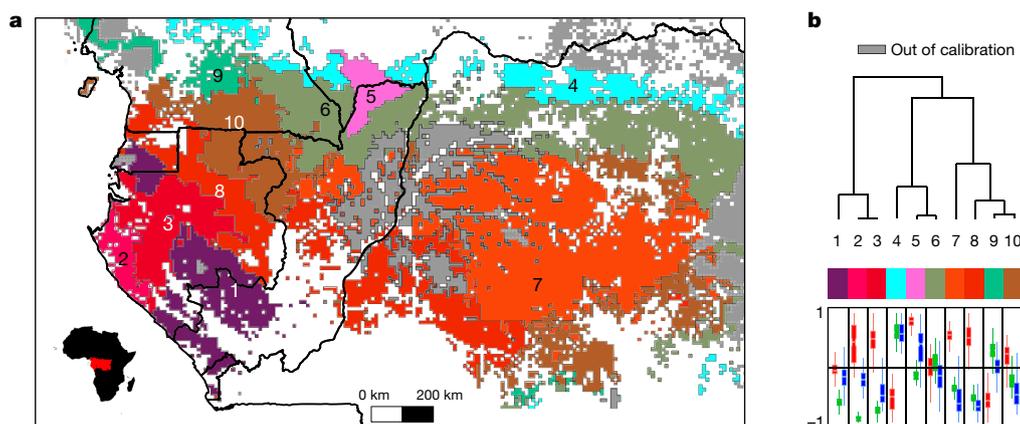
To ease practical applications, we performed hierarchical clustering of the predicted floristic gradients (pixel scores on the first five CA axes), which are continuous by nature, and identified ten major types of forest (Fig. 3; Extended Data Table 1). The strongest floristic dissimilarity appeared between Atlantic forests (types 1–3) and the other forest types (4–10), within which semi-deciduous seasonal forests were clearly distinguished (types 4–6). We also observed functional convergences among floristically dissimilar types of forest and vice versa. For example, despite having a regional species pool similar to deciduous forests (types 4 and 6), sandstone forests (type 5) have a functional composition that is closer to remote forest groups (for example, types 2, 3, 7 and 8), with a high wood density and low deciduousness. Soil filtering modifies the relative abundance of species (rather than their presence or absence<sup>31</sup>), favouring suitable functional attributes in poor sandy soils<sup>11</sup>. By contrast, although Atlantic forests (types 1–3) have little taxonomic affinity with the east–central and southern forests (types 7 and 8), they show a similar functional composition owing to

climate conditions that are more similar, as represented on the first predictive CC (Extended Data Table 1). This confirms that although taxonomic composition has an important biogeographical component, the functional composition of tree communities can converge in similar environmental conditions.

### Vulnerability to global change

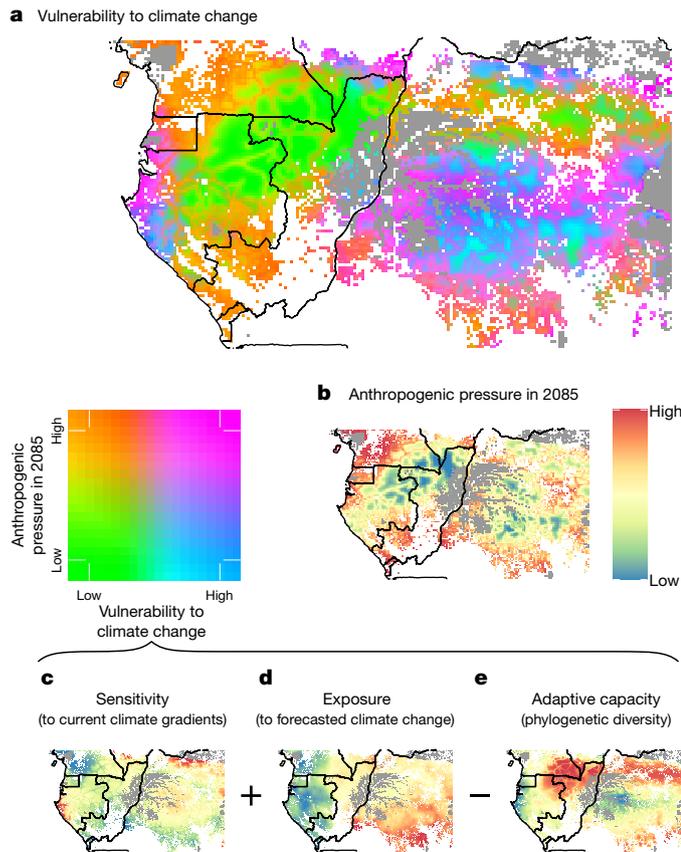
For the 10 forest types, most climate models predict current climate conditions either to virtually disappear from central Africa (for example, types 2 and 5; Extended Data Fig. 6), or to move at spatial and temporal scales that are incommensurate with tree dispersal ability (for example, types 4 and 6). This suggests that current forest communities will not be able to track their present climate envelopes and will face the emergence of novel climates, making the prediction of taxon distributions under future climate projection highly risky<sup>32</sup>. We thus assessed the vulnerability of central African forests to climate change through their sensitivity, exposure and adaptive capacity, following the recommendation of the Intergovernmental Panel on Climate Change (IPCC)<sup>33</sup>.

We quantified sensitivity at the community level using the inverse of the current climate niche breadth of taxa (Fig. 4c) and assuming that assemblages dominated by taxa with narrow environmental tolerances will be more vulnerable to upcoming changes<sup>34</sup>. Sensitivity appeared to be high in coastal Gabon (type 2), where a high level of species endemism exists<sup>35</sup>, and in the driest northern margin of central African forests. Recent studies consistently showed that drier tropical forests exhibited larger functional changes than wetter forests in response to a long-term



**Fig. 3 | Main forest types across central Africa and their functional composition.** a, Forest-type classification obtained by hierarchical clustering of the predicted floristic gradients. Colours represent an RGB composite of the mean values of the three functional traits per forest type (see Fig. 2); that is, wood density (red), deciduousness (green) and maximum diameter (blue). Thus, similar colours illustrate a similar functional composition. b, Taxonomic relationships among the forest types illustrated by a clustering dendrogram (top) and a box plot of the standardized predicted functional composition over

the 12,295 grid cells (bottom), with wood density in red, deciduousness in green and maximum diameter in blue (median is reported at the centre, the lower and upper hinges correspond to the first and third quartiles and the two whiskers extend from these two quartiles to the largest and smallest values, at most 1.5 times the interquartile range from the hinge). Forest-type names and additional information are provided in Extended Data Table 1. Clustering uncertainty is reported in Supplementary Fig. 1.



**Fig. 4 | Predicted vulnerability of central African tree communities to global changes.** **a**, Composite map of the vulnerability to climate change and of the forecasted human-induced forest-disturbance intensity by 2085. Areas in magenta are predicted to be the most vulnerable to both climate change and anthropogenic pressure; areas in green are predicted to be the least vulnerable to both climate change and anthropogenic pressure; areas in blue are predicted to be the most vulnerable to climate change but the least vulnerable to anthropogenic pressure; and areas in orange are predicted to be the least vulnerable to climate change but the most vulnerable to anthropogenic pressure. **b**, Projected human-induced forest-disturbance intensity in 2085. **c–e**, Vulnerability to climate change was estimated as the sensitivity to current climate (c) plus the exposure to forecasted climate changes by 2085 (under the RCP 4.5 scenario) (d) minus the adaptive capacity of tree communities using phylogenetic diversity as a proxy (e).

drought in west Africa<sup>36</sup> and are likely to be more sensitive to global warming<sup>21</sup>. By contrast, forests from northwest Cameroon showed a relatively low sensitivity to current climate conditions, probably because these forests are dominated by widespread tree taxa that are adapted to anthropogenic pressure (Fig. 2). Long-lived pioneer taxa—typical of these human-disturbed forests—are also expected to be favoured by a possible acceleration in forest dynamics induced by global change<sup>37,38</sup>.

Exposure to climate change was quantified as the extent to which the current climate determinants (CCI–CC3) are expected to change by 2085, using 18 unique bias-corrected climate model combinations (under the IPCC Assessment Report 5 (AR5) RCP 4.5 scenario; see Extended Data Fig. 7 for other scenarios). We found that exposure to climate change was mostly driven by an increase in drought stress and maximum temperature<sup>4,39</sup> (Supplementary Fig. 2). The central and east part of central African forests are predicted to be the most exposed, particularly in the south of the Democratic Republic of the Congo (Fig. 4d). Note, however, that climate-change predictions in central Africa are uncertain because meteorological data for model validation are lacking<sup>4</sup> (Supplementary Fig. 3).

Finally, we assessed the adaptive capacity of tree communities through their evolutionary potential. We first found highly significant niche conservatism along the first two climate components ( $P < 0.002$ ). This indicates that closely related taxa tend to share similar climate niche spaces at the regional scale, and suggests that they could be affected similarly by future climate change. We therefore assumed that higher local phylogenetic diversity provides a wider range of potential responses to novel climate conditions<sup>40</sup>, in a similar manner to the insurance hypothesis<sup>41</sup>. We thus used the phylogenetic diversity of predicted tree assemblages as a proxy of their adaptive capacity to climate change. Undisturbed semi-deciduous and transitional forests (types 6 and 10 in Fig. 3) appeared phylogenetically more diverse than evergreen forests (Fig. 4e). A recent study in Amazonia<sup>42</sup> also found a peak of phylogenetic diversity at an intermediate level of precipitation, at which dry- and wet-adapted lineages are mixing. As expected<sup>43</sup>, we also found that human-disturbed areas tended to have a low phylogenetic diversity.

The resulting vulnerability of tree communities to climate change did not correlate with the expected human impact on forests in 2085 ( $\rho = -0.08$ ), which was assessed here by using country-specific projections of human population growth (Fig. 4a, Extended Data Fig. 8). Vulnerability to climate change is expected to be higher for communities that are dominated by hard-wooded taxa ( $\rho = 0.46$  with wood density, Supplementary Table 1). By contrast, the forecasted human impact on forests is predicted to be higher in already disturbed communities; that is, those that are dominated by light-wooded taxa with a large potential size ( $\rho = -0.4$  and  $0.43$  for wood density and maximum size, respectively). However, because we did not account for the appearance of new roads by 2085, we may have underestimated the effect of future anthropogenic activities in remote, currently undisturbed forests. Vulnerability to both climate change and anthropogenic activities (pink colour in Fig. 4a) is predicted to be high for forests of coastal Gabon, in large areas of forests from Democratic Republic of the Congo, and in the northern margin of the forest domain. Forests from Cameroon and in the south of the Republic of the Congo mostly appear vulnerable owing to the high expected human impact on forests by 2085 (orange patches in Fig. 4a). By contrast, the tri-national Sangha transboundary forest complex and the northeastern part of Gabon appeared to be the least vulnerable area in the region (large green patch in Fig. 4a). Globally, the Democratic Republic of the Congo, where most of the central African forests are located, mainly contains forests that are predicted to be vulnerable to climate change and/or to anthropogenic pressure (blue to pink patches in Fig. 4a).

### Conclusions and perspectives

Although some country-specific vegetation patterns were already suggested by phytogeographers (for example, refs. <sup>44,45</sup>), here we provide what is to our knowledge the first synoptic view of central African forest composition at a fine resolution, based on a vast amount of quantitative data. Unveiling the functional composition of central African forests provides key insights into their functioning, dynamics and carbon uptake potential, and the ways in which they could respond to global change. Accounting for the functional characteristics of forests can considerably reduce uncertainty in large-scale models of vegetation<sup>46</sup> or improve remote sensing approaches; for example, by assimilating large-scale variation in wood density into forest carbon maps<sup>47</sup>. Our maps may also help scientists to design representative sampling to better understand the long-term impact of climate change on tree species and stand dynamics; for example, monitoring underrepresented forest types or areas that are highly vulnerable to climate change.

The forest types and vulnerability maps should guide the development of new land-use plans that preserve the full range of evolutionary and functional potential of today's forests—or, at least, that maintain their connectivity—to attenuate the threats related

to expected changes. In central Africa, protected areas and logging concessions, which together cover almost half of the forest domain (14.9% and 32.2%, respectively; Extended Data Fig. 9), are important to consider in such plans. Protected areas do not equally cover the 10 identified types of forest (4–54%; Extended Data Table 1) and should therefore be extended to reach a better representativity. How estimated vulnerability should be accounted for when designing protected areas, for example, by extending the network in vulnerable areas to minimize the loss of biodiversity, or in areas with low anthropogenic pressure to improve their protection, is subject to debate<sup>48</sup>. Logging concessions can also contribute to the maintenance of forest cover and functions, providing that they are well managed<sup>49,50</sup>, and are likely to act as biodiversity corridors between protected areas at present<sup>51</sup>. However, this will only prove effective in the long term if they strictly comply with legislation and, ideally, with standard certification requirements. These good practices are especially important in forests that are dominated by evergreen taxa with a high wood density, in which disturbances may have a higher impact on community composition. In areas that are expected to be under high anthropogenic pressure, forest connectivity could be preserved by promoting agroforestry and restoration programmes, strictly controlling access to logging roads and stabilizing shifting agriculture<sup>52</sup>. Across central Africa, the highest uncertainties for the future of forests remain in the Democratic Republic of the Congo, where substantial areas that belong to the state are not yet attributed to any land-use category and should warrant particular attention owing to their high vulnerability (Fig. 4).

