

1 **TreeGOER: a database with globally observed environmental ranges for 48,129 tree species**

2 Roeland Kindt

3 [R.Kindt@cifor-icraf.org](mailto:R.Kindt@cifor-icraf.org)

4 <https://orcid.org/0000-0002-7672-0712>

5

6 Trees and forest genetic resources and biodiversity, World Agroforestry, CIFOR-ICRAF

7 Correspondence: World Agroforestry (ICRAF), PO Box 30677, Nairobi 00100, Kenya

8

9 **Abstract**

10 The BIOCLIM algorithm provides a straightforward method to estimate the effects of climate change on the  
11 distribution of species. Estimating the core ranges of species from 5% and 95% quantiles of bioclimatic  
12 variables, the algorithm remains widely used even when more sophisticated methods of species  
13 distribution modelling have become popular. Where sufficient representative observations are available, I  
14 expect that BIOCLIM correctly identifies locations that would not be suitable in a future climate. To  
15 accommodate climate change investigations based on BIOCLIM for 48,129 tree species (a substantial subset  
16 of known tree species), I developed the TreeGOER (Tree Globally Observed Environmental Ranges)  
17 database, providing information on environmental ranges for 38 bioclimatic, 8 soil and 3 topographic  
18 variables. The database can be accessed from: <https://doi.org/10.5281/zenodo.7922928>. Statistics that  
19 include 5% and 95% quantiles were estimated for a cleaned and taxonomically standardized occurrence  
20 data set with different methods of outlier detection, with estimates for roughly 45% of species being based  
21 on 20 or more observation records. Inferred core bioclimatic ranges of species along global temperature  
22 and moisture index gradients and across continents follow the known global distribution of tree diversity  
23 such as its highest levels in moist tropical forests and the 'odd man out' pattern of lower levels in Africa. To  
24 demonstrate how global analyses for large numbers of tree species can easily be done in R with  
25 TreeGOER, here I present two case studies. The first case study investigated latitudinal trends of tree  
26 vulnerability and compared these with previous results obtained for urban trees. The second case study  
27 focused on tropical areas, compared trends in different longitudinal zones and investigated patterns for the  
28 moisture index. TreeGOER is expected to benefit researchers conducting biogeographical and climate  
29 change research for a wide range of tree species at a variety of spatial and temporal scales.

30

## 31 1 Introduction

32

33 Trees are of immense importance to ecological systems, the global economy and to human livelihoods and  
34 wellbeing (Di Sacco et al., [2021](#); Rivers et al., [2022](#)). At least a quarter of known tree species have been  
35 documented to be useful (Kindt et al., [2023](#)), whereby they provide the matrix of many terrestrial  
36 ecosystems, define agroforestry systems and play key roles in climate regulation through carbon and water  
37 cycling (Keppel et al., [2021](#); van Noordwijk et al., [2021](#)). Given their importance and the ensuing climate  
38 change crisis (Ripple et al., [2017](#); Lyon et al., [2021](#)), it has become essential to anticipate the effects of  
39 climate change on their future suitability. Estimating the future suitability of tree species is not only  
40 relevant to devise adaptation strategies for trees in forests and agroforestry systems (Meybeck et al.,  
41 [2021](#)), but also provides insights in mitigation pathways by linking assemblages of future-suitable species  
42 with their carbon sequestration potentials (Jucker et al., [2022](#); Rius et al., [2023](#); Duguma et al., [2023](#)). It is  
43 equally important that massive ecological restoration initiatives such as the UN Decade on Ecosystem  
44 Restoration or the Bonn Challenge (Chapman et al., [2020](#); Höhl et al., [2020](#); van Noordwijk et al., [2020](#))  
45 factor in climate change effects – this has been explicitly included in the *International Principles and*  
46 *Standards for the Practice of Ecological Restoration* via Principle 3 that “*Ecological Restoration Practice Is*  
47 *Informed by Native Reference Ecosystems, while Considering Environmental Change*” (Gann et al., [2019](#)).

48 To estimate the suitability of tree species in future climates, species distribution models (SDMs) need to be  
49 calibrated, and to calibrate these models, observations are required that represent the range of  
50 environmental conditions under which a species can occur. Booth ([2018](#)) has argued that for most tree  
51 species, the only option to learn about climate change impacts are correlative SDMs such as models  
52 described by Guisan et al. ([2017](#)). For each tree species, presence observations are therefore required  
53 which ideally should characterize the full range of environmental conditions where a species can occur.

54 GlobalTreeSearch (GTS) was the first global database aimed at listing all known tree species and is updated  
55 regularly (Beech et al., [2017](#); [https://tools.bgci.org/global\\_tree\\_search.php](https://tools.bgci.org/global_tree_search.php)). GTS uses the tree definition of  
56 IUCN’s Global Tree Specialist Group of “*a woody plant with usually a single stem growing to a height of at*  
57 *least two metres, or if multi-stemmed, then at least one vertical stem five centimetres in diameter at breast*  
58 *height*”. GTS excludes hybrid species, cycads, tree ferns and tree-like Poaceae, Bromeliaceae and Musaceae  
59 species. Updated regularly, the most recent downloadable version listed 57,958 species (version 1.6 of April  
60 2022). Searching for available occurrence records for all known tree species, Serra-Diaz et al. ([2017](#)) found  
61 49,206 tree species with available records from which they retained 15,140 species with at least 20 records  
62 after data cleaning. Keppel et al. ([2021](#); see their Appendix S3) compiled a global standardized list of 58,044  
63 tree species informed partially by GTS version 1.4. Their analysis of georeferenced records available in GBIF,  
64 the largest database of occurrence data available among those they analyzed, showed that 48,970 species  
65 (84.4%) had at least one record.

66 Achieving acceptable SDM calibrations requires a minimum number of observation records, with this  
67 number of records widely debated. Results obtained from Wisz et al. ([2008](#)) suggest that a minimum of 30  
68 records are required in combination with superior modelling algorithms to achieve acceptable SDM  
69 calibration results. Van Proosdij et al. ([2016](#)) in an African study documented lower limits that ranged from  
70 14 for narrow-ranged to 25 for widespread species. Varela et al. ([2014](#)) suggest that for optimal  
71 performance of non-filtered data, 50 observations may be required but that environmental filtering results  
72 in smaller required sample sizes (but obviously filtering requires a larger initial sample size). A substantially  
73 higher number of records is suggested by Feeley and Silman ([2011](#)) as a consequence of the spatio-  
74 temporal aggregation of collections, with their observation that time-sequenced collections of 75-100  
75 occurrences was on average equivalent to 25 randomly subsampled occurrences. Santini et al. ([2021](#)) when  
76 assessing the reliability of SDM in climate change research made the argument to use large sample sizes

77 (tentatively 200 – 500 points) because of uncertainties for most species about ecologically-meaningful  
78 predictor variables, biases in sampling and questions whether a species was in equilibrium with its  
79 environment.

80 Although studies such as Elith et al. (2006) and Wisz et al. (2008) found that more sophisticated SDM  
81 algorithms such as MAXENT or boosted regression trees outperformed BIOCLIM, this pioneering method  
82 remains widely used, especially through the concept of bioclimatic variables that it developed (Booth,  
83 2018). BIOCLIM also remains to be used within the context of ensemble modelling frameworks (Marmion et  
84 al., 2008; Kindt, 2018; Brummit et al., 2020). One of its utilities comes from mapping areas where different  
85 algorithms reach consensus that a species would be suitable, thus showing areas with which the easily-  
86 understood BIOCLIM agrees. A recent example of using a BIOCLIM algorithm to infer the effects of climate  
87 change is the global study of urban tree vulnerability by Esperon-Rodriguez et al (2022); this study will be  
88 discussed in further detail among the case studies. The Climate Assessment Tool of Botanic Gardens  
89 Conservation International that was launched near the end of 2022 ([https://www.bgci.org/resources/bgci-  
90 hosted-data-tools/climate-assessment-tool/](https://www.bgci.org/resources/bgci-hosted-data-tools/climate-assessment-tool/)) also uses a version of the BIOCLIM algorithm, but modified  
91 the system to identify locations near the edge of the known distribution through 1% - 10% and 90% - 99%  
92 percentile ranges, whereas the default BIOCLIM algorithm uses 5%- 95% quantile ranges to identify the  
93 core bioclimatic domain, and minimum – 5% and 95% - maximum ranges to define marginal bioclimatic  
94 domains (e.g., Lindenmayer et al., 1996).