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-021-03483-6>.

- Diffenbaugh, N. S. & Giorgi, F. Climate change hotspots in the CMIP5 global climate model ensemble. *Clim. Change* **114**, 813–822 (2012).
- United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables*. Working Paper No. ESA/P/WP/248 (2017).
- Malhi, Y., Adu-Bredu, S., Asare, R. A., Lewis, S. L. & Mayaux, P. African rainforests: past, present and future. *Phil. Trans. R. Soc. B* **368**, 20120312 (2013).
- James, R., Washington, R. & Rowell, D. P. Implications of global warming for the climate of African rainforests. *Phil. Trans. R. Soc. B* **368**, 20120298 (2013).
- Abernethy, K., Maisels, F. & White, L. J. Environmental issues in Central Africa. *Annu. Rev. Environ. Resour.* **41**, 1–33 (2016).
- Hubau, W. et al. Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature* **579**, 80–87 (2020).
- De Wasseige, C., Tadoum, M., Atyi, E. & Doumenge, C. *The Forests of the Congo Basin: Forests and Climate Change* (Weyrich, 2015).
- Stévant, T. et al. A third of the tropical African flora is potentially threatened with extinction. *Sci. Adv.* **5**, eaax9444 (2019).
- Parmentier, I. et al. The odd man out? Might climate explain the lower tree  $\alpha$ -diversity of African rain forests relative to Amazonian rain forests? *J. Ecol.* **95**, 1058–1071 (2007).
- Réjou-Méchain, M. et al. Regional variation in tropical forest tree species composition in the Central African Republic: an assessment based on inventories by forest companies. *J. Trop. Ecol.* **24**, 663–674 (2008).
- Réjou-Méchain, M. et al. Tropical tree assembly depends on the interactions between successional and soil filtering processes. *Glob. Ecol. Biogeogr.* **23**, 1440–1449 (2014).
- Fayolle, A. et al. Geological substrates shape tree species and trait distributions in African moist forests. *PLoS One* **7**, e42381 (2012).
- Fayolle, A. et al. Patterns of tree species composition across tropical African forests. *J. Biogeogr.* **41**, 2320–2331 (2014).
- Droissart, V. et al. Beyond trees: biogeographical regionalization of tropical Africa. *J. Biogeogr.* **45**, 1153–1167 (2018).
- Sosef, M. S. et al. Exploring the floristic diversity of tropical Africa. *BMC Biol.* **15**, 15 (2017).
- Parmentier, I. et al. Predicting alpha diversity of African rain forests: models based on climate and satellite-derived data do not perform better than a purely spatial model. *J. Biogeogr.* **38**, 1164–1176 (2011).
- Bry, X., Trottier, C., Verron, T. & Mortier, F. Supervised component generalized linear regression using a PLS-extension of the fisher scoring algorithm. *J. Multivariate Anal.* **119**, 47–60 (2013).
- ter Steege, H. et al. Continental-scale patterns of canopy tree composition and function across Amazonia. *Nature* **443**, 444–447 (2006).
- Slik, J. W. et al. Soils on exposed Sunda shelf shaped biogeographic patterns in the equatorial forests of Southeast Asia. *Proc. Natl Acad. Sci. USA* **108**, 12343–12347 (2011).
- Philippon, N. et al. The light-deficient climates of western Central African evergreen forests. *Environ. Res. Lett.* **14**, 034007 (2019).
- Sullivan, M. J. P. et al. Long-term thermal sensitivity of Earth's tropical forests. *Science* **368**, 869–874 (2020).
- Beale, C. M., Lennon, J. J. & Gimona, A. Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proc. Natl Acad. Sci. USA* **105**, 14908–14912 (2008).
- Deblauwe, V. et al. Remotely sensed temperature and precipitation data improve species distribution modelling in the tropics. *Glob. Ecol. Biogeogr.* **25**, 443–454 (2016).
- Maguire, K. C. et al. Controlled comparison of species- and community-level models across novel climates and communities. *Proc. R. Soc. B* **283**, 20152817 (2016).
- Morin-Rivat, J. et al. Present-day central African forest is a legacy of the 19th century human history. *eLife* **6**, e20343 (2017).
- Ricklefs, R. E. Intrinsic dynamics of the regional community. *Ecol. Lett.* **18**, 497–503 (2015).
- Violle, C. et al. Let the concept of trait be functional! *Oikos* **116**, 882–892 (2007).
- Díaz, S. et al. The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).
- Rüger, N. et al. Demographic trade-offs predict tropical forest dynamics. *Science* **368**, 165–168 (2020).
- Ouédraogo, D.-Y. et al. The determinants of tropical forest deciduousness: disentangling the effects of rainfall and geology in central Africa. *J. Ecol.* **104**, 924–935 (2016).
- Shipley, B. *From Plant Traits to Vegetation Structure: Chance and Selection in the Assembly of Ecological Communities* (Cambridge University Press, 2010).
- Feeley, K. J. & Silman, M. R. Biotic attrition from tropical forests correcting for truncated temperature niches. *Glob. Change Biol.* **16**, 1830–1836 (2010).
- Parry, M. et al. *Climate Change 2007 – Impacts, Adaptation, and Vulnerability: Contribution of Working Group II to the Fourth Assessment Report of the IPCC* (Cambridge University Press, 2007).
- Foden, W. B. et al. Identifying the world's most climate change vulnerable species: a systematic trait-based assessment of all birds, amphibians and corals. *PLoS One* **8**, e65427 (2013).
- Lachenaud, O., Stévant, T., Ikabanga, D., Ndjabounda, E. C. N. & Walters, G. The littoral forests of the Libreville area (Gabon) and their importance for conservation: description of a new endemic species (Rubiaceae). *Plant Ecol. Evol.* **146**, 68–74 (2013).
- Aguirre-Gutiérrez, J. et al. Drier tropical forests are susceptible to functional changes in response to a long-term drought. *Ecol. Lett.* **22**, 855–865 (2019).
- Claeys, F. et al. Climate change would lead to a sharp acceleration of Central African forests dynamics by the end of the century. *Environ. Res. Lett.* **14**, 044002 (2019).
- McDowell, N. G. et al. Pervasive shifts in forest dynamics in a changing world. *Science* **368**, eaaz9463 (2020).
- Zhou, L. et al. Widespread decline of Congo rainforest greenness in the past decade. *Nature* **509**, 86–90 (2014).
- Purvis, A. Phylogenetic approaches to the study of extinction. *Annu. Rev. Ecol. Syst.* **39**, 301–319 (2008).
- Yachi, S. & Loreau, M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proc. Natl Acad. Sci. USA* **96**, 1463–1468 (1999).
- Neves, D. M. et al. Evolutionary diversity in tropical tree communities peaks at intermediate precipitation. *Sci. Rep.* **10**, 1188 (2020).
- Letcher, S. G. Phylogenetic structure of angiosperm communities during tropical forest succession. *Proc. R. Soc. B* **277**, 97–104 (2010).
- Letouzey, R. Notice de la carte phytogéographique du Cameroun au 1:500000 (Institut de la Carte Internationale de la végétation Toulouse-France et Institut de la recherche agronomique (Herbier National) Yaoundé-Cameroun, 1985).
- Boulvert, Y. Carte phytogéographique de la République Centrafricaine (feuille ouest-feuille est) à 1 000 000 (Editions de l'ORSTOM, 1986).
- Fyllas, N. M., Quesada, C. A. & Lloyd, J. Deriving plant functional types for Amazonian forests for use in vegetation dynamics models. *Perspect. Plant Ecol. Evol. Syst.* **14**, 97–110 (2012).
- Mitchard, E. T. A. et al. Markedly divergent estimates of Amazon forest carbon density from ground plots and satellites. *Glob. Ecol. Biogeogr.* **23**, 935–946 (2014).
- Visconti, P., Pressey, R. L., Bode, M. & Segan, D. B. Habitat vulnerability in conservation planning—when it matters and how much. *Conserv. Lett.* **3**, 404–414 (2010).
- Putz, F. E. et al. Sustaining conservation values in selectively logged tropical forests: the attained and the attainable. *Conserv. Lett.* **5**, 296–303 (2012).
- Gourlet-Fleury, S. et al. Tropical forest recovery from logging: a 24 year silvicultural experiment from Central Africa. *Phil. Trans. R. Soc. B* **368**, 20120302 (2013).
- Clark, C. J., Poulsen, J. R., Malonga, R. & Elkan, P. W. Jr. Logging concessions can extend the conservation estate for Central African tropical forests. *Conserv. Biol.* **23**, 1281–1293 (2009).
- Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A. & Hansen, M. C. Classifying drivers of global forest loss. *Science* **361**, 1108–1111 (2018).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

## Methods

### Data reporting

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

### Floristic and functional trait data

Forestry data were extracted from management forest inventories conducted in 105 logging concessions covering around  $1.6 \times 10^5$  km<sup>2</sup> (Extended Data Fig. 1). Most companies followed a standardized inventory protocol similar to that described previously<sup>53</sup>. In most cases, it consisted of continuous and parallel transects 20 m or 25 m wide, often 2–3 km apart, and subdivided into rectangular 0.4- or 0.5-ha plots. Overall, the full dataset had a total of 192,972 plots. Within each plot, trees with a diameter at breast height (DBH)  $\geq 30$  cm were allocated into 10-cm wide diameter classes and identified at the species or genus level whenever possible through either commercial or local names<sup>53</sup>. Independent analyses performed on 298 scientific plots ( $\geq 1$  ha in size) showed that the floristic gradients obtained with large trees are representative of the ones obtained with trees  $\geq 10$  cm in diameter (Pearson  $r > 0.94$ ; Supplementary Fig. 4). Overall, around  $7 \times 10^6$  trees were recorded. Taxonomy was revised and homogenized using the African Flowering Plants Database<sup>54</sup> and the Angiosperm Phylogeny Group III for orders and families<sup>55</sup>. A total of 1,092 taxa were recorded in the original dataset. Extensive botanical controls demonstrated that the patterns of both intra ( $\alpha$ )- and inter ( $\beta$ )- plot diversity extracted from these data were highly reliable<sup>53</sup>.

For the purpose of the present paper, we conducted an additional assessment according to botanical experts and by comparing the distributional range of our taxa with that in other datasets<sup>54,56</sup> to select a set of species and genera deemed to be reliably identified over the whole study area ( $n = 195$ ). The abundances of these taxa were aggregated in  $10 \times 10$ -km<sup>2</sup> grid cells across the study area, but we only kept the taxa occurring in at least 5% of the cells to discard taxa that cannot be studied at the regional scale ( $n = 2$ ). For the statistical analyses, we kept the  $10 \times 10$ -km<sup>2</sup> grid cells having a field plot sampling area  $\geq 10$  ha and where the selected taxa represented at least 75% of the total number of individuals originally inventoried to ensure that our dataset was representative of the within-cell tree community composition. The final dataset contains  $6.1 \times 10^6$  tree individuals belonging to 193 taxa, of which 96 were analysed at the species and 97 at the genus levels (Supplementary Table 2), recorded in 185,665 plots aggregated in  $1,57110 \times 10$ -km<sup>2</sup> grid cells. Overall, the selected taxa represented 90% of the total number of individuals originally inventoried in the selected grid cells.

For each taxon, we compiled information on three functional traits. First, we extracted an average wood density using the global wood density database<sup>57,58</sup> as well as other wood density data<sup>59</sup>. Wood density is an integrative trait that reflects a trade-off between tree growth potential and mortality risk<sup>28</sup> and is thus highly informative on community dynamics<sup>60</sup>. It ultimately directly affects the amount of carbon that can be stored in the vegetation<sup>61</sup>. Second, we extracted the leaf phenology (deciduous or evergreen) of all taxa from the large unpublished CoForTraits database<sup>62</sup>. This database compiles information on more than 1,000 species from central Africa with values extracted from the literature (mostly from local floras, academic papers and unpublished theses). When several values were available for a given species from different sources, we attributed the one with the maximum of occurrences (ambiguities were left as unknown). At the genus level, we first computed this step for all species belonging to the genus and then attributed the phenology with the maximum of occurrences at the species level to the genus so that all congeneric species have the same weight in the phenology attribution. This approach relies on the assumption that leaf phenological traits are highly phylogenetically conserved<sup>63</sup>. For a few taxa ( $n = 5$ ), the phenology information was

obtained from Ouédraogo et al.<sup>30</sup> and following these authors we considered *Lophira alata* Banks ex C. F. Gaertn. and *Musanga cecropioides* R. Br. as leaf exchangers; that is, with a trait value of 0.5, intermediate between evergreen (0) and deciduous (1). Leaf phenology is one of the few traits considered in dynamic global vegetation models as it affects the dynamics of forest productivity<sup>64</sup>. In particular, deciduousness indicates that tree photosynthetic activity, and thus growth, is seasonally depressed, which has a direct effect on carbon, water and nutrient cycling<sup>65</sup>. Deciduousness has often been interpreted as a strategy to avoid water stress and is thus expected to depend on climate and soil conditions<sup>30,66</sup>. Finally, we used the original inventory data to calculate the maximum diameter as the 95th percentile value of the diameter distribution for each taxon. Maximum potential diameter, which is often used as a proxy of maximum height<sup>67</sup>, informs both on tree competitive ability for light and on the carbon sequestration potential. At the community level, it is expected to vary along gradients of productivity and disturbance<sup>68</sup>. Leaf phenology was successfully assigned to 89% of the taxa (98% of the individuals), wood density to 91% of the taxa (96% of the individuals) and maximum diameter to all taxa.