95 Hijmans and Graham (2006) did not recommend using BIOCLIM for climate change investigations and  
96 instead endorsed more complex algorithms such as generalized additive models (GAM) and maximum  
97 entropy models (MAXENT). However, their arguments (as in their discussion of their Figure 3) combined  
98 issues of sample sizes and representative sampling besides methodological differences between  
99 algorithms. They stated that “*Bioclim can be used as a conservative approach, for example, in the context of*  
100 *reserve planning. It will likely underestimate future ranges, but there is a high probability that areas*  
101 *identified as suitable for a species will be correctly identified*”. My opinion is that BIOCLIM may instead  
102 overestimate future ranges compared to other algorithms (such as GAM and MAXENT). I illustrate my  
103 argument via Figure 1 where the environmental niche of a species has an ellipsoid shape as used in  
104 previous theoretical discussions as by Hijmans and Graham (2006), Etherington (2019) or Erickson and  
105 Smith (2023). The same arguments can be made, however, for niches with convex or concave hull shapes,  
106 as used for example in climate change studies by Pironon et al. (2019) or van Zonneveld et al. (2023). With  
107 a large enough sample size where more complex model calibrations can approximate the true ellipsoid  
108 niches well, the BIOCLIM algorithm tends to overestimate suitable conditions in the zones labelled as ‘B1’ in  
109 Figure 1. If one further accepts the additional condition for more complex approaches, that for each  
110 individual explanatory variable the conditions at the tails of the distribution should not be predicted to be  
111 suitable (put alternatively, accepting the ecological justification of the BIOCLIM algorithm and deciding not  
112 to include marginal bioclimatic conditions), then zones labelled as ‘B2’ in Figure 1 would be modelled as not  
113 being suitable regardless of the SDM. The overprediction of suitable conditions in the ‘B1’ zones by  
114 BIOCLIM becomes more prominent in situations where environmental variables are more strongly  
115 correlated (Figure 1a), whereas where variables are correlated less most of the suitable area predicted by  
116 BIOCLIM corresponds to the ‘A’ zone (Figure 1b).

117 A general practice when calibrating species distribution models is to select explanatory variables that are  
118 correlated less with methods such as the Variance Inflation Factor analysis (Ranjitkar et al., 2014; de Sousa  
119 et al., 2019; Fremout et al., 2020). The justification for these selections come from modelling complications  
120 such as overfitting and required sample sizes that multiply with the number of variables used (Erickson and  
121 Smith 2023). A shortcoming of selecting a subset of less-correlated variables, however, is that correlations  
122 may be different in future climates and that therefore different subsets of variables could generate  
123 different future projections (Braunish et al., 2013). My observations from Figure 1 seem to point in a similar

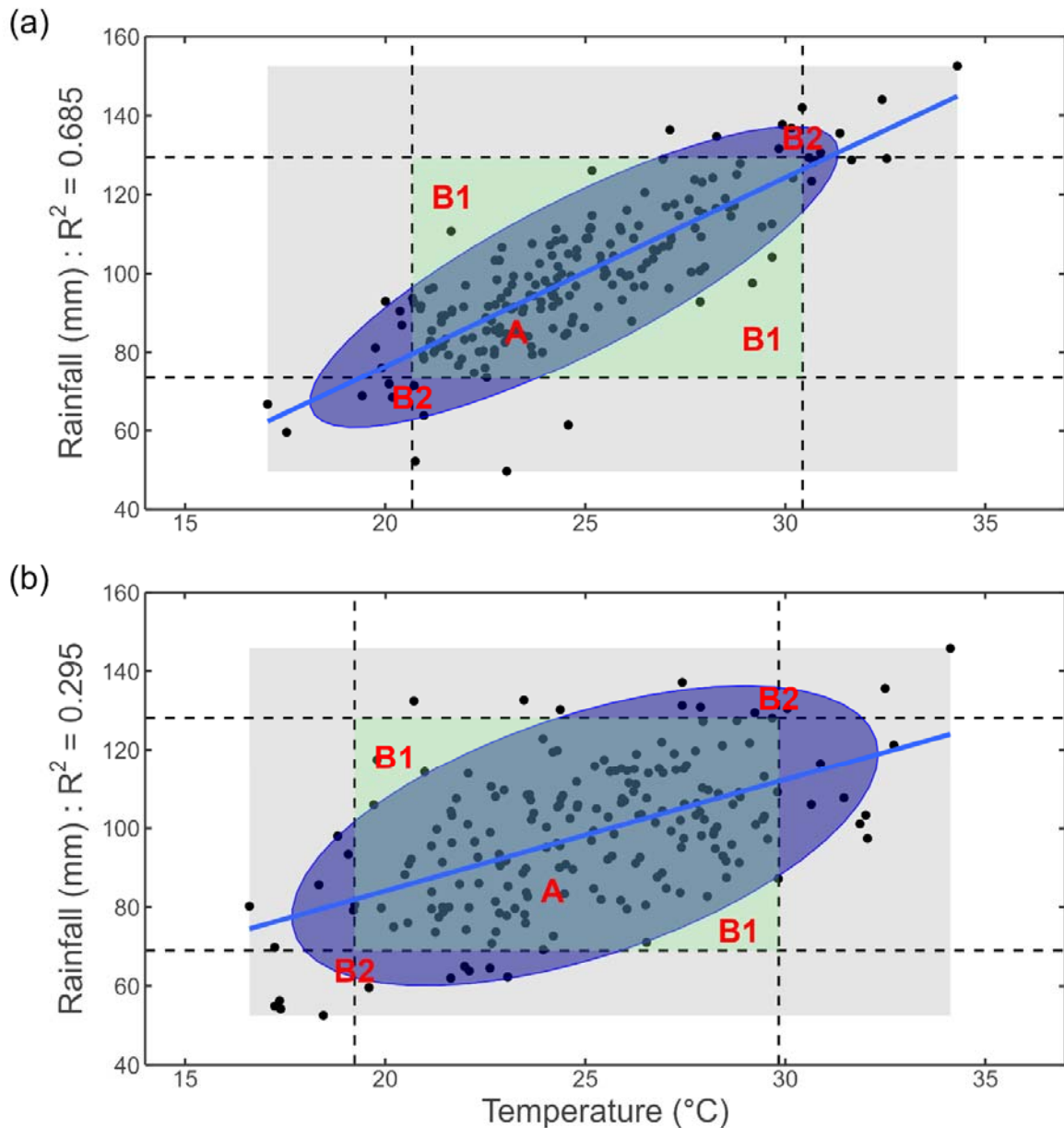
124 direction that using a larger set of more-correlated variables provides an advantage by allowing more  
125 complex algorithms to correctly identify the 'B1' zones that BIOCLIM fails to identify as unsuitable.

126 For smaller or biased samples, I partially agree with Hijmans and Graham (2006) that BIOCLIM may  
127 underestimate future ranges (agreeing only partially as at the same time, BIOCLIM could continue to  
128 overestimate ranges in the 'B1' zones). But in these situations, one may also require that more  
129 sophisticated algorithms do not extrapolate beyond observed ranges depicted by the grey rectangle in  
130 Figure 1 (and see also discussions on SDM transferability as by Charney et al. (2021) for American tree  
131 species). Consequently, also the more sophisticated models would then be expected to underestimate the  
132 future range. (That more sophisticated models may extrapolate is a phenomenon that we experienced  
133 when developing a Climate Change Atlas for Africa (<https://atlas.worldagroforestry.org/>), where  
134 extrapolation from ensemble suitability models tended to result in large area differences if allowed.  
135 Realizing this, we included different versions of maps where extrapolations were allowed and where they  
136 were not, so that users would be aware of such differences.)

137 In conclusion, where sample sizes and their representation approximate the true 90% ranges of a particular  
138 species, I would expect that the BIOCLIM algorithm can be used as a quick but incomplete filter to identify  
139 species that are not suitable (occurring outside the ellipsoid and 90% interval) at a location under baseline  
140 or future conditions. The filter is incomplete by correctly identifying unsuitable species but failing to  
141 identify a larger set of unsuitable species, with more failures to be expected in situations where  
142 correlations between explanatory variables are high.

143 In the next sections, I will describe the TreeGOER database that allows implementation of the BIOCLIM  
144 method for nearly 50,000 tree species. I will also illustrate how the database can be used in two global  
145 investigations of the effects of climate change.

146



147

148 FIGURE 1

149 Comparison of environmental niches estimated by BIOCLIM and by an ellipsoid-fitting model where the  
150 species' fundamental niche is elliptically-shaped following a multivariate normal distribution (a) with  
151 relatively high correlation between explanatory variables and (b) with relatively low correlation. The  
152 distribution of observation points for each subfigure was generated using function `MASS::mvrnorm`  
153 similar to R scripts used by Etherington (2019) but with different covariance matrices and 200 observations  
154 in each simulation. Dashed lines correspond to 5% and 95% quantile limits and delineate the core  
155 bioclimatic niche that BIOCLIM identifies. The ellipsoid niche was estimated via  
156 `ggplot2::stat_ellipse` with level set at 0.90. 'A' indicates a zone where BIOCLIM and the  
157 ellipsoid-fitting model agree that the species is suitable, 'B1' a zone where only BIOCLIM predicts that the  
158 species is suitable, and 'B2' a zone where only the ellipsoid-fitting model predicts the species is suitable.

159

## 2 COMPILING THE DATABASE

160  
161

### 162 Presence observations

163 Presence observations were processed from a recently (within the current decennium, March 2021)  
164 compiled data set of 44,267,164 occurrences from the Global Biodiversity Information Facility (GBIF;  
165 <https://www.gbif.org/>) that were collated by Keppel et al. (2021); their appendices S2 and S3) after they  
166 had compiled a list of 58,044 tree species that was informed by GlobalTreeSearch version 1.4. These  
167 occurrences are available from a GBIF occurrence download (GBIF.org 2021;  
168 <https://doi.org/10.15468/dl.77gcvq>). I selected this data set as GBIF was identified by Keppel et al. (2021)  
169 as holding the largest available occurrence data set for tree species.

170 The data set was subjected to the following series of data quality checks, where records that did not satisfy  
171 listed criteria were removed. First, I retained records where information on geographic coordinates  
172 ('decimalLongitude' and 'decimalLatitude') were available. Second, records were removed where  
173 information was not available on the 'basisOfRecord' and where its value was 'FOSSIL\_SPECIMEN'. Third,  
174 records were removed where information on the collection year was not available. Fourth, observations  
175 with a collection year prior to 1946 were removed, following the suggestion of the documentation of the  
176 `CoordinateCleaner` package (version 2.0-20; Zizka et al., 2019) that GBIF records from and before the  
177 second world war period are often very imprecise. Fifth, records were only retained if the geographical  
178 coordinates corresponded to a terrestrial 30 arc-seconds ( $\approx 1$  km at the equator) raster cell of the  
179 WorldClim 2.1 database (Fick and Hijmans, 2017). Sixth, records were removed if they were flagged by the  
180 `CoordinateCleaner::clean_coordinates` function for the 'capitals', 'centroids', 'equal', 'gbif',  
181 'institutions', 'zeros' and 'duplicates' tests, thereby flagging records such as those corresponding to country  
182 capitals or duplicated coordinate records for the same species.

183 A final series of data quality checks verified that the country location documented by GBIF in the  
184 'countryCode' field corresponded to the country location of Natural Earth (NE) 'Admin 0 – Countries' vector  
185 layers (first I created an 'iso3c' country code variable from the GBIF 'iso2c' country code via the  
186 `countrycode` package version 1.4.0). I retained all records where the countries matched for the 1:110  
187 million NE layer (<https://www.naturalearthdata.com/downloads/110m-cultural-vectors/>; version 5.1.1  
188 downloaded in September 2022). From the records where countries did not match, I retained all records  
189 where the countries matched for the NE 1:10m layer ([https://www.naturalearthdata.com/downloads/10m-  
190 cultural-vectors/](https://www.naturalearthdata.com/downloads/10m-cultural-vectors/); version 5.1.1 downloaded in November 2022). By using the small scale layer first, I  
191 allowed for some buffers around country boundaries, but such buffers can be considered acceptable given  
192 the spatial precision of country GIS layers and locality data (Tack et al., 2022). The NE 1:10 million layer did  
193 not show the distribution of certain GBIF country codes as for example 'GF' (French Guyana, with  
194 corresponding occurrences mainly mapped by NE in the multipolygon with the country code of 'FRA',  
195 indicating France) or 'BQ' (Bonaire, Sint Eustatius and Saba, corresponding in NE to The Netherlands).  
196 Recognizing these situations, I also retained occurrence records for the following matches between the  
197 GBIF and NE country codes: AX-ALD, BQ-NLD, CX-IOA, EH-SAH, GF-FRA, GP-FRA, MQ-FRA, PS-PSX, RE-FRA,  
198 SJ-NOR, SS-SDS and YT-FRA. What was also the case was that some countries and territories mapped by NE  
199 had no corresponding country code in the GBIF data set, for example for the separately mapped Somaliland  
200 and military bases in Cyprus. I therefore also retained occurrence records for country code matches of: CU-  
201 USG, CY-CNM, CY-CYN, CY-WSB and SO-SOL. After realizing that no records had been retained for Namibia, I  
202 discovered that the GBIF country code for Namibia of 'NA' had been interpreted as missing data when  
203 reading in the GBIF data set in R. To allow for records from Namibia, I assumed that all records with  
204 'missing' GBIF country codes and with occurrences between longitudes 10 and 30 degrees East and  
205 latitudes between 15 and 30 South corresponded to observations from Namibia. For these records, the  
206 entire process of quality checking was repeated starting from the second step documented above.

207

## 208 Taxonomy

209 After data quality checks, I created a master list of the 48,518 unique species names from the retained  
210 occurrence records. This result indicated a relatively mild effect of data cleaning by retaining 99.1% of the  
211 48,970 species with at least one observation reported by Keppel et al. (2021) for the same GBIF data set  
212 (<https://doi.org/10.15468/dl.77gcvq>). Via the WorldFlora package (version 1.10; Kindt 2020), the  
213 master list was standardized to World Flora Online (WFO; Borsch et al., 2020; taxonomic backbone version  
214 2021.12 downloaded from <http://www.worldfloraonline.org/downloadData>) and, for those species that  
215 could not be matched to WFO, to the World Checklist of Vascular Plants (WCVP; Govaerts et al., 2021;  
216 taxonomic backbone version 8 downloaded from <http://sftp.kew.org/pub/data-repositories/WCVP/>). Since  
217 these were the same protocols that had been used to create the 2022 version of the Agroforestry Species  
218 Switchboard (Kindt et al., 2022), I first directly matched 47,728 species with species that were encountered  
219 when preparing the *Switchboard*, followed by using the same matching procedures for the remaining 790  
220 species. The *TreeGOER\_Taxonomy* file documents for each of the 48,518 species to which WFO or  
221 WCVP taxon they were standardized, further including other details such as the identifier of the taxonomic  
222 backbone database and whether the match was made as a direct match or a manual match suggested by  
223 fuzzy matching. As some of the submitted names were identified as synonyms, the number of unique  
224 standardized names was 48,129, directly corresponding to the species with documented environmental  
225 ranges in the *TreeGOER* database.

226

## 227 Environmental variables

228 Table S1 in the Supporting Information lists the 51 environmental variables that are covered by *TreeGOER*  
229 together with links to references where more details can be obtained about each variable. The bioclimatic,  
230 soil and topographic variables can be characterized as abiotic variables and as a consequence, *TreeGOER* is  
231 useful particularly to estimate the ‘abiotically potential range’ of a species (Booth, 2016 and references  
232 therein).

233 A set of 19 bioclimatic layers was obtained at 30 arc-seconds resolution from WorldClim 2.1 (Fick and  
234 Hijmans, 2017). Also downloaded from WorldClim 2.1 was the elevation data used when creating the  
235 database. I included this elevation data to allow comparisons with elevation ranges documented in other  
236 databases or floras, but I strongly advise against using elevation as a predictor variable for suitability in  
237 future conditions given its strong correlation with temperature (typically predicted to increase in future  
238 climates whereas elevation remains constant; elevation may also be correlated with other factors that  
239 affect tree physiology such as frequencies of fog; see also the discussion of ‘indirect variables’ by Guisan  
240 and Zimmermann (2000)) – unfortunately this is a mistake that remains to be made in various climate  
241 change studies.