### Climate and soil data

We considered 24 climatic predictors derived from the open Climatic Research Unit (CRU) dataset<sup>69</sup> (Extended Data Table 2). We decided to rely on the CRU dataset as other datasets, such as WorldClim<sup>70</sup>, contain erroneous observations for some climatic stations (for example, Ngoundi in Cameroon) that severely affected our model. Furthermore, our cross-validation approach showed that the CRU database led to higher correlations between observed and predicted taxa abundances, correspondence analyses scores and community weighted trait values than the WorldClim<sup>70</sup> and CHIRPS<sup>71</sup> databases (results not shown).

Soil maps have been published at the country scale in central Africa and their homogenization is very challenging. We thus relied on a global dataset, the Harmonized World Soil Database (HWSD)<sup>72</sup>, to attribute a soil type to each grid cell. A cross-validation analysis of our joint distribution model revealed that soil significantly improved predictions, mostly due to the contrast between Arenic Acrisols and the other soil types (Supplementary Fig. 5). To keep the model parsimonious and maximize its robustness, we thus merged all soil categories but the Arenic Acrisols soils into a single category and discarded the permanently flooded areas as mapped in the open European Space Agency Climate Change Initiative (ESA-CCI) landcover product (v.1.6), where no tree inventory data were available.

### Human-induced forest-disturbance intensity

Many studies have attempted to spatialize human impacts on the environment at a large scale. In most cases, these human footprint maps have consisted of cumulative threat maps, assuming, for instance, population density and infrastructure effects<sup>73–75</sup>. Moreover, most of these maps relied on population statistics obtained at the level of administrative entities, resulting in human footprint indices with sharp changes at administrative boundaries<sup>76</sup>. We thus developed a statistical model to link the probability for a forest pixel  $i$  to be affected by anthropogenic activities depending on human population density and road proximity through nonlinear relationships. This resulted in a spatially continuous index representing human-induced forest-disturbance intensity that can be projected in space and/or time following predefined scenarios of human population dynamics (Extended Data Fig. 8).

We calibrated this index with the ‘Settlement Points’ dataset produced under the Global Rural Urban Mapping Project (Grumpv1). This dataset provides estimates of human population (counts, in number of people (individuals)) for the year 2000 using a proportional allocation gridding algorithm (1-km<sup>2</sup> grid) based on more than 1,000,000 national and subnational geographic units. Focusing on central Africa, we compared this product with the Natural Earth Populated Places product (v.3.0.0; <http://www.naturalearthdata.com/downloads/>

10m-cultural-vectors/10m-populated-places/; last accessed 7 October 2018) derived from the LandScan (<https://earthworks.stanford.edu/catalog/stanford-yj715rc4110#iso-metadata-reference-info>) dataset (pixels with fewer than 200 individuals per km<sup>2</sup> were discarded). The total number of populated points in central Africa (longitude 5.6 to 39.8, latitude -9.8 to 7.5 in decimal degrees) was 807 and 376 for the Grumpv1 and Natural Earth products, respectively. We thus performed a random manual check of the populated places present in Grumpv1 and absent from Natural Earth (the reverse rarely occurred) using Google Earth images and found that in all cases Grumpv1 was correct. We finally used the Grumpv1 dataset, which mostly provides information on populated places with more than around 1,000 people. Because smaller populations may have a substantial impact on forests, we added to this dataset the populated locations of the category ‘towns’ from OpenStreetMap (<https://data.maptiler.com/downloads/planet/#1.59/-17.3/19.7>; last accessed 2 October 2018), assuming by default that they all contained 500 people (OpenStreetMap does not provide systematic information on population size).

The road network was extracted from the Global Roads Open Access Data Set, v.1 (<https://sedac.ciesin.columbia.edu/data/set/groads-global-roads-open-access-v1>; last accessed 14 September 2018), a dataset combining road data by country. Note that logging roads are not fully represented in this dataset, so we may underestimate their effect in this study. A few roads from the northern Republic of the Congo were corrected using data from OpenStreetMap. Preliminary analyses revealed that further accounting for the railway and river networks did not improve predictions of tree taxon and community distributions.

Our index was thus calculated as followed. Let  $z_i, i = 1, \dots, n$  be  $n$  random variables indicating the disturbance status of pixel  $i$ : 0 if the pixel is undisturbed and 1 if disturbed. We assumed that  $z_i$  is distributed as a Bernoulli variable:

$$z_i = \text{Bern}(p_i), \text{ with } p_i = \frac{\text{IP}_i(\theta)}{\text{IP}_i(\theta) + \text{IR}_i^r},$$

where  $\text{IP}_i(\theta)$  is a synthetic index describing the influence of the population density of all populated places on pixel  $i$  (see below),  $\theta$  is an unknown parameter to be inferred, and  $\text{IR}_i^r$  expresses the road influence on pixel  $i$ , defined as the normalized square root distance of pixel  $i$  to the nearest road  $r$ :

$$\text{IR}_i^r = \frac{\min_{r \in R} \sqrt{\text{DR}_i^r}}{\max_{i=1, \dots, n} \left( \min_{r \in R} \sqrt{\text{DR}_i^r} \right)}$$

where DR denotes the distance to the nearest road in the study area and  $R$  denotes all roads in the study area.

Population influence,  $\text{IP}_i^\theta$ , is defined as the normalized square root of the weighted sum of the population size of place  $j$ . Note that the weight depends on both the distance between pixel  $i$  and populated place  $j$ ,  $\delta_{ij}$ , and on the population size  $N_j$ :

$$\text{IP}_i^\theta = \frac{\sqrt{\sum_j^n N_j e^{-\frac{\delta_{ij}}{\log(N_j)^\theta} + 1}}}{\max_j \sqrt{\sum_j^n N_j e^{-\frac{\delta_{ij}}{\log(N_j)^\theta} + 1}}}$$

We finally calibrated the  $\theta$  parameter using two reference areas of around 190,000 km<sup>2</sup> (Supplementary Fig. 6). These two areas were chosen because they cover contrasting conditions, are well known by our team and were found to be little influenced by atmospheric pollution in the MODIS data. Degraded versus intact forests were identified from a recently published MODIS-based regional vegetation map<sup>20</sup>. Using a likelihood optimization approach in these two areas, we obtained

$\theta = 1.27$  and  $1.71$  in calibration areas 1 and 2, respectively, indicating that under a similar anthropogenic context, forests tend to be disturbed at a greater distance from sources of anthropogenic disturbance in the second calibration area. The final human-induced forest-disturbance intensity index was thus calculated with  $\theta = 1.49$ , the average estimate for the two calibration areas, over the whole central African forest domain, thus avoiding the risk of artefacts related to atmospheric pollution, from which satellite products suffer, especially over Gabon.

This index, built independently from our floristic dataset, outperformed previously published indices to predict floristic composition in our study area. Using a simple linear model, with individual anthropogenic indices as single predictors, the mean wood density of tree communities was better predicted with our new index ( $r = 0.33$ ) than with the WorldPop<sup>77</sup> ( $r = 0.30$ ), LandScan ( $r = 0.15$ ) and Venter<sup>74</sup> ( $r = 0.23$ ) indices. Similarly, using a simple generalized linear model with a Poisson distribution to predict the abundance of *Musanga cecropioides*—the most widespread and abundant short-lived pioneer taxon over central African forests—revealed a better performance of our index ( $r = 0.35$ ) compared to previous indices ( $r = 0.31, 0.11$  and  $0.26$  for WorldPop, LandScan and Venter, respectively).

### Statistical model

To predict the joint taxa distributions we relied on a recently developed methodology called supervised component generalized linear regression (SCGLR)<sup>17</sup>, which identifies the most predictive dimensions among a large set of potentially multicollinear predictors. SCGLR is an extension of partial least-squares regression (PLSR) to the uni- and multivariate generalized linear framework. PLSR is particularly well suited for analysing a large array of correlated predictor variables, and many studies have demonstrated its ability for prediction in various biological fields, such as genetics<sup>78</sup> and ecology<sup>79</sup>. Although PLSR is well adapted for continuous variables, SCGLR is suited for non-Gaussian outcomes and noncontinuous covariates. It is a model-based approach that extends PLSR<sup>80</sup>, principal component analysis (PCA) on instrumental variables<sup>81</sup>, canonical correspondence analysis<sup>82</sup> and other related empirical methods by maximizing a trade-off between goodness of fit of the model and the quantity of information that the components capture from the climatic variables. This information is measured through an indicator of ‘structural relevance’ (SR)<sup>83</sup>, which uses bundles of highly correlated variables to attract components to rich and robust informational dimensions.

The components are sought as  $K$  linear combinations of environmental variables common to all species with coefficient vectors denoted  $u = u_1, \dots, u_K$  (under norm and orthogonality constraints). SCGLR also estimates the corresponding  $q \times K$  (number of species by number of components) matrix of unknown parameters  $\gamma$  to maximize the following convex sum:

$$s \log \psi(u, \gamma) + (1 - s) \log \phi_l(u)$$

where  $\psi$  is the likelihood and  $\phi_l$  is the SR.  $s$  and  $l$  are tuning parameters.  $s$  is related to the trade-off between goodness of fit and structural relevance.  $l$  is a non-negative scalar related to the narrowness of the bundles of climatic variables the components are wanted to align with. The  $K$  climatic components (CCs) are then equal to  $\text{CC}_k = Xu_k, k = 1, \dots, K$  and can be understood as the main environmental directions predicting all species simultaneously, whereas  $\gamma_j, j = 1, \dots, q$  are the magnitude of the effects of the  $K$  components on the abundances of each species. Then, the species abundances of each taxon  $j = 1, \dots, 193$  on the grid cell  $i = 1, \dots, 1,571$  are modelled using a generalized linear Poisson regression such that:

$$y_{ij} \sim P(S_i \lambda_{ij})$$

$$\log(\lambda_{ij}) = X_i \beta_j + T_i \alpha_j = X_i u_j + T_i \alpha_j = \text{CC}_i \gamma_j + T_i \alpha_j$$

# Article

where  $X$  denotes climatic variables (Extended Data Table 2),  $S_i$  is an offset corresponding to the number of plots within each grid cell, and  $T$  is a set of covariates known to affect species abundances: here, the soil type and the human-induced forest-disturbance intensity index, as well as its logarithm to account for nonlinear responses.

The number of components ( $K$ ) as well as the tuning parameters ( $l$  and  $s$ ) must appropriately be chosen. This was done with a 10% cross-validation procedure in which the criterion used was the harmonic mean of the mean square prediction error (MSPE) across the 194 taxa. A dedicated R package, SCGLR<sup>84</sup>, is available (see also <https://github.com/SCnext/SCGLR>).

To assess the predictive power of our model, we performed a leave-one-block-out cross-validation in which our dataset was divided into 40 spatial clusters identified with a Ward's hierarchical clustering<sup>85</sup> of plot coordinates<sup>86</sup> (Supplementary Fig. 7). All clusters but one were used for training the model (that is, calibration dataset) and the remaining cluster was used for validating the model. We repeated this procedure 40 times such that all clusters were used once in the validation dataset and participated in the model assessment. Model validation was performed through the use of nonparametric Spearman's rank correlation coefficients between observations and predictions. For individual taxon abundances, correlations were estimated using observed and predicted abundance per taxon. For taxon assemblages, a correspondence analysis (CA) was performed on the grid cell  $\times$  observed species abundance matrix, providing the observed CA axes. The predicted site scores on each CA axis were then obtained by projecting the grid cell  $\times$  predicted species abundance matrix in the observed CA planes. Correlations were computed on the observed and predicted site scores (that is, loadings) enabling us to assess the ability of our model to predict the main floristic gradients across our area. Finally, for the three functional traits, correlations were estimated on the grid-cell-based community weighted mean (CWM) traits<sup>27</sup> calculated on observed and predicted taxon assemblages.