242 With the *envirem* package (version 2.3; Title and Bemmels, 2017), I created an additional set of 18  
243 bioclimatic variables at the same resolution as the WorldClim 2.1 layers (this was also the resolution of the  
244 other environmental layers described below). This was done after downloading all historical monthly  
245 temperature and rainfall layers from WorldClim 2.1 and creating monthly extraterrestrial solar radiation  
246 files for 1985 (at the centre of the 1970 to 2000 period covered by WorldClim 2.1) via  
247 `envirem::ETSolradRasters`. The two topographic variables created by the same authors and made  
248 available via their ENVIREM website (<https://envirem.github.io/>) were also included among the  
249 environmental layers.

250 With function `BiodiversityR::ensemble.PET.season` (version 2.14-3; Kindt and Coe, 2006;  
251 Kindt, 2018) the maximum climatological water deficit (MCWD) was calculated, corresponding to the dry

252 season with the largest difference between precipitation and potential evapotranspiration (PET). As inputs  
253 in the analysis, monthly PET layers were first created via `envirem::monthlyPET`. The calculations  
254 involved a similar procedure as used by Chave et al. (2014) who modelled and used this bioclimatic variable  
255 to model aboveground biomass of tropical trees, but the `BiodiversityR` procedure allows to  
256 distinguish more than one dry season at a location. Zuidema et al. (2022) provided recent evidence for the  
257 importance of MCWD in tree growth with data from a pantropical tree-ring network.

258 Physical and chemical soil properties were obtained via SoilGrids 2.0 (Poggio et al., 2021) by processing  
259 1000 m aggregated data ([https://files.isric.org/soilgrids/latest/data\\_aggregated/1000m/](https://files.isric.org/soilgrids/latest/data_aggregated/1000m/); accessed in  
260 November 2022). I obtained eight soil layers that correspond to all chemical and physical soil properties,  
261 except for 'Coarse Fragments (CFVP)' which I did not include because the Model Efficient Coefficients for  
262 this variable were significantly lower (and seeing also the significantly lower  $R^2$  values reported by Hengl et  
263 al., 2017). Using a similar procedure as described by Hannah et al (2020), for each soil variable the average  
264 was calculated of strata within the top 1 m (0 - 5 cm, 5 - 15 cm, 15 - 30 cm, 30 - 60 cm and 60 - 100 cm).  
265 The average layers were reprojected to the resolution of the bioclimatic layers with `terra::project`  
266 (version 1.6-47; Hijmans 2022), using a bioclimatic layer as template.

267 Also included among the environmental variables were decimal longitude and latitude, directly obtained  
268 from the occurrence data set.

269

## 270 Outlier detection methods

271 I used two methods for outlier detection based on Tukey's (1977) fences method and available via  
272 `BiodiversityR::ensemble.outliers` (version 2.15-1). In one application of the methodology  
273 termed 'method 1' here, the default parameters of the function were used for `n_min` (the minimum  
274 number of environmental variables required to flag an outlier record) of 5 and `fence.k` (the fence  
275 multiplier of the interquartile range) of 2.5. Van Zonneveld et al. (2018) provides the justification for  
276 method 1, which was recently also used by van Zonneveld et al. (2023). As a more strict outlier detection  
277 method, I also used a 'method 2' where `n_min` was set as 2 and `fence.k` as 1.5, the latter as in the  
278 original Tukey method. I used both these methods to flag outliers for the 19 bioclimatic variables obtained  
279 directly via WorldClim 2.1, not including other bioclimatic layers as these were partially calculated from  
280 WorldClim 2.1 data.

281

## 282 Calculation of global environmental ranges and niche breadths

283 Global environmental ranges were determined for each species separately. Allowing for different  
284 synonyms, all observations for the same standardized species name were pooled together and afterwards  
285 duplicated observations from the same 30 arc-seconds grid cell were removed. This procedure reflects a  
286 low spatial thinning process whereby some of the spatial sampling biases of species occurrence records can  
287 be removed (Aiello-Lammens et al., 2015). Except for longitude and latitude that were obtained directly,  
288 values for the other variables were obtained by extracting the 30 arc-seconds raster layers at the locations  
289 of the cleaned occurrence data set via `terra::extract`.

290 After removing outliers flagged by method 1, I calculated the minimum, maximum, median, mean, first  
291 quartile and third quartile values via `base::summary`. Quantile estimates at probabilities of 0.05 (Q05)  
292 and 0.95 (Q95) were calculated via `Qtools::midquantile` (version 1.5.6; Geraci, 2016). Q05 and Q95  
293 limits are used to define the core distribution of a species in the BIOCLIM method (Lindenmayer et al.,  
294 1996; Booth, 2014). Where the number of observations was between 3 and 5000, I calculated the  $P$  value  
295 of a Shapiro-Wilk test of normality via `stats::shapiro.test`.



296 Relative niche breadth (scaled afterwards as percentage) was calculated for each variable and each species  
297 by dividing the species-specific 90% quantile range by the global range; the latter was calculated as the  
298 difference between the maximum of all the Q95 values and the minimum of all the Q05 values.

299 If outlier method 1 identified some outliers, then I also calculated the Q05 and Q95 values for the full data  
300 set where outliers were not removed. If outlier method 2 identified outliers, I calculated the Q05 and Q95  
301 values for the data without the respective outliers.

302

### 303 **Statistical software**

304 When developing TreeGOER and this article, all data processing was done in R (version 4.2.1; R Core  
305 Team, [2022](#)). Figures included here were obtained via the `ggplot2` (version 3.3.6; Wickham, [2009](#)) and  
306 `sf` (version 1.0-8; Pebesma, [2018](#)) packages. Country outlines in the figures correspond to the Natural  
307 Earth 1:110m vector layer used to develop the database.

308

### 309 3 DATABASE OVERVIEW AND ACCESS

310

311 The TreeGOER database documents the bioclimatic ranges for 48,129 species (Table 1). Soil ranges are  
 312 documented for a smaller number of species (46,608 – 47,635; Table S1 in the Supporting Information),  
 313 which I attribute to the mask used in SoilGrids that removes built-up, water and glacier areas (Buchhorn et  
 314 al., 2020). Roughly 45% of species retained 20 observation records or more, 38% retained a minimum of 30  
 315 records (this is one of the thresholds identified as a minimum data requirement for species distribution  
 316 modelling, see introduction), whereas roughly a quarter of the retained species had fewer than 10 records  
 317 (Table 1). Outlier detection method 2 resulted in lower species counts, with roughly a third of species  
 318 having 30 records or more and over 3,700 species having no records. Not removing outliers had a relatively  
 319 mild effect on retained species numbers, with cumulative percentages being maximum one percent higher  
 320 for bin sizes for 10 observations or more. Given this relatively mild effect, I recommend utilizing  
 321 TreeGOER statistics derived from data sets where outliers were removed. In the remainder of this article,  
 322 I will therefore use statistics obtained via Method 1. In the below, I will refer to a ‘range’ as defined by the  
 323 environmental range from Q05 to Q95.

324 Tables S1 and S2 in the Supporting Information provide information on the distribution of niche breadths  
 325 for the different variables. For most of the variables, most species and often more than half of them had  
 326 niche breadths within the 10-25% interval. Notable exceptions for the full data set (Table S1) were bio04,  
 327 continentality, monthCountByTemp10 (Tmo10), Longitude and Latitude that had most species in the 0-  
 328 0.5% interval. For species with at least 30 observations (Table S2), only Tmo10 had most species in the 0-  
 329 0.5% interval, with the second-lowest highest count for latitude in the 2.5-5% bin.

330

331 **TABLE 1.** Number of tree species in TreeGOER for two different methods of outlier detection (Method 1  
 332 and Method 2) and without excluding outliers, calculated for bioclimatic variable bio01. Bin sizes were  
 333 defined by the number of observations per tree species (N). Values between brackets show cumulative  
 334 percentages.

| Observations    | Method 1      | Method 2      | (All)         |
|-----------------|---------------|---------------|---------------|
| N ≥ 5000        | 256 [0.5]     | 234 [0.5]     | 258 [0.5]     |
| 1000 ≥ N > 5000 | 989 [2.6]     | 872 [2.3]     | 997 [2.6]     |
| 500 ≥ N > 1000  | 1,045 [4.8]   | 950 [4.3]     | 1,061 [4.8]   |
| 200 ≥ N > 500   | 2,748 [10.5]  | 2,444 [9.3]   | 2,785 [10.6]  |
| 100 ≥ N > 200   | 3,613 [18.0]  | 3,294 [16.2]  | 3,648 [18.2]  |
| 50 ≥ N > 100    | 5,084 [28.5]  | 4,657 [25.9]  | 5,192 [29.0]  |
| 30 ≥ N > 50     | 4,704 [38.3]  | 4,332 [34.9]  | 4,804 [38.9]  |
| 20 ≥ N > 30     | 4,009 [46.6]  | 3,811 [42.8]  | 4,108 [47.5]  |
| 10 ≥ N > 20     | 7,065 [61.3]  | 6,715 [56.7]  | 7,131 [62.3]  |
| 5 ≥ N > 10      | 6,253 [74.3]  | 6,337 [69.9]  | 6,525 [75.9]  |
| N = 4           | 2,504 [79.5]  | 1,509 [73.0]  | 1,976 [80.0]  |
| N = 3           | 2,585 [84.9]  | 3,577 [80.5]  | 2,371 [84.9]  |
| N = 2           | 3,150 [91.4]  | 2,825 [86.3]  | 3,149 [91.4]  |
| N = 1           | 4,124 [100.0] | 2,849 [92.3]  | 4,124 [100.0] |
| N = 0           | 0 [100.0]     | 3,723 [100.0] | 0 [100.0]     |

335

336

337 **TABLE 2.** Number of tree species in TreeGOER for different global zones defined by the number of months  
 338 with average temperature > 10 °C (Tmo10), the Climatic Moisture Index (CMI), Latitude (LAT) and Longitude  
 339 (LON) and where suitability was predicted from 90% ranges of the variable defining the zone. The CMI  
 340 classification matches dryland zones defined by the aridity index, including dry subhumid (DS), semi-arid  
 341 (SA), arid (A) and hyperarid (HA) zones. Species were counted separately if their 90% range was contained  
 342 entirely in the zone ('endemic'), if they reached their upper distribution limits (Q95) in the zone, if they  
 343 reached their lower distribution limits (Q05) in the zone or if the limits of the zone was contained entirely  
 344 with the species 90% range ('within range'). Values between brackets correspond to species with 30  
 345 observations or more. Figure S3 in the Supporting Information provides a zoomable map that shows the  
 346 global distribution of the different zones.

| Zone  | Endemic         | Upper limit    | Within range  | Lower limit    |
|---|-----------------|----------------|---------------|----------------|
| <b>Number of months with average temperature &gt; 10 °C</b> |                 |                |               |                |
| Tmo10 = 12  | 38,131 [11,929] | 0 [0]          | 7,143 [4,957] | 0 [0]          |
| 8 ≤ Tmo10 < 12  | 382 [48]        | 1,455 [986]    | 2,690 [1,626] | 4,453 [3,331]  |
| 4 ≤ Tmo10 < 8   | 354 [140]       | 462 [349]      | 1,542 [828]   | 2,494 [1,779]  |
| 1 ≤ Tmo10 < 4   | 13 [0]          | 19 [13]        | 1,031 [455]   | 830 [673]      |
| Tmo10 < 1   | 0 [0]           | 0 [0]          | 668 [261]     | 349 [203]      |
| <b>Climatic moisture index</b>                              |                 |                |               |                |
| CMI ≥ 0.5   | 2,367 [64]      | 15,413 [7,431] | 0 [0]         | 0 [0]          |
| 0 ≤ CMI < 0.5   | 8,899 [829]     | 12,458 [6,485] | 5,713 [3,910] | 10,574 [3,982] |
| -0.35 ≤ CMI < 0   | 1,605 [51]      | 3,299 [1,707]  | 7,190 [4,807] | 10,694 [5,481] |
| -0.5 ≤ CMI < -0.35 (DS)                                     | 201 [1]         | 830 [444]      | 5,646 [3,876] | 4,578 [2,497]  |
| -0.8 ≤ CMI < -0.5 (SA)                                      | 625 [193]       | 715 [538]      | 581 [416]     | 5,697 [3,766]  |
| -0.95 ≤ CMI < -0.8 (A)                                      | 198 [47]        | 78 [32]        | 133 [81]      | 1,074 [803]    |
| CMI < -0.95 (HA)  | 0 [0]           | 4 [0]          | 4 [1]         | 175 [85]       |
| <b>Latitude</b>   |                 |                |               |                |
| LAT > 23.5  | 2,880 [1,531]   | 2,760 [1,619]  | 0 [0]         | 0 [0]          |
| -23.5 ≤ LAT ≤ 23.5  | 36,943 [11,520] | 3,168 [2,233]  | 323 [269]     | 2,436 [1,350]  |
| LAT < -23.5   | 2,377 [1,536]   | 0 [0]          | 0 [0]         | 3,491 [2,502]  |
| <b>Longitude</b>  |                 |                |               |                |
| LON < -30   | 21,581 [9,168]  | 0 [0]          | 0 [0]         | 1,246 [748]    |
| -30 ≤ LON ≤ 65  | 8,543 [3,128]   | 303 [202]      | 943 [546]     | 403 [251]      |
| LON > 65  | 16,356 [5,144]  | 1,345 [797]    | 0 [0]         | 0 [0]          |

347

348 I subdivided the globe in five zones based on the number of months with average temperature above 10 °C  
 349 (Tmo10) inspired by the Tmo10 thresholds used in the Köppen-Trewartha climate classification system  
 350 (Belda et al., [2014](#)) to differentiate between tropical, subtropical, temperate, boreal and polar climates  
 351 whereas I ignored dry climate criteria from that system. Analysing 90% ranges, the vast majority of species  
 352 (45,274 representing 94.1% of species in TreeGOER) occurred in the 'tropical' zone where Tmo10 was  
 353 exactly 12. More than 38,000 species (80.8%) were only observed within this zone (Table 2).

354 A second subdivision of the globe that I implemented was via the Climatic Moisture Index (CMI), where I  
 355 expanded the dryland classification system based on the aridity index developed by the United Nations  
 356 Environment Programme ([1997](#)) with three humid zones (Table 2). A large number of species (> 17,000)  
 357 occurred in any of the three humid zones. Nearly 9,000 species exclusively occurred in the zone defined by  
 358 0 ≤ CMI < 0.5 where precipitation was equal to double the PET. Fewer species (< 8,000) occurred in the  
 359 semi-arid zone, mostly species that reached their lowest limits there.

360 The subdivision based on latitude used the approximate positions of the tropics of Cancer and Capricorn. As  
 361 for the results with Tmo10, a great majority of species occurred in tropical areas with most of them (nearly

362 37,000 species) only occurring in that zone (Table 2). Roughly equal amounts of species occurred outside  
363 the tropics.

364 I developed the zones for longitude especially to differentiate between the three main blocks of continental  
365 Africa, the Neo-tropics and South-East Asia of tropical moist forests, the most species-rich terrestrial biome  
366 of the planet (Couvreur [2015](#) ; Hagen et al., [2021](#); Gatti et al., [2022](#)). With the great majority (96.6%) of  
367 species occurring exclusively in one of the longitudinal zones (Table 2), a clear pattern emerged of the  
368 western zone including the highest number of species and the central zone (including Africa) having  
369 substantially lower richness.

370 Table S3 in the Supporting Information cross tabulates the Tmo10 zones with the other zones. The cross  
371 tabulations clearly show that most tree species occur in tropical climates, with one exception for northern  
372 latitudes where more species occur in subtropical climates ( $8 \leq Tmo10 < 12$ ). Given that TreeGOER  
373 contains fewer than 49,000 tree species, the crosstabulation of the tropical zone with the longitudinal  
374 zones approximated patterns observed for the native distribution of over 58,000 tree species in the *State of*  
375 *the World's Trees* (BGCI, [2021](#)) reasonably well, where 23,631 species were observed in the Neotropics and  
376 9,237 species in the Afrotropics. A comparison with the BGCI results from the East was more complicated  
377 given the latitudinal and longitudinal ranges for Indo-Malaya (13,739 species) and Australasia (7,442  
378 species).