Taxon abundances and community composition were predicted across the entire study area in a regular 10-km grid. To predict the floristic composition of the existing forests, we first used the ESA-CCI landcover product (v.1.6) to only keep grid cells that are likely to be forested (that is, category 'broadleaved evergreen'). Then, we only selected grid cells that had a combination of predictor values similar to those in the calibration dataset. To do this, we used a three-dimensional (3D) convex hull algorithm on the three climatic components to exclude all the grid cells that had a combination of predictors different from that represented in the calibration dataset. This resulted in 12,295 grid cells, representing 85% of the central African forests; that is, an area of around 1,230,000 km<sup>2</sup>.

We finally used the Ward's hierarchical clustering method to classify the predicted floristic composition into broad floristic types. Group classification was done on the first five axes of a CA performed on predicted taxon abundances, accounting for 77% of the total inertia. The number of retained types was chosen based on our expert knowledge. The uncertainty associated with this classification was then assessed through Gaussian finite mixture models<sup>87</sup> (repeated 500 times) using a spherical, equal volume model (EII).

## Spatially explicit null models

Whenever predictors and observations are spatially structured, model errors of type I (false positive associations) are inflated<sup>88</sup>, hindering our capacity to extrapolate predictions in space or time<sup>22</sup>. We thus built a spatialized null model to test the degree to which the successfulness of our predictions resulted from an actual relationship with climatic variables or was simply due to spatial costructures between taxon distributions and climatic gradients that arose by chance. We used the RGeostats R package<sup>89</sup> to simulate sets of SCGLR CCs having similar spatial properties to those of the observed CCs as well as similar spatial costructures between them. This step consisted of

fitting theoretical variograms and covariograms to empirical ones to simulate random realizations that can be then used as 'null' spatialized predictors (Supplementary Figs. 8, 9). We simulated 500 sets of 'null' spatialized predictors and used them as predictors in our GLMs using the leave-one-block-out cross-validation described above. The resulting correlations between observed and predicted taxon abundances, and axes scores (for taxon assemblages) were finally compared with the correlations obtained when observed climatic predictors were considered. The resulting  $P$  values were calculated as the number of times a simulated correlation was higher than the observed one, divided by the total number of realizations ( $n = 501$ ).

## Forest vulnerability to global change

Vulnerability to climate change, as assessed through the IPCC framework, is composed of three components: (1) sensitivity, (2) exposure and (3) adaptive capacity to climate change.

Sensitivity to climate change, Sensitivity<sub>clim</sub>, was first estimated at the taxon level in a similar way to that described previously<sup>34</sup>. For each taxon, we calculated the mean of the weighted standard deviation (SDw) of the three climatic components (CCs) at the present time, using locally observed taxon abundances as weights. SDw thus represents the width of the climatic niche currently occupied by the taxa. Taxon-specific climate sensitivity was then measured as  $1/\text{SDw}$  (it increases as niche width decreases). To upscale tree sensitivity to climate change at the community level and over our study area, sensitivity was measured as the CWM of taxon-specific climate-sensitivity scores, using predicted taxon assemblages.

Exposure to climate change, Exposure<sub>clim</sub>, was assessed using projected changes in climate from 18 unique climate model combinations provided by the AFRICLIM v3.0 dataset<sup>90</sup> (last accessed 3 February 2020). These models corresponded to pairwise combinations of five regional climate models (RCMs) driven by 10 general circulation models (GCMs), with an unequal number of GCMs models per RCM (10 models for the Swedish Meteorological and Hydrological (SMHI) RCM, four for the Climate Limited-area Modelling Community (CLMCOM) RCM, two for the Royal Netherlands Meteorological Institute (KNMI) RCM, one for the Canadian Centre for Climate Modelling (CCCMA) RCM and one for the Danish Meteorological Institute (DMI) RCM). These models were generated using change-factor downscaling approaches to model spatial variation at local scales while correcting for differences between observed and simulated baseline climates (see ref. <sup>90</sup> for more details). We here concentrated on one representative concentration pathway of the IPCC AR5 (RCP 4.5) for the late 21st century (2071–2100; hereafter named 2085) and reconstructed the three SCGLR selected CCs from the climatic predictions as follows: let  $X_{\text{rcp4.5}}$  be the predicted future climatic conditions and let  $m = \bar{X}$  and  $S = \text{sd}(X)$  be the mean and standard deviation matrices of the current climatic conditions. The predictive climatic components under future scenarios are then equal to  $f_{\text{rcp4.5}} = (X_{\text{rcp4.5}} - m)S\hat{u}$ , where  $\hat{u}$  represents SCGLR CCs. We then calculated the Euclidean distance between the 3 current and the 3 predicted CCs for each of the 18 models and then estimated the exposure to climate change as the mean distance over the 18 models.

Adaptive capacity to climate change, Adaptive<sub>clim</sub>, was assessed through the phylogenetic diversity of predicted assemblages at the genus level. We used a recently published dated phylogeny<sup>91</sup>, covering 167 out of our 180 genera (representing 94% of predicted individuals). We first tested if the studied taxa exhibited a significant conservatism in their climate niches using Abouheif's permutation tests<sup>92</sup> on the taxa-specific score ( $\gamma$ ) values on the three SCGLR climate components ( $\gamma$  represents the influence of a CC on a given taxa distribution; see above). We then measured the phylogenetic diversity (PD) of predicted assemblages at the genera level using the Chao's PD index with an order  $q$  of 1 (equivalent to the Shannon index)<sup>93</sup> that we used as a proxy of adaptive capacity.

Vulnerability to climate change,  $Vulnerability_{clim}$ , was finally estimated as the sum of the three standardized (st) (0 to 1) components:

$$Vulnerability_{clim} = \left( Sensitivity_{clim}^{st} + Exposure_{clim}^{st} - Adaptive_{clim}^{st} \right).$$

$Vulnerability_{clim}$  theoretically ranges from -1 (low vulnerability) to 2 (high vulnerability) and, owing to the standardization of its three components, it expresses a relative vulnerability over the study area and is thus little affected by the IPCC scenario chosen (RCP 4.5 or 8.5) because different scenarios predict different amplitudes of changes but similar spatial patterns (Extended Data Fig. 7).

Forecasted human impact on forests in 2085 was assessed using our human-induced forest-disturbance intensity index combined with country-specific projections of human populations in 2085. We assigned to each current town a country-specific relative population increase drawn from the United Nations World Population Prospects<sup>2</sup> and rebuilt our index based on this modified dataset. This approach did not account for new roads that might be established by 2085, and thus tended to underestimate the increase in anthropogenic pressure.

### Software and packages

All analyses were performed and figures were created with R<sup>94</sup>, mostly using the `ade4`<sup>95</sup>, `alphashape3d`<sup>96</sup>, `ggplot2`<sup>97</sup>, `raster`<sup>98</sup>, `RGeostats`<sup>99</sup>, `entropart`<sup>99</sup> and `SCGLR`<sup>84</sup> (<https://github.com/SCnext/SCGLR/>) packages. Data are archived in a public repository (<https://doi.org/10.18167/DVN1/UCNCA7>).

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

### Data availability

All maps and data used for this article are accessible online in a public repository at <https://doi.org/10.18167/DVN1/UCNCA7>. Raw floristic data are, however, archived in a private data repository, owing to the highly sensitive nature of commercial inventory data, and access may be granted for research purposes using the form provided in the public repository.

### Code availability

R scripts are available at <https://github.com/MaximeRM/ScriptNature>.

53. Réjou-Méchain, M. et al. Detecting large-scale diversity patterns in tropical trees: can we trust commercial forest inventories? *For. Ecol. Manage.* **261**, 187–194 (2011).
54. African Plant Database v.3.4.0 (Conservatoire et Jardin Botaniques de la Ville de Genève and South African National Biodiversity Institute, Pretoria, accessed 10 February 2017).
55. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**, 105–121 (2009).
56. Dauby, G. et al. RAINBIO: a mega-database of tropical African vascular plants distributions. *PhytoKeys* **74**, 1–18 (2016).
57. Chave, J. et al. Towards a worldwide wood economics spectrum. *Ecol. Lett.* **12**, 351–366 (2009).
58. Zanne, A. E. et al. Data from: towards a worldwide wood economic spectrum. *Dryad* <https://doi.org/10.5061/dryad.234> (2009).
59. Gourlet-Fleury, S. et al. Environmental filtering of dense-wooded species controls above-ground biomass stored in African moist forests. *J. Ecol.* **99**, 981–990 (2011).
60. Westoby, M. & Wright, I. J. Land-plant ecology on the basis of functional traits. *Trends Ecol. Evol.* **21**, 261–268 (2006).
61. Chave, J. et al. Improved allometric models to estimate the aboveground biomass of tropical trees. *Glob. Change Biol.* **20**, 3177–3190 (2014).
62. Bénédict, F. et al. CoForTraits, African plant traits information database v.1.0, <https://doi.org/10.18167/DVN1/Y2BIZK> (2013).
63. Davies, T. J. et al. Phylogenetic conservatism in plant phenology. *J. Ecol.* **101**, 1520–1530 (2013).
64. Cramer, W. et al. Global response of terrestrial ecosystem structure and function to CO<sub>2</sub> and climate change: Results from six dynamic global vegetation models. *Glob. Change Biol.* **7**, 357–373 (2001).
65. Menzel, A. Phenology: its importance to the global change community. *Clim. Change* **54**, 379 (2002).
66. Borchert, R., Rivera, G. & Hagnauer, W. Modification of vegetative phenology in a tropical semi-deciduous forest by abnormal drought and rain. *Biotropica* **34**, 27–39 (2002).
67. Kraft, N. J. B., Valencia, R. & Ackerly, D. D. Functional traits and niche-based tree community assembly in an Amazonian forest. *Science* **322**, 580–582 (2008).
68. Schamp, B. S. & Arssen, L. W. The assembly of forest communities according to maximum species height along resource and disturbance gradients. *Oikos* **118**, 564–572 (2009).
69. New, M., Lister, D., Hulme, M. & Makin, I. A high-resolution data set of surface climate over global land areas. *Clim. Res.* **21**, 1–25 (2002).
70. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surface for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
71. Funk, C. et al. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* **2**, 150066 (2015).
72. Nachtergaele, F. et al. The harmonized world soil database. In *Proc. 19th World Congress of Soil Science, Soil Solutions for a Changing World* (eds Gilkes, R. & Prakongkep, N.) 34–37 (International Union of Soil Sciences, 2010).
73. Woolmer, G. et al. Rescaling the human footprint: a tool for conservation planning at an ecoregional scale. *Landsc. Urban Plan.* **87**, 42–53 (2008).
74. Venter, O. et al. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nat. Commun.* **7**, 12558 (2016).
75. Geldmann, J., Joppa, L. N. & Burgess, N. D. Mapping change in human pressure globally on land and within protected areas. *Conserv. Biol.* **28**, 1604–1616 (2014).
76. Linard, C., Gilbert, M., Snow, R. W., Noor, A. M. & Tatem, A. J. Population distribution, settlement patterns and accessibility across Africa in 2010. *PLoS One* **7**, e31743 (2012).
77. Lloyd, C. T. et al. Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data* **3**, 108–139 (2019).
78. Boulesteix, A.-L. & Strimmer, K. Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Brief. Bioinform.* **8**, 32–44 (2007).
79. Carrascal, L. M., Galván, I. & Gordo, O. Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos* **118**, 681–690 (2009).
80. Tenenhaus, M. *La Régression PLS: Théorie et Pratique* (Editions Technip, 1998).
81. Sabatier, R., Lebreton, J. D. & Chessel, D. in *Multivariate Data Analysis* (eds Coppi, R. & Bolasco, S.) 341–352 (1989).
82. Ter Braak, C. J. F. in *Theory and Models In Vegetation Science* (eds Prentice, I. C. & van der Maarel, E.) 69–77 (Springer, 1987).
83. Bry, X. & Verron, T. THEME: THEmatic model exploration through multiple co-structure maximization. *J. Chemometr.* **29**, 637–647 (2015).
84. Cornu, G., Mortier, F., Trottier, C. & Bry, X. SCGLR: supervised component generalized linear regression. R version 3.0 <https://cran.r-project.org/web/packages/SCGLR/index.html> (2016).
85. Ward, J. H. Jr. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244 (1963).
86. Ploton, P. et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 4540 (2020).
87. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: clustering, classification and density estimation using gGaussian finite mixture models. *R. J.* **8**, 289–317 (2016).
88. Dormann, C. F. et al. Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography* **30**, 609–628 (2007).
89. Renard, D. et al. RGeostats: the geostatistical package v.11.0. 1 <http://rgeostats.free.fr/> (MINES ParisTech, 2014).
90. Platts, P. J., Omeny, P. A. & Marchant, R. AFRICLIM: high-resolution climate projections for ecological applications in Africa. *Afr. J. Ecol.* **53**, 103–108 (2015).
91. Janssens, S. B. et al. A large-scale species level dated angiosperm phylogeny for evolutionary and ecological analyses. *Biodivers. Data J.* **8**, e39677 (2020).
92. Abouheif, E. A method for testing the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* **1**, 895–909 (1999).
93. Chao, A., Chiu, C.-H. & Jost, L. Phylogenetic diversity measures based on Hill numbers. *Phil. Trans. R. Soc. B* **365**, 3599–3609 (2010).
94. R Core Team. *R: a language and environment for statistical computing*. (R Foundation for Statistical Computing, 2017).
95. Chessel, D., Dufour, A. B. & Thioulouse, J. The `ade4` package – I: one-table methods. *R News* **4**, 5–10 (2004).
96. Lafarge, T. & Pateiro-Lopez, B. `alphashape3d`: implementation of the 3D alpha-shape for the reconstruction of 3D sets from a point cloud. R version 1.3.1 <https://cran.r-project.org/web/packages/alphashape3d/index.html> (2017).
97. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
98. Hijmans, R. J. `raster`: geographic data analysis and modelling. R version 3.4-5 <https://cran.r-project.org/web/packages/raster/index.html> (2017).
99. Marcon, E. & Hérault, B. `entropart`: An R package to measure and partition diversity. *J. Stat. Softw.* **67**, 1–26 (2015).
100. Dufrière, M. & Legendre, P. Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol. Monogr.* **67**, 345–366 (1997).