379 The latitudinal differences clearly follow global latitudinal gradients in biodiversity as for example reviewed  
380 by Kinlock et al. ([2018](#)) or Nishizawa et al. ([2022](#)). Patterns observed in TreeGOER also agree with  
381 observations that tropical rain forests are the most species-rich terrestrial biomes of the planet and with  
382 the “Odd man out” of lower plant and tree diversity in Africa (Couvreur, [2015](#) ; Raven et al., [2020](#); Hagen et  
383 al., [2021](#); de Miranda et al., [2022](#)).

384 The version of the TreeGOER database described in this article is publicly archived on Zenodo under a CC-  
385 BY 4.0 license so that it can be freely used, shared and modified as long as appropriate credit is given to the  
386 database (citing this article and <https://doi.org/10.5281/zenodo.7922928>). The database is stored as  
387 different text files delimited by the pipe “|” character, including the main file with the environmental  
388 ranges (TreeGOER\_2023.txt), a file with information on taxonomy and standardization methods  
389 (TreeGOER\_Taxonomy.txt; via the ‘SID’ field more taxonomic details can be easily obtained by linking to the  
390 taxonomic backbone databases of WFO or WCVP) and the identifications of the GBIF records involved in  
391 calculating the ranges (TreeGOER\_GBIFID.txt; to reduce file size, for 126 species with over 10,000  
392 observations, a random subset is provided of 10,000 observations). Also provided are files that show the  
393 distribution of species in the Tmo10 and CMI zones with codes reflecting the columns of Table 2; these data  
394 can be used for filtering suitable species for different zones.

395

396

## 397 4 CASE STUDIES

398

399 To showcase possible applications of the TreeGOER database in climate change investigations, I  
400 developed two case studies. In both case studies, baseline bioclimatic data were extracted from the raster  
401 data used to prepare TreeGOER. I used future climate data at 2.5 arc-minutes resolution from the full set  
402 of 23 downscaled General Circulation Model (GCM) outputs available for the 2050s and for shared socio-  
403 economic pathway 3-7.0 (a higher emissions scenario; Meinshausen et al., [2020](#)) from WorldClim 2.1  
404 ([https://www.worldclim.org/data/cmip6/cmip6\\_clim2.5m.html](https://www.worldclim.org/data/cmip6/cmip6_clim2.5m.html); accessed in January-March 2023),  
405 including ACCESS-CM2, ACCESS-ESM1-5, BCC-CSM2-MR, CanESM5, CanESM5-CanOE, CMCC-ESM2, CNRM-  
406 CM6-1, CNRM-CM6-1-HR, CNRM-ESM2-1, EC-Earth3-Veg, EC-Earth3-Veg-LR, GFDL-ESM4, GISS-E2-1-G, GISS-  
407 E2-1-H, INM-CM4-8, INM-CM5-0, IPSL-CM6A-LR, MIROC-ES2L, MIROC6, MPI-ESM1-2-HR, MPI-ESM1-2-LR,  
408 MRI-ESM2-0 and UKESM1-0-LL. Estimates for bioclimatic variables directly available from WorldClim were  
409 calculated as median values after extracting the 23 separate layers at the point locations of the case  
410 studies. The extended set of `envirem` variables was calculated at point locations for the different GCMs  
411 via `BiodiversityR::ensemble.envirem.masterstack`,  
412 `BiodiversityR::ensemble.envirem.solradstack`, and `BiodiversityR::envirem.run`.  
413 Prior to these calculations, monthly layers of extraterrestrial solar radiation were generated for 2050 via  
414 `envirem::ETSolradRasters`. Similar to the bioclimatic data from WorldClim, I calculated medians  
415 afterwards across the 23 GCMs.

416 I selected a subset of environmental variables for the case studies that included `bio01` (mean annual  
417 temperature), `bio12` (annual precipitation), `bio05` (maximum temperature of the warmest month), `bio06`  
418 (minimum temperature of the coldest month), `bio16` (precipitation of the wettest quarter), `bio17`  
419 (precipitation of the driest quarter), `climaticMoistureIndex` (CMI), `monthCountByTemp10` (`Tmo10`) and  
420 `growingDegDays5` (`GDD5`). I selected these variables as they corresponded to bioclimatic variables used  
421 previously for BIOCLIM modelling of the future climatic adaptability of tree species by Nogués-Bravo et al.  
422 ([2014](#)) and Booth ([2016](#)) (`GDD5`, `bio06` and a moisture index) or otherwise used by Esperon-Rodriguez et al.  
423 ([2022](#)) (`bio01`, `bio05`, `bio06`, `bio12` and `bio17`). I expanded the set with `bio16` to add a humid period variable  
424 to the dry period variable of `bio17` analogous to having the cold period and warm period variables of `bio05`  
425 and `bio06`. I also added `Tmo10` as it was used for zonation when describing TreeGOER above, but this  
426 variable was redundant for case study 2.

427 I used a subset of the TreeGOER database where 5% and 95% statistics have been calculated from 20 or  
428 more observations; these were available for 22,448 species (Table 1). Esperon-Rodriguez et al. ([2022](#)) used  
429 the same criterion to create a species subset for their study. From my discussion of Figure 1 in the  
430 introduction, my assumption would be that the results obtained for the case studies are conservative, but  
431 this assumption also implies that the available records were representative for the environmental ranges of  
432 most species for the selected variables, and that these variables characterize the suitable niche well (see a  
433 discussion by Santini et al., [2021](#)).

434

## 435 Case study 1: Testing the effect of climate change for urban locations

436

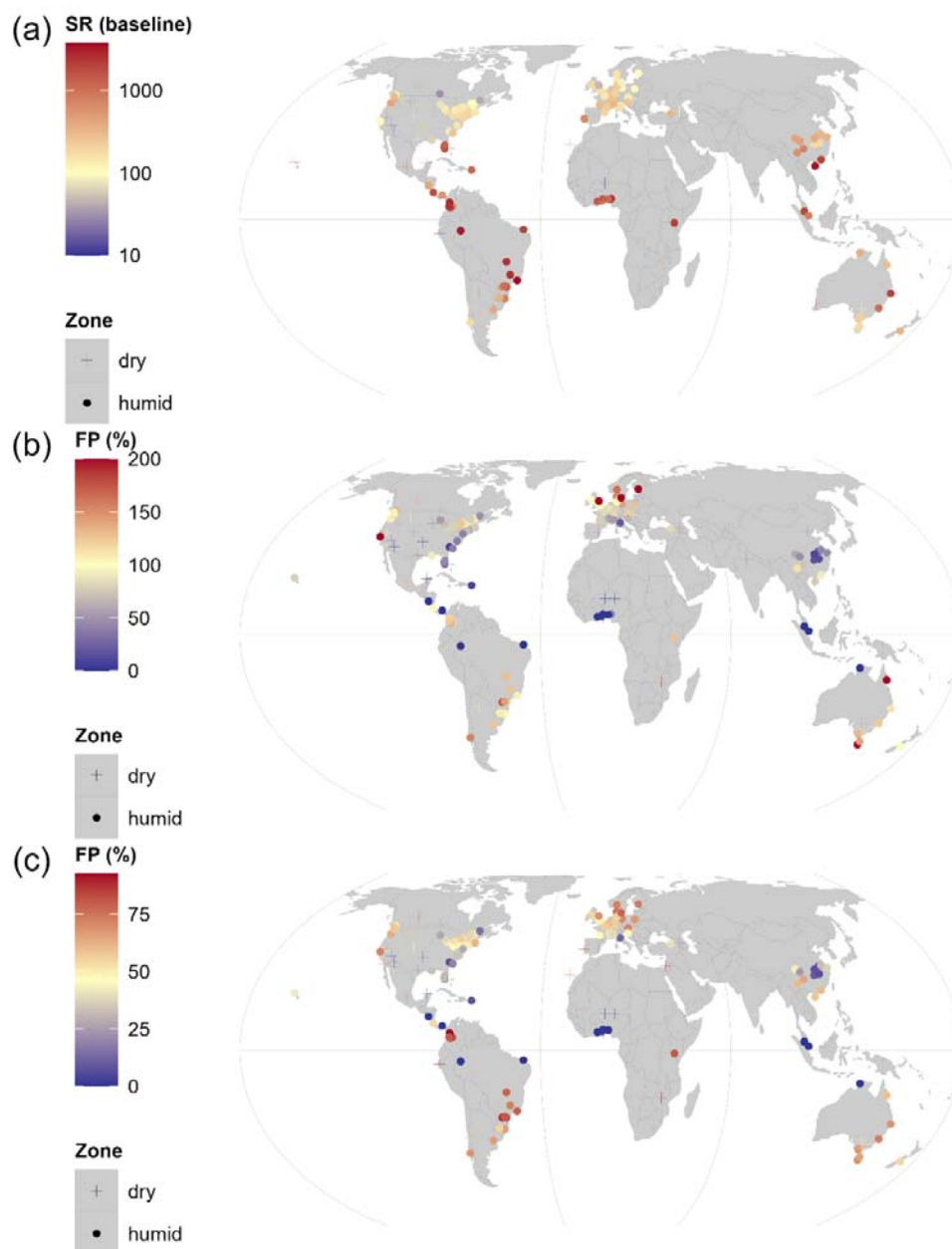
437 Esperon-Rodriguez et al. (2022) conducted a global study on the climate change vulnerability of 3,129  
438 urban trees and shrubs. These authors also used median values for the 2050s, albeit for a different high  
439 emissions scenario (RCP 6.0) and a set of ten GCM projections specific for urban climates (Zhao et al.,  
440 2021). Their analyses were based on 5% and 95% quantiles as well, but they selected 95% thresholds for  
441 temperature-based bioclimatic variables and 5% thresholds for precipitation-based ones (see above which  
442 variables they selected). A fundamental part of their analyses is that their estimates were done for each city  
443 separately for documented city-species observations (Ossola et al., 2020). Where the results that they  
444 present are for 164 cities, I excluded three cities as the geographical coordinates were outside the baseline  
445 30 arc-seconds raster terrestrial coverage. Given these differences in methodologies, my main objective  
446 here, besides showcasing that straightforward climate change analyses via TreeGOER can be conducted,  
447 therefore was to check whether still similar results would be obtained. I especially checked the Esperon-  
448 Rodriguez et al. (2022) finding that cities at low latitudes would be more vulnerable to climate change.

449 I estimated future species richness by two different methods. The first method ('filter 1') identified all  
450 species suitable to the future climate at the city location. The second method ('filter 2') first filtered species  
451 suitable to the baseline climate and then proceeded to further filter those species that remained suitable to  
452 the future climate. The difference between the two methods can be understood as allowing species to  
453 expand their ranges to occupy previously unoccupied sites, or not allowing migration to track suitable  
454 climatic conditions (Boisvert-Marsh and de Blois, 2021; Lima et al., 2022).

455 With filter 1, locations furthest from the equator often were predicted to have larger proportions of species  
456 richness in the future climate (FP; Figure 2b; Data S5 in the Supporting Information). With filter 2, every city  
457 was predicted to lose some species, but generally more so at lower latitudes (Figure 2c).

458 The results shown here (Figure 3) confirm the general pattern of increased vulnerability at lower latitudes  
459 that Esperon-Rodriguez et al. (2022) had documented. When investigating vulnerability (calculated as  $100 - \min(c(FP, 100))$ ) separately for three longitudinal zones, however, in the west the highest  
460 vulnerability manifested at mid-latitudes in the North. What I also observed was that cities near the  
461 equator with lower annual mean temperatures (a pattern that can be explained by their higher altitudes;  
462 see also Liang et al. (2022) and their discussion on the annual mean temperature as dominant predictor of  
463 global tree species richness patterns), such as Nairobi for the central zone and Bogota in the west, had a  
464 significantly lower risk. An effect of the moisture index was not immediately obvious with dryland and  
465 humid locations showing the full range of risks. What should not be forgotten both for the results of  
466 Esperon-Rodriguez et al. (2022) and those shown here is that the latitudinal pattern only explains a  
467 relatively small part of the variation among cities.

469

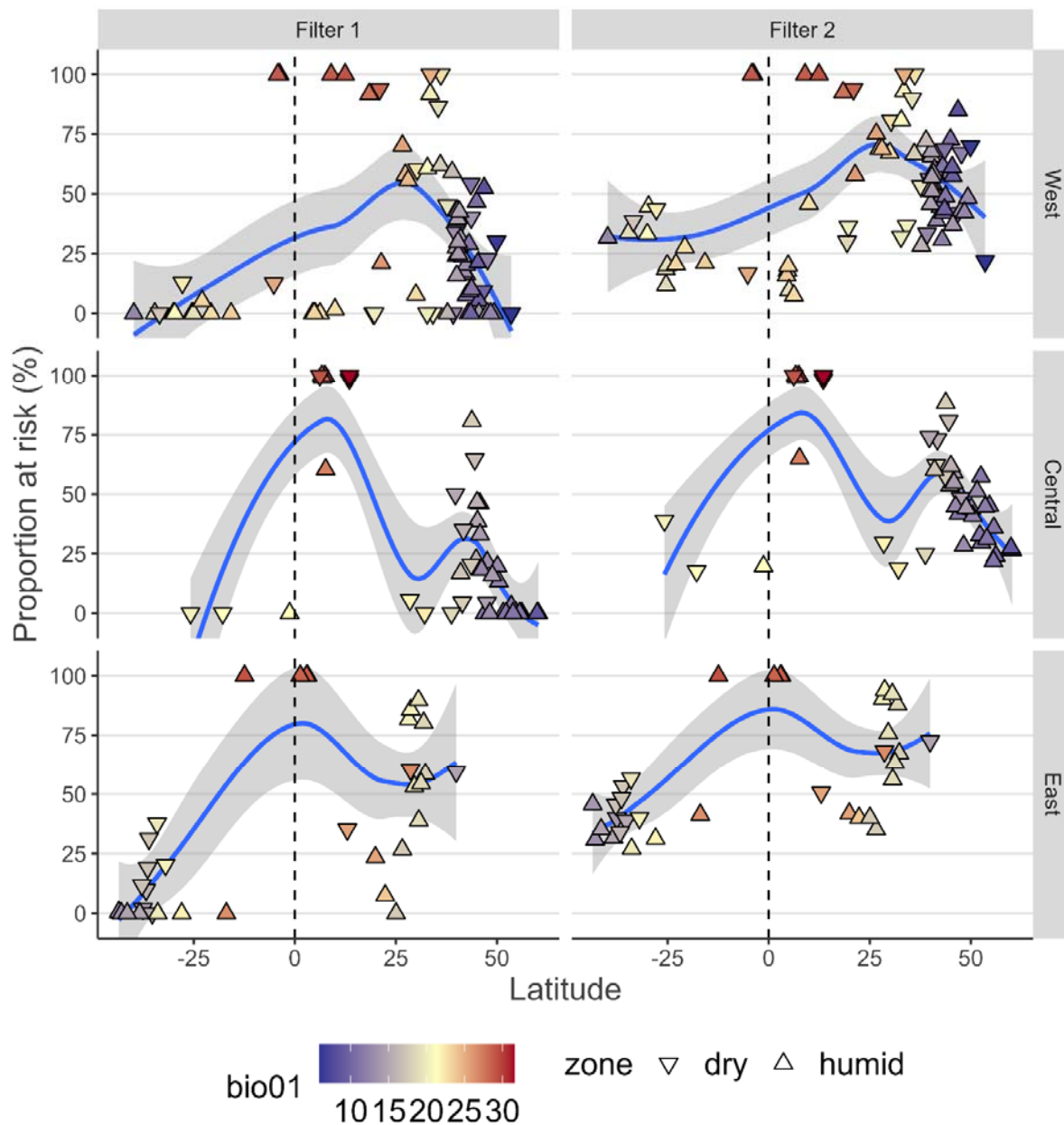


470

471 **FIGURE 2**

472 Predicted species richness (SR) of suitable tree species in the baseline climate (a) and (b-c) predicted future  
473 proportions (FP) suitable in the future climate for 161 locations of cities (Data S5 in the Supporting  
474 Information). Panel (b) shows results where all suitable species were included, with proportions above  
475 200% depicted as 200% for six locations. Panel (c) shows results where only species suitable in the baseline  
476 climate at the city location were investigated for future suitability, a “no migration” scenario. Dryland  
477 locations were defined by a CMI < -0.35 conform global dryland definitions. Map lines obtained from a  
478 Natural Earth 1:110m vector layer (see methods) delineate study areas and do not necessarily depict  
479 accepted national boundaries. The maps were created in R with Equal Earth projection.