**Acknowledgements** We thank the 105 forest companies that provided access, albeit restricted, to their inventory data for research purposes and members of the central African plot network (<https://central-african-plot-network.netlify.app/>), Y. Yalibanda, F. Allah-Barem, F. Baya, F. Boyemba, M. Mbasi Mbula, P. Berenger, M. Mazengue, V. Istace, I. Zombo, E. Forni, Nature+ and the CEB-Precious Woods company for giving access to the scientific inventories described in Supplementary Fig. 4, some of which were funded by the AFD and the FFEM (for example, DynAFFor and P3FAC projects). We thank J. Chave, P. Couteron, S. Lewis and M. Tadesse for their comments and discussions on previous versions, B. Sultan for useful discussions on climate projections, O. J. Hardy for advice on phylogenetical analyses, B. Locatelli for advice on vulnerability analyses, G. Vieilledet for discussions on the human-induced forest-disturbance intensity index and A. Stokes for English editing. This work was supported by the CoForTips project (ANR-12-EBID-0002) funded by the ERA-NET BiodivERsA, with the national funders ANR, BELSPO and WWF, as part of the 2012

# Article

BiodivERsA call for research proposals, the GAMBAS project funded by the French National Research Agency (ANR-18-CE02-0025) and the project 3DForMod funded by the UE FACCE ERA-GAS consortium (ANR-17-EGAS-0002-01). This study is a contribution to the research program of LMI DYCOFAC (Dynamique des écosystèmes continentaux d'Afrique Centrale en contexte de changements globaux).

**Author contributions** Conceptualization: M.R.-M., F.M., R.P. and S.G.-F.; data curation: G.C. and F.B.; formal analysis: M.R.-M. and F.M.; project administration: C.G.; writing (original draft): M.R.-M., F.M., R.P. and S.G.-F.; writing (review and editing): all authors.

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

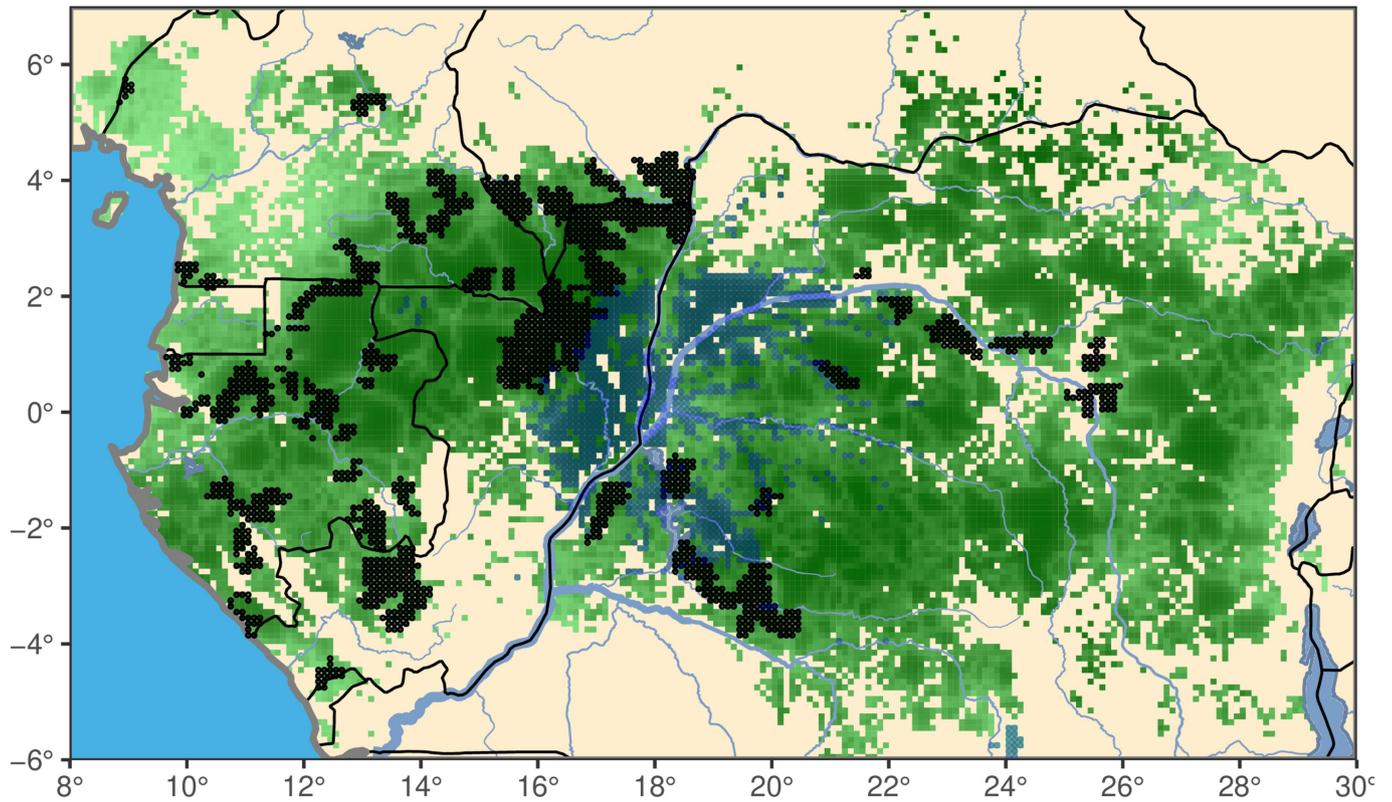
**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03483-6>.

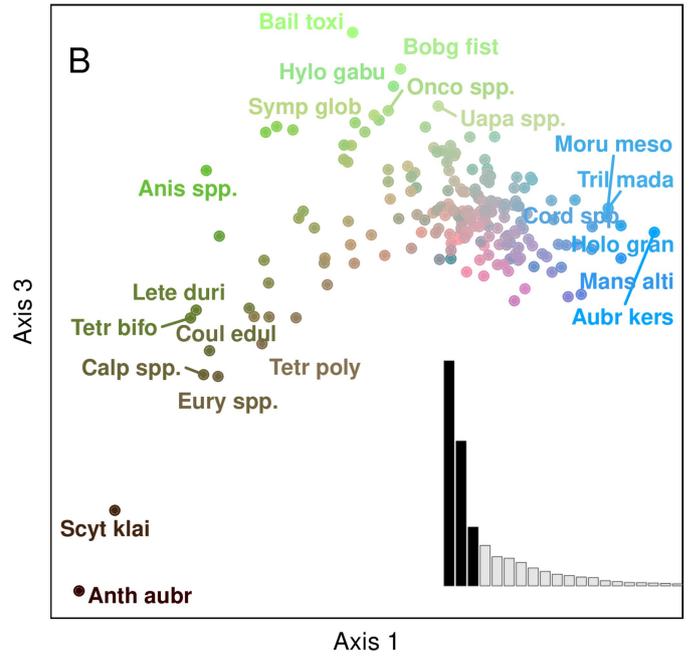
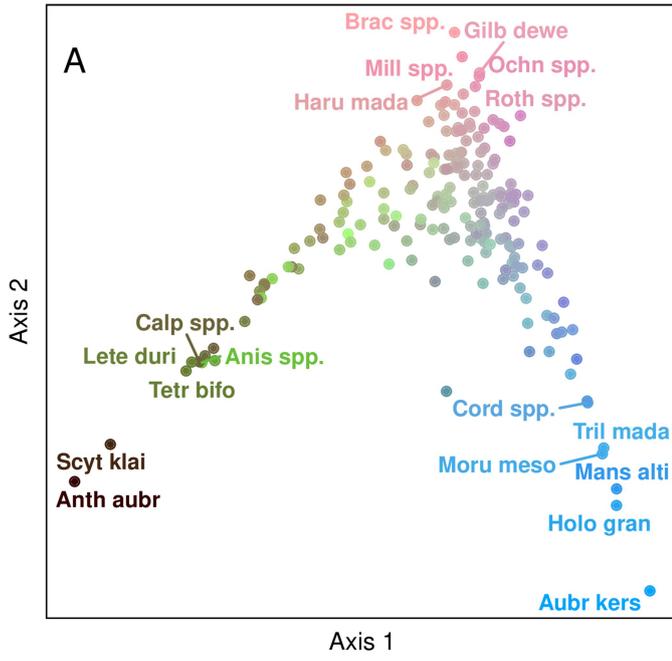
**Correspondence and requests for materials** should be addressed to M.R.-M.

**Peer review information** *Nature* thanks Jonas Geldmann, Marion Pfeifer, Philip Platts and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

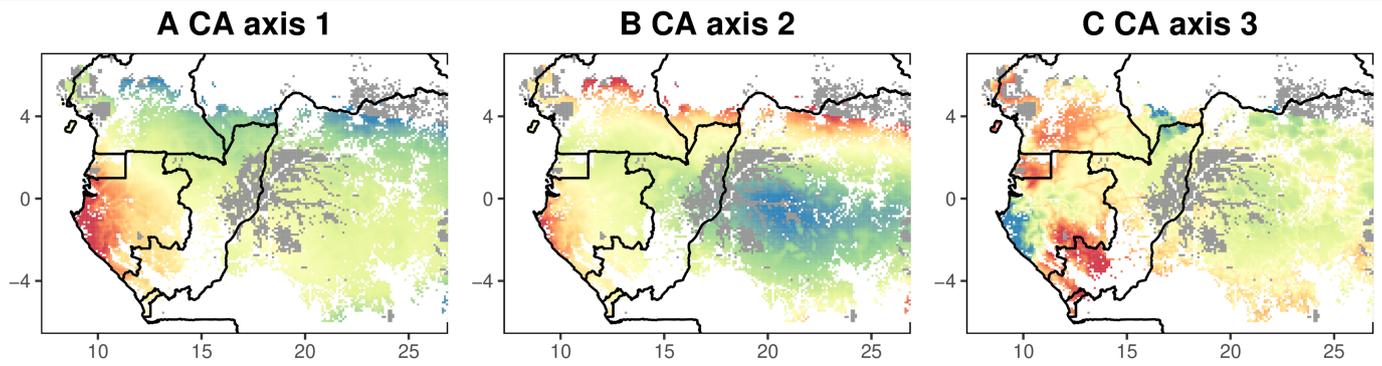


**Extended Data Fig. 1 | Study area and sampling plots.** In green, the current distribution of tropical forests following the ESA-CCI landcover product (v.1.6), with a dark-green-to-white gradient representing anthropogenic pressure (see Methods) and non-forested areas represented in beige. The sampling grid cells ( $n = 1,57110 \times 10\text{-km}^2$  grid cells) are in black and the flooding forests, as proposed by the ESA-CCI landcover, are in blue.



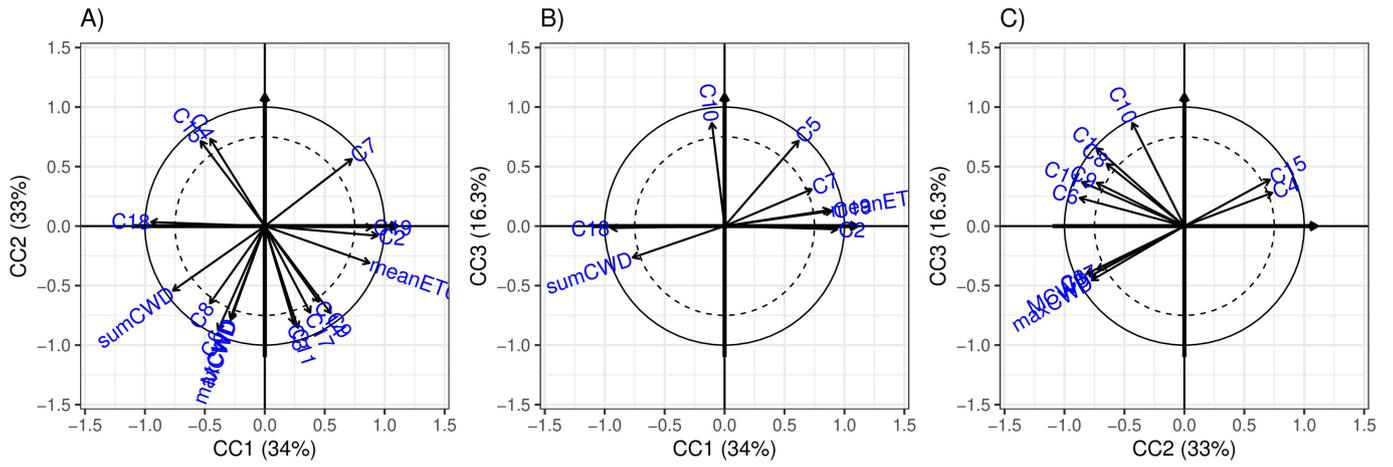
**Extended Data Fig. 2 | Taxon CA planes 1–2 and 1–3 with labels for the 12 most representative taxa on each axis. a**, Planes 1–2. **b**, Planes 1–3. Colour code corresponds to that reported in Fig. 1. The first eigenvalues are reported

in **b**, highlighting in black the first three axes. Taxon codes and scores of the 193 taxa are provided in Supplementary Table 2.

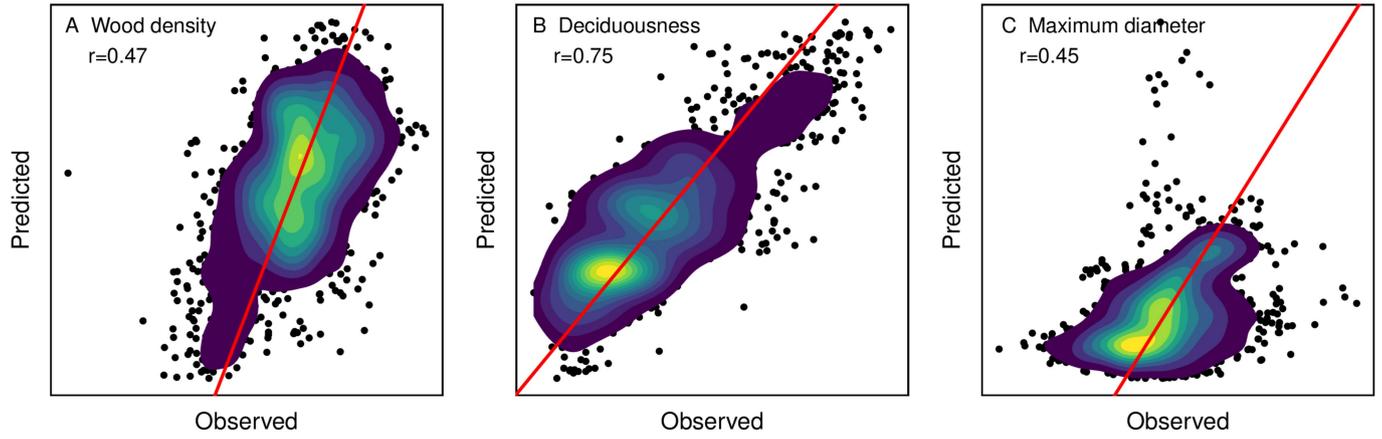


**Extended Data Fig. 3 | Individual predicted floristic gradients illustrated by the three first axes of the correspondence analysis performed on predicted taxon abundances. a–c, CA axis 1 (a), CA axis 2 (b) and CA axis 3 (c).**

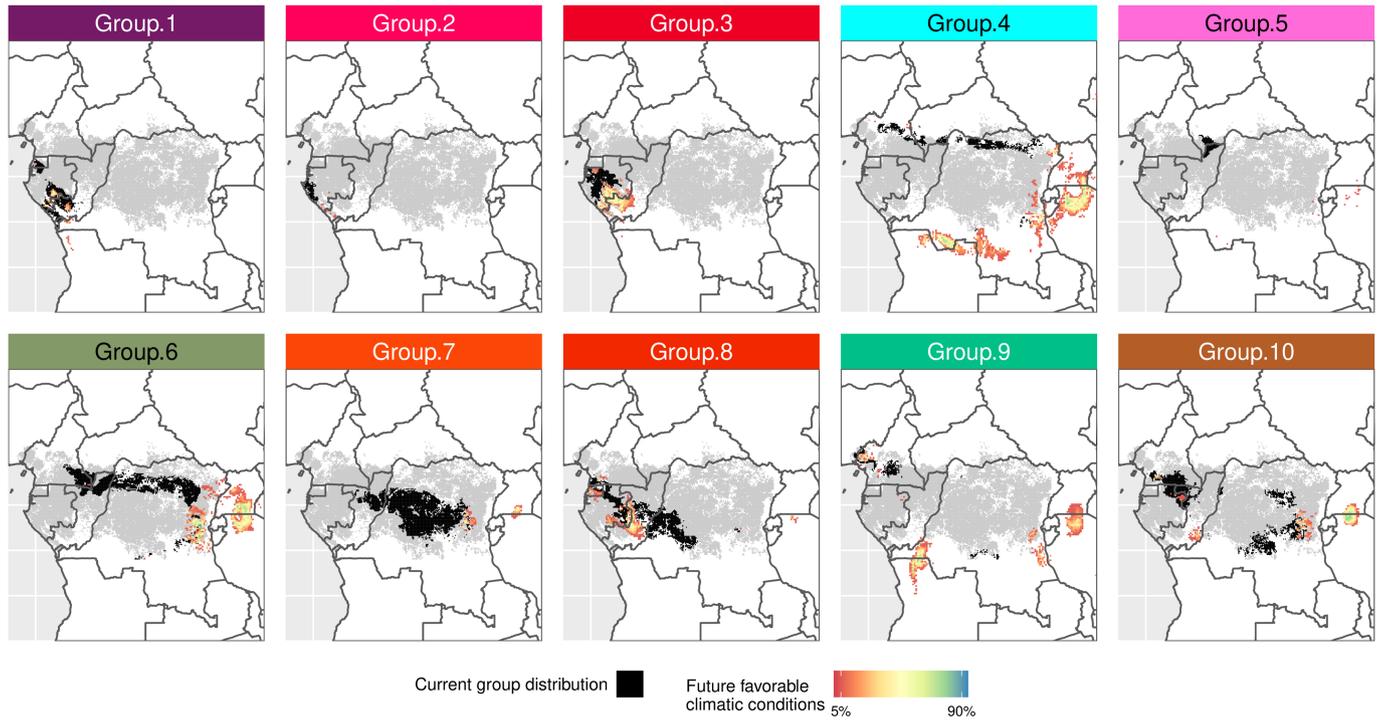
A composite map of these three axes is given in Fig. 1 and the corresponding taxon CA planes are provided in Extended Data Fig. 2.



**Extended Data Fig. 4 | Plans 1–2, 1–3 and 2–3 of the SCGLRCCs. a, CC1 and CC2. b, CC1 and CC3. c, CC2 and CC3.** All climatic variables with a correlation of less than 0.75 with the two components (dashed circle) were excluded for the sake of clarity. For abbreviations, see Extended Data Table 2.



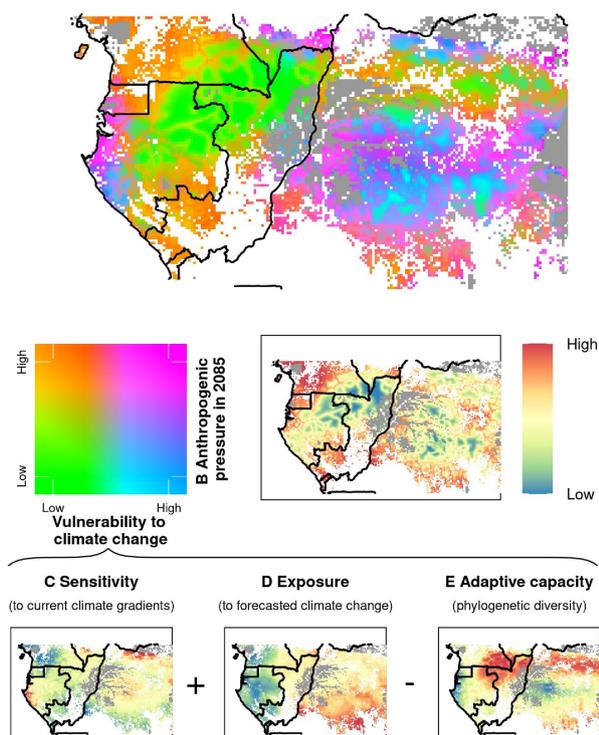
**Extended Data Fig. 5 | Spatial cross-validation results of the predictions of functional assemblages. a–c,** The observed and predicted community weighted mean trait values within the  $1,57110 \times 10\text{-km}^2$  grid cells are given for wood density (a), deciduousness (b) and maximum diameter (c). The 1:1 line is displayed in red.



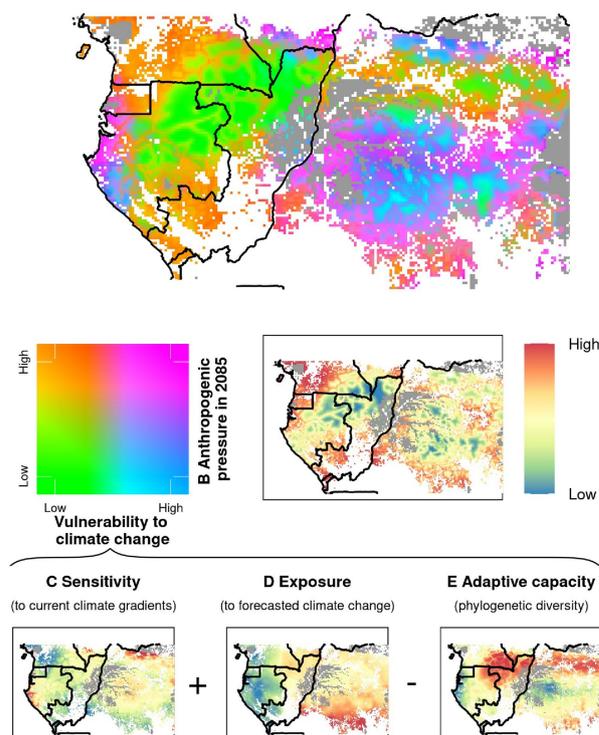
**Extended Data Fig. 6 | Projected changes under the RCP 4.5 scenario in 2085 of the climatic conditions of the 10 forest types.** Areas for which climate models predict similar climatic components (CCs) values to those currently found within forest types (in black) are illustrated with a colour gradient indicating the level of agreement amongst the 18 climate models

(as a percentage; no colour indicates that none of the original 18 climate models predicted similar conditions). More specifically, we used 3D concave hull (alpha shape) models to assess where the combinations of current CCs corresponding to each forest type are predicted to be represented in 2085.

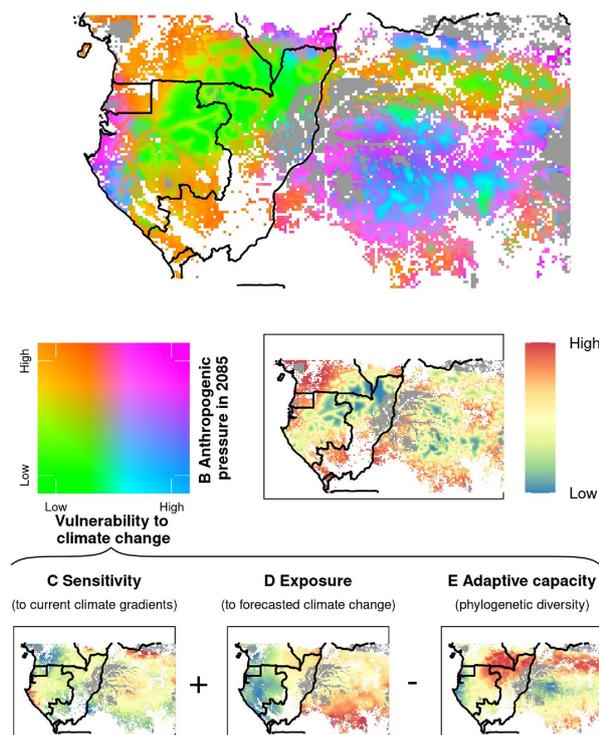
**A Vulnerability to global change (RCP 4.5 2055)**



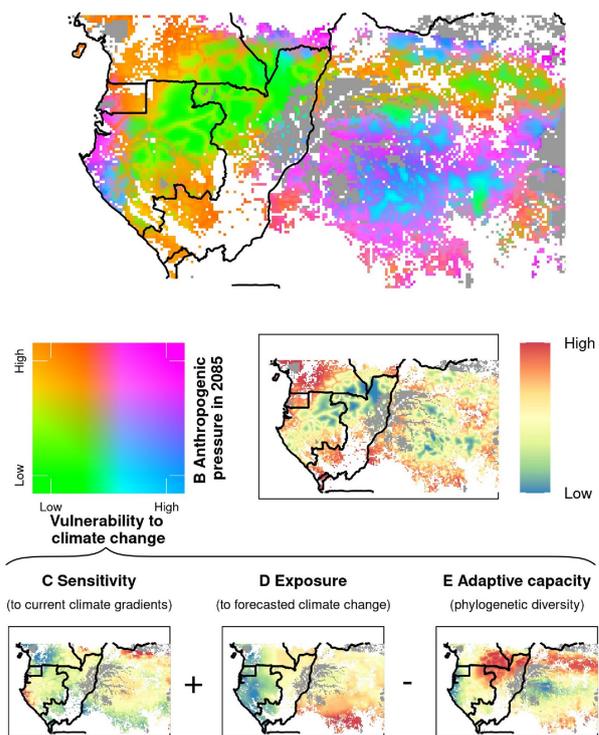
**B Vulnerability to global change (RCP 8.5 2055)**



**C Vulnerability to global change (RCP 4.5 2085)**

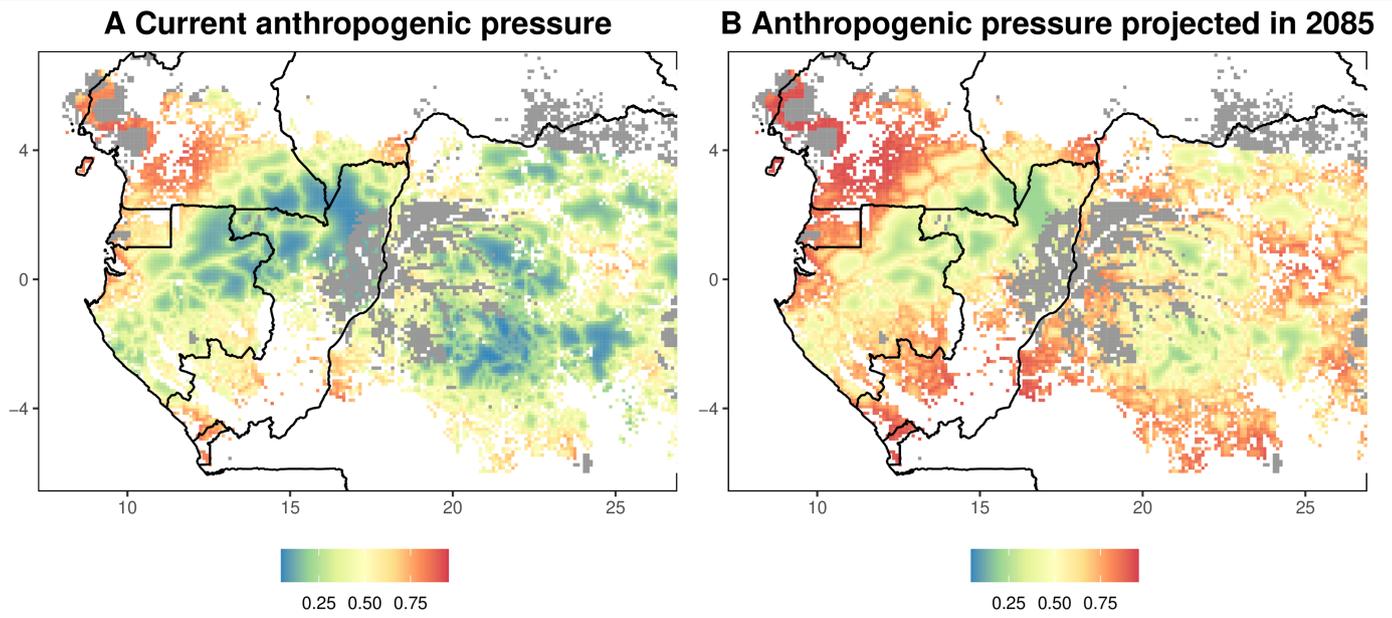


**D Vulnerability to global change (RCP 8.5 2085)**

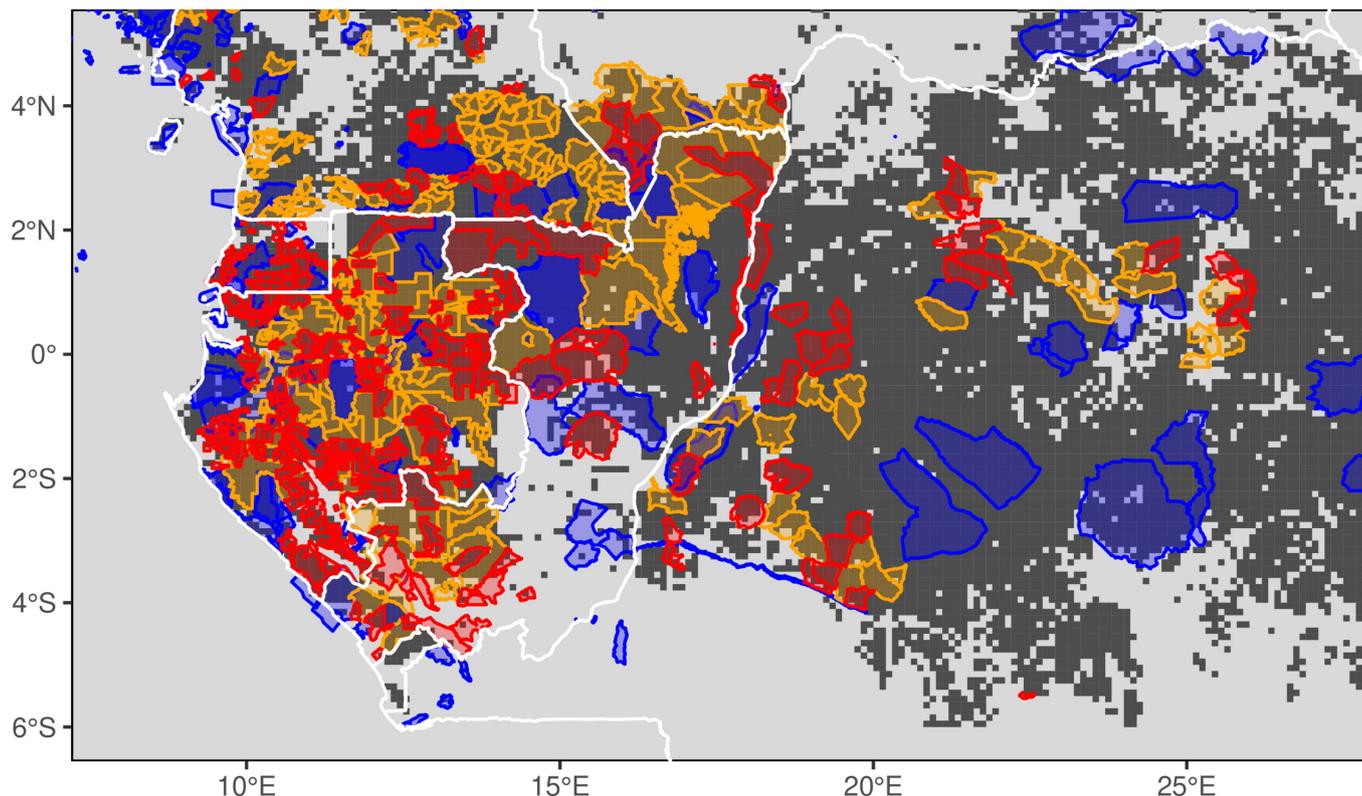


**Extended Data Fig. 7 | The vulnerability map under two different RCP scenarios and for two different years. a–d, Vulnerability maps under RCP 4.5 in 2055 (a), RCP 8.5 in 2055 (b), RCP 4.5 in 2085 (c) and RCP 8.5 in 2085 (d). As can be seen, the predicted vulnerability is little affected by the IPCC scenario**

chosen because it expresses a relative vulnerability over the study area and, if different scenarios predict different amplitudes of climate change, spatial patterns of climate exposure remains similar (see Methods).



**Extended Data Fig. 8 | Current and projected anthropogenic pressure over central Africa. a, b,** Current (a) and projected (b) anthropogenic pressure predicted from our index of human-induced forest-disturbance intensity.



**Extended Data Fig. 9 | Protected area network and areas dedicated to logging activities in central Africa.** The protected area network is shown in blue; areas dedicated to logging are shown in orange and red. Data on protected areas were obtained from the World Database on Protected Areas (<https://www.iucn.org/theme/protected-areas/our-work/world-database-protected-areas>, last accessed 14 August 2018), excluding marine, hunting and game-oriented areas, except for the Democratic Republic of the Congo, for which data from the World Resource Institute were used and downloaded from ArcGIS hub (<https://hub.arcgis.com/datasets/1bcd463cbb6549c9a0676edb9>

f751f9b, last accessed 1 June 2019). Logging activity data were provided by the Observatoire des Forêts d'Afrique Centrale based on an unpublished work completed in June 2018, except for the Democratic Republic of the Congo, for which more updated data (June 2019) were provided by the AGEDUFOR national project. Areas in orange illustrate forest concessions that are known to have, or to be in the process of having, an officially validated sustainable forest management plan. Red areas illustrate forest areas that are currently dedicated to logging but that either do not have an official management plan or have an uncertain status.

# Article

**Extended Data Table 1 | Characteristics of the floristic groups**

Group	Name	Main families	Representative taxa	Area	PA	Logging	Phum	CC1	CC2	CC3
1	Atlantic highland evergreen	Fabaceae (19%), Burseraceae (17%), Myristicaceae (13%)	Anisophyllea spp., Baillonella toxisperma, Aucoumea klaineana, Bobgunnia fistuloides, Testulea gabonensis	79,400	9	70	0.48	-3.4	3.4	-1.1
2	Atlantic coastal evergreen	Fabaceae (27%), Burseraceae (17%), Myristicaceae (15%)	Anthostema aubryanum, Scytopetalum klaianum, Calpocalyx spp., Coula edulis, Tetraberlinia bifoliolata	17,700	54	36	0.35	-5.2	0.1	4.6
3	Atlantic inland evergreen	Fabaceae (27%), Burseraceae (15%), Myristicaceae (14%)	Calpocalyx spp., Letestua durissima, Eurypetalum spp., Coula edulis, Tetraberlinia bifoliolata	60,800	22	69	0.35	-3.8	-0.1	3.0
4	Margin semideciduous	Malvaceae (16%), Fabaceae (13%), Cannabaceae (11%)	Aubrevillea kerstingii, Holoptelea grandis, Mansonia altissima, Trilepisium madagascariense, Morus mesozygia	87,600	4	20	0.42	4.1	2.4	2.0
5	Evergreen-semideciduous on sandstone	Fabaceae (15%), Sapotaceae (15%), Annonaceae (9%)	Manilkara spp., Oldfieldia africana, Balanites wilsoniana, Autranella congolensis, Synsepalum spp.	22,200	23	80	0.27	3.2	2.0	1.0
6	Semideciduous	Fabaceae (20%), Annonaceae (10%), Malvaceae (8%)	Pericopsis elata, Fernandoa adolfi friderici, Dasylepis seretii, Desplatsia spp., Entandrophragma cylindricum	206,400	10	34	0.26	2.6	-0.3	-0.2
7	Central evergreen	Fabaceae (33%), Annonaceae (9%), Olacaceae (9%)	Millettia spp., Brachystegia spp., Ochna spp., Gilbertiodendron dewevrei, Rothmannia spp.	265,900	23	9	0.22	0.6	-3.8	-2.3
8	Mixed evergreen	Fabaceae (30%), Olacaceae (10%), Myristicaceae (8%)	Diogoa zenkeri, Elaeis guineensis, Cryptosepalum spp., Bikinia spp., Ochthocosmus spp.	158,200	10	45	0.40	-1.4	-2.2	-0.2
9	Degraded semideciduous	Fabaceae (14%), Cannabaceae (13%), Urticaceae (8%)	Pseudospondias spp., Musanga cecropioides, Pterygota spp., Ricinodendron heudelotii, Afzelia spp.	40,000	10	6	0.73	1.1	2.0	1.6
10	Semideciduous-evergreen transition	Fabaceae (22%), Annonaceae (10%), Olacaceae (8%)	Uapaca spp., Musanga cecropioides, Annickia spp., Croton spp., Pseudospondias spp.	180,000	15	29	0.37	0.4	-0.3	-1.5

For each floristic group, information is provided on the three most abundant families (Angiosperm Phylogeny Group (APG) III classification, except for the subfamilies Caesalpiniaceae and Mimosaceae, which were considered here independently owing to their different ecological strategies), the five most representative taxa (that is, taxa having the highest A score of the Dufréne and Legendre index<sup>100</sup>), the total area (km<sup>2</sup>) covered by each group, the percentage of the area covered by protected areas (PA) and dedicated to logging activities (Logging), the mean probability of being affected by human activities (Phum, this study) and the mean value of the three climatic components (CCs) that best explain the current distribution of central African trees (this study).

**Extended Data Table 2 | Climatic predictors**

CODE	Description	Mean (range)	CC1	CC2	CC3
C1	Annual Mean Temperature (°C)	24.7 (22.7 - 26.4)	-0.01	-0.54	0.43
C2	Mean Diurnal Range (°C)	9.4 (6.2 - 11)	0.88	0.01	0
C3	Isothermality (C2/C7) (* 100) (unitless)	76.6 (53.1 - 89.8)	-0.06	-0.67	-0.16
C4	Temperature Seasonality (Coefficient of Variation of kelvin values) (%)	1.5 (0.7 - 3.6)	-0.21	0.54	0.08
C5	Max Temperature of Warmest Month (°C)	31.1 (28.6 - 33.9)	0.38	-0.03	0.51
C6	Min Temperature of Coldest Month (°C)	18.8 (16.5 - 22)	-0.15	-0.76	0.06
C7	Temperature Annual Range (C5-C6) (°C)	12.3 (9.4 - 16)	0.53	0.32	0.1
C8	Mean Temperature of Wettest Quarter (°C)	24.5 (22 - 26.6)	-0.21	-0.42	0.27
C9	Mean Temperature of Driest Quarter (°C)	24.2 (20.9 - 27.2)	0.3	-0.53	0.13
C10	Mean Temperature of Warmest Quarter (°C)	25.6 (23.9 - 27.3)	-0.01	-0.19	0.75
C11	Mean Temperature of Coldest Quarter (°C)	23.8 (20.5 - 25.5)	0.08	-0.73	0.14
C12	Annual Precipitation (mm)	1733.5 (1219.7 - 2983)	-0.26	-0.06	0.06
C13	Precipitation of Wettest Month (mm)	263.2 (195.7 - 608.7)	-0.43	0.05	0.11
C14	Precipitation of Driest Month (mm)	31.9 (0 - 112.4)	0.2	-0.4	-0.16
C15	Precipitation Seasonality (Coefficient of Variation) (%)	52.1 (21.7 - 84.5)	-0.29	0.51	0.15
C16	Precipitation of Wettest Quarter (mm)	665 (435.8 - 1273.7)	-0.28	0.01	0.15
C17	Precipitation of Driest Quarter (mm)	137.7 (2 - 405.2)	0.14	-0.53	-0.13
C18	Precipitation of Warmest Quarter (mm)	434.3 (220.7 - 816.1)	-0.9	0	0
C19	Precipitation of Coldest Quarter (mm)	302.8 (0.7 - 1332.9)	0.8	0	0.02
meanETO	mean monthly evapotranspiration <sup>a</sup> (mm)	133.8 (109.9 - 146.6)	0.76	-0.1	0.02
meanCWB	mean climatic water balance <sup>b</sup> (mm)	10.6 (-19.2 - 134.3)	-0.52	0.01	0.02
sumCWD	total climatic water deficit <sup>c</sup> (mm)	-1617.8 (-4466.3 - -55.8)	-0.59	-0.29	-0.07
maxCWD	maximum cumulative water deficit <sup>d</sup> (mm)	-299.3 (-596 - -21.9)	-0.08	-0.6	-0.21
MCWD	maximum climatic water deficit <sup>e</sup> (mm)	-312.3 (-596 - -37.7)	-0.08	-0.62	-0.18

Correlations with the three climatic components (CCs) are given in the last three columns (see also Extended Data Fig. 4).

<sup>a</sup>meanETO was calculated using the Hargreaves formula with  $\text{meanETO} = 1/n(\sum_{i=1}^n \text{ETO}_i)$ , where  $\text{ETO}_i$  is the evapotranspiration of month  $i$  calculated as  $\text{ETO}_i = 0.0023 \times 0.408 \text{RA}_i \times (\text{Tavg}_i + 17.8) \times \text{TD}_i^{0.5}$ , in which  $\text{RA}_i$  is the mean extrasolar radiation of month  $i$  in  $\text{MJ m}^{-2} \text{d}^{-1}$ ,  $\text{Tavg}_i$  is the average daily temperature of month  $i$  in °C, computed as the average of the mean maximum and minimum temperature of month  $i$ , and  $\text{TD}_i$  is the mean temperature range of month  $i$  in °C, computed as the difference between the mean maximum and minimum temperature of month  $i$ .

<sup>b</sup>meanCWB =  $1/n(\sum_{i=1}^n P_i - \text{ETO}_i)$ , where  $P_i$  is the precipitation of month  $i$ .

<sup>c</sup>sumCWD =  $\sum_{i=1}^n \text{CWD}_i$  and  $\text{maxCWD} = \max(\text{CWD})$ , where  $\text{CWD}_i = \sum_{j=1}^i \text{WD}_j$ , with  $\text{WD}_i = \text{WD}_{i-1} + P_i - \text{ETO}_i$  if  $(\text{WD}_{i-1} + P_i - \text{ETO}_i) < 0$  or  $\text{WD}_i = 0$  if  $(\text{WD}_{i-1} + P_i - \text{ETO}_i) \geq 0$ . To compute  $\text{CWD}_i$ , the wettest month was set as  $i = 1$  at the grid-cell level.

<sup>e</sup>MCWD =  $\sum_{i=1}^n \min(0, P_i - \text{ETO}_i)$ .

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Taxonomy was revised and homogenized using the African Flowering Plants Database (<http://www.ville-ge.ch/musinfo/bd/cjb/africa/index.php?langue=an>, last access on 01/09/2019).

For the human-induced forest disturbance index we used data from the "Global Rural Urban Mapping Project" (<https://sedac.ciesin.columbia.edu/data/set/grump-v1-settlement-points>; last access the 01/10/2018), the Natural Earth Populated Places product (version 3.0.0; <http://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-populated-places/>; last access the 07/10/2018) derived from the LandScan (<https://earthworks.stanford.edu/catalog/stanford-yj715rc4110#iso-metadata-reference-info>; last access the 05/10/2018) dataset, from OpenStreetMap (<https://data.maptiler.com/downloads/planet/#1.59/-17.3/19.7>; last access 02/10/2018) and from the Global Roads Open Access Data Set, version 1 (<https://data.maptiler.com/downloads/planet/>

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<i>Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.</i>
Data exclusions	<i>Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Replication	<i>Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.</i>
Randomization	<i>Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.</i>
Blinding	<i>Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.</i>

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	<i>Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).</i>
Research sample	<i>State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.</i>
Sampling strategy	<i>Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.</i>
Data collection	<i>Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.</i>
Timing	<i>Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.</i>
Data exclusions	<i>If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.</i>
Non-participation	<i>State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.</i>
Randomization	<i>If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.</i>

# Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	In this study, we use a massive compilation of forest inventory data providing information on the abundance distribution of 193 dominant tree taxa in 185,665 plots (6 million trees). We jointly model their spatial distribution and provide the first transnational benchmark maps of the floristic and functional composition of central African forests. Based on these predictions and on global change scenarios, we then predicted the expected vulnerability of central African forests to global change by 2085.
Research sample	We studied tropical trees with a diameter at breast height (DBH) $\geq$ 30 cm. The sample choice was based on the availability of commercial inventories and the data is meant to represent tree communities in central Africa.
Sampling strategy	The sampling strategy consisted of continuous and parallel transects 20 m or 25 m wide, often 2-3 km apart, and subdivided into rectangular 0.4 or 0.5-ha plots. The total number of plots was of 18,665 are hence considered to represent most central African forests
Data collection	Data were collected by forest companies in 105 logging concessions in order to build management plans. Most forest companies were assisted and trained by European consultant firms for these inventories (e.g. CIRAD or Foret Resource and Management).
Timing and spatial scale	Data were collected over an area of ca. 160,000 square km from 1996 to 2014.
Data exclusions	We discarded i) species and genera deemed to be not reliably identified over the whole study area; ii) taxa occurring in less than 5% of the grid cells because they cannot be studied at the regional scale; iii) grid cells having a field plot sampling area $<$ 10 ha and where the selected taxa represented less than 75% of the total number of individuals originally inventoried to ensure that our dataset was representative of the within-grid cell tree community composition.  All this selection procedure (detailed in the methods) was done independently and before the statistical analyses in order to perform analyses on a high-quality dataset.
Reproducibility	Data are available at <a href="http://dx.doi.org/10.18167/DVN1/UCNCA7">http://dx.doi.org/10.18167/DVN1/UCNCA7</a> and codes are available at <a href="https://github.com/MaximeRM/ScriptNature">https://github.com/MaximeRM/ScriptNature</a> . Analyses are thus fully reproducible. Due to its huge size, the raw dataset is, however, hardly reproducible at short term.
Randomization	Data were acquired through systematic inventories designed to best represent the different forest types within the forest concession.
Blinding	Because data were acquired through a systematic design without any prior stratification or information, data acquisition can be considered here as a blinding process.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

## Field work, collection and transport

Field conditions	<i>Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).</i>
Location	<i>State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).</i>
Access & import/export	<i>Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).</i>
Disturbance	<i>Describe any disturbance caused by the study and how it was minimized.</i>

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

- n/a Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Human research participants
- Clinical data
- Dual use research of concern

## Methods

- n/a Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

## Antibodies

Antibodies used *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s) *State the source of each cell line used.*

Authentication *Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*

Commonly misidentified lines (See [ICLAC](#) register) *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

## Palaeontology and Archaeology

Specimen provenance *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*

Specimen deposition *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals *For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*

Wild animals *Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Field-collected samples *For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*

Ethics oversight *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*

Recruitment

*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*

Ethics oversight

*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration

*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*

Study protocol

*Note where the full trial protocol can be accessed OR if not available, explain why.*

Data collection

*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*

Outcomes

*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

## Dual use research of concern

Policy information about [dual use research of concern](#)

### Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

- | No                       | Yes                      |                            |
|--------------------------|--------------------------|----------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | Public health              |
| <input type="checkbox"/> | <input type="checkbox"/> | National security          |
| <input type="checkbox"/> | <input type="checkbox"/> | Crops and/or livestock     |
| <input type="checkbox"/> | <input type="checkbox"/> | Ecosystems                 |
| <input type="checkbox"/> | <input type="checkbox"/> | Any other significant area |

### Experiments of concern

Does the work involve any of these experiments of concern:

- | No                       | Yes                      |   |
|--------------------------|--------------------------|---|
| <input type="checkbox"/> | <input type="checkbox"/> | Demonstrate how to render a vaccine ineffective                             |
| <input type="checkbox"/> | <input type="checkbox"/> | Confer resistance to therapeutically useful antibiotics or antiviral agents |
| <input type="checkbox"/> | <input type="checkbox"/> | Enhance the virulence of a pathogen or render a nonpathogen virulent        |
| <input type="checkbox"/> | <input type="checkbox"/> | Increase transmissibility of a pathogen                                     |
| <input type="checkbox"/> | <input type="checkbox"/> | Alter the host range of a pathogen  |
| <input type="checkbox"/> | <input type="checkbox"/> | Enable evasion of diagnostic/detection modalities                           |
| <input type="checkbox"/> | <input type="checkbox"/> | Enable the weaponization of a biological agent or toxin                     |
| <input type="checkbox"/> | <input type="checkbox"/> | Any other potentially harmful combination of experiments and agents         |

## ChIP-seq

### Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links  
May remain private before publication.

For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.

Files in database submission

Provide a list of all files available in the database submission.

Genome browser session  
(e.g. [UCSC](#))

Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.

## Methodology

Replicates

Describe the experimental replicates, specifying number, type and replicate agreement.

Sequencing depth

Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.

Antibodies

Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.

Peak calling parameters

Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.

Data quality

Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.

Software

Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.

## Flow Cytometry

### Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

### Methodology

Sample preparation

Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.

Instrument

Identify the instrument used for data collection, specifying make and model number.

Software

Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.

Cell population abundance

Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.

Gating strategy

Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

## Magnetic resonance imaging

### Experimental design

Design type

Indicate task or resting state; event-related or block design.

Design specifications

Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.

Behavioral performance measures

State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).

## Acquisition

Imaging type(s)

Field strength

Sequence & imaging parameters

Area of acquisition

Diffusion MRI  Used  Not used

## Preprocessing

Preprocessing software

Normalization

Normalization template

Noise and artifact removal

Volume censoring

## Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis:  Whole brain  ROI-based  Both

Statistic type for inference (See [Eklund et al. 2016](#))

Correction

## Models & analysis

n/a	Involvement in the study
<input type="checkbox"/>	<input type="checkbox"/> Functional and/or effective connectivity
<input type="checkbox"/>	<input type="checkbox"/> Graph analysis
<input type="checkbox"/>	<input type="checkbox"/> Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis

Multivariate modeling and predictive analysis