480



481

482 **FIGURE 3**

483 Predicted proportions of species at risk for 161 locations of cities (Figure 4; the Central zone has longitudes  
484 between -30 and 65). With filter 1, all suitable species were included. With filter 2, only species suitable in  
485 in the baseline climate at the city location were investigated. The colour scale corresponds to bio01, the mean  
486 annual temperature. Dryland locations were defined by a CMI < -0.35 conform global dryland definitions.  
487 Smoothed regression curves were added via `ggplot2::geom_smooth` (version 3.3.6) with the `loess`  
488 method. Dashed vertical reference lines corresponds to the equator.

489



490

## 491 Case study 2: Testing the effect of climate change for tropical locations

492

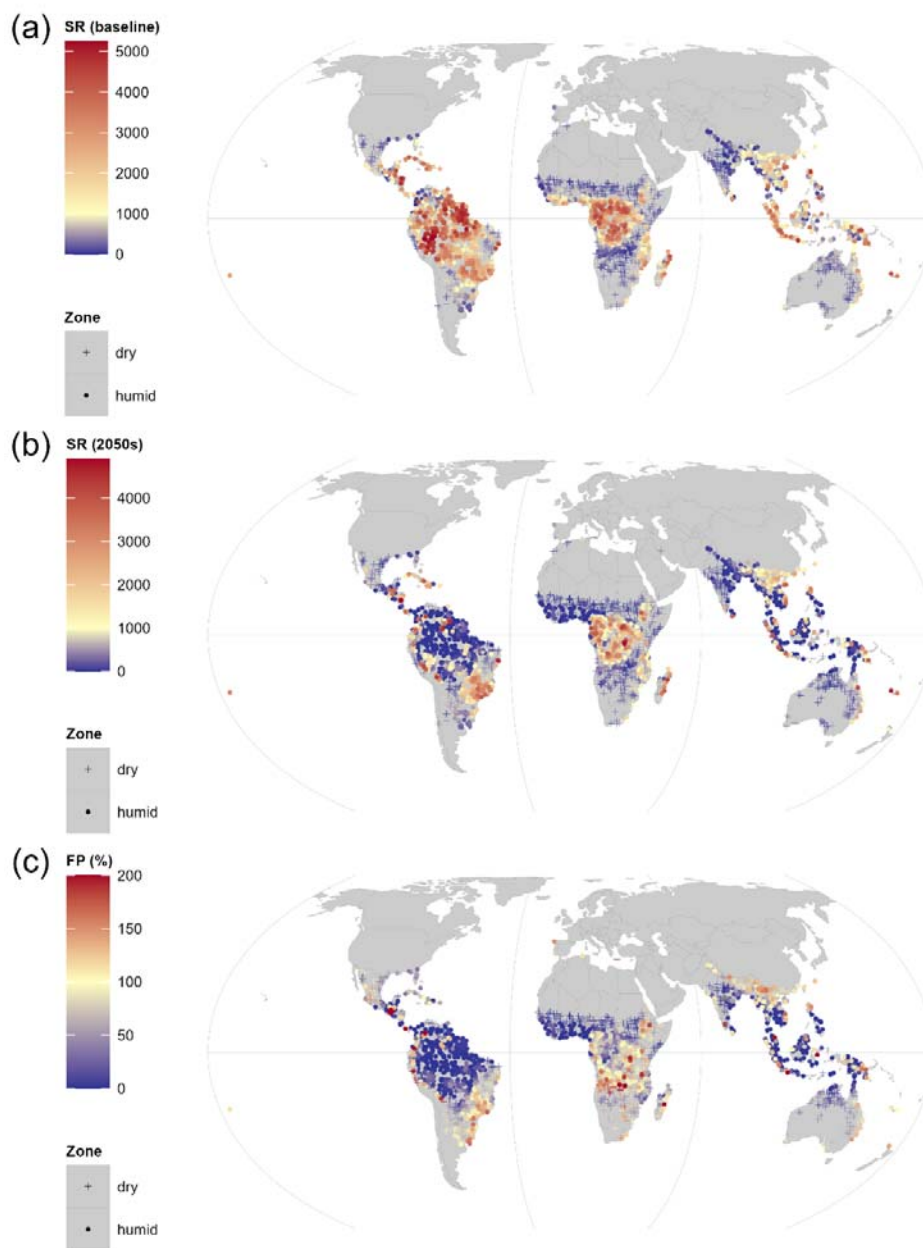
493 In the second case study, I focused the climate change modelling on tropical areas ( $T_{mo10} = 12$ ) where I  
494 also excluded arid and hyperarid zones ( $CMI < -0.8$ ; Table 2). This is the global area that is richest overall in  
495 plant and tree species, especially in humid areas (Couvreur [2015](#); Keppel et al. [2021](#); Gatti et al., [2022](#);  
496 Table S4 in the Supporting Information). I randomly sampled 2000 locations in this study area via  
497 `dismo::randomPoints` (version 1.3-9; Hijmans et al., [2022](#); the function uses a weighted sampling  
498 adjustment for longitude-latitude coordinate systems). Besides mapping overall patterns and differences of  
499 tree species richness in the baseline and future climate, I focused on patterns related to the moisture index  
500 as one of the explanatory variables for differences in species richness (this bioclimatic variable explains  
501 differences in total species richness between zones as documented and discussed in the previous section  
502 (Tables 2 and S4), but no obvious pattern had emerged in the first case study). The moisture index can also  
503 be linked to inter- and intra-specific variations in carbon sequestration potential (Jucker et al., [2022](#)).

504 Results for the baseline and future climate are available from Table S6 in the Supplementary Information  
505 and are shown in figures 4 and 5. Under changed climate conditions for the middle of the 21<sup>st</sup> century and  
506 the higher emissions scenario, significant effects of climate change can be observed and especially so in  
507 South America (Figure 4 bc). These patterns can be seen clearly when comparing results for the different  
508 longitudinal zones separately (Figure 5). Strong declines in species richness can be seen in the west and  
509 especially so in humid zones with  $CMI > 0$  (Figure 5). Also in the east, an overall trend of strong decline in  
510 species richness can be observed. The trend is less pronounced in central Africa, which can be partially  
511 explained by a larger fraction of species being suitable in this zone in future conditions.

512 Figure 5a clearly suggests an effect of water availability, with dryland locations typically having lower  
513 predicted species richness. Figure 5c suggests that various dryland locations may become suitable to a  
514 larger range of species, for example in dryland areas in southern and eastern Africa, but not in west Africa.

515 The highest species richness observed in the baseline climate was 5,252 in a Brazilian location in the  
516 western Amazon (longitude: -72.4042; latitude: -7.9042). All ten observations of richness above 5,000 were  
517 in Brazil. The highest observation of 4,635 species outside South America was in the Philippines at rank 42.  
518 The highest species richness was in a similar range between 5,000 and 6,000 species as shown by Keppel et  
519 al. ([2021](#)) across 463 geographic regions. What is interesting is that the results shown here are based on  
520 point estimates ( $\alpha$ -diversity), whereas the Keppel et al. ([2021](#)) results refer to regional statistics ( $\gamma$ -  
521 diversity). Cai et al. ([2023](#)) who recently conducted a sophisticated modelling exercise to predict plant  
522 diversity arrived at maximum estimates near 5,000 (and also important to note is that the study identified  
523 bioclimatic variables as the most important drivers of diversity; see also Keil and Chase ([2019](#)) and Liang et  
524 al. ([2022](#)) with similar observations for global studies of tree diversity). As my calculations had involved all  
525 candidate species regardless their native distribution and I had expected that this could partially explain  
526 differences with the other studies, I conducted an analysis where only species native to a particular country  
527 were investigated for baseline and future suitability (Figures S6 and S7 in the Supplementary Information;  
528 data on native country distribution was obtained from GlobalTreeSearch; Beech et al., [2017](#)). Generally, the  
529 same patterns can be seen for native species as with the full set of species, but with maximum species  
530 richness values now below 2,500.

531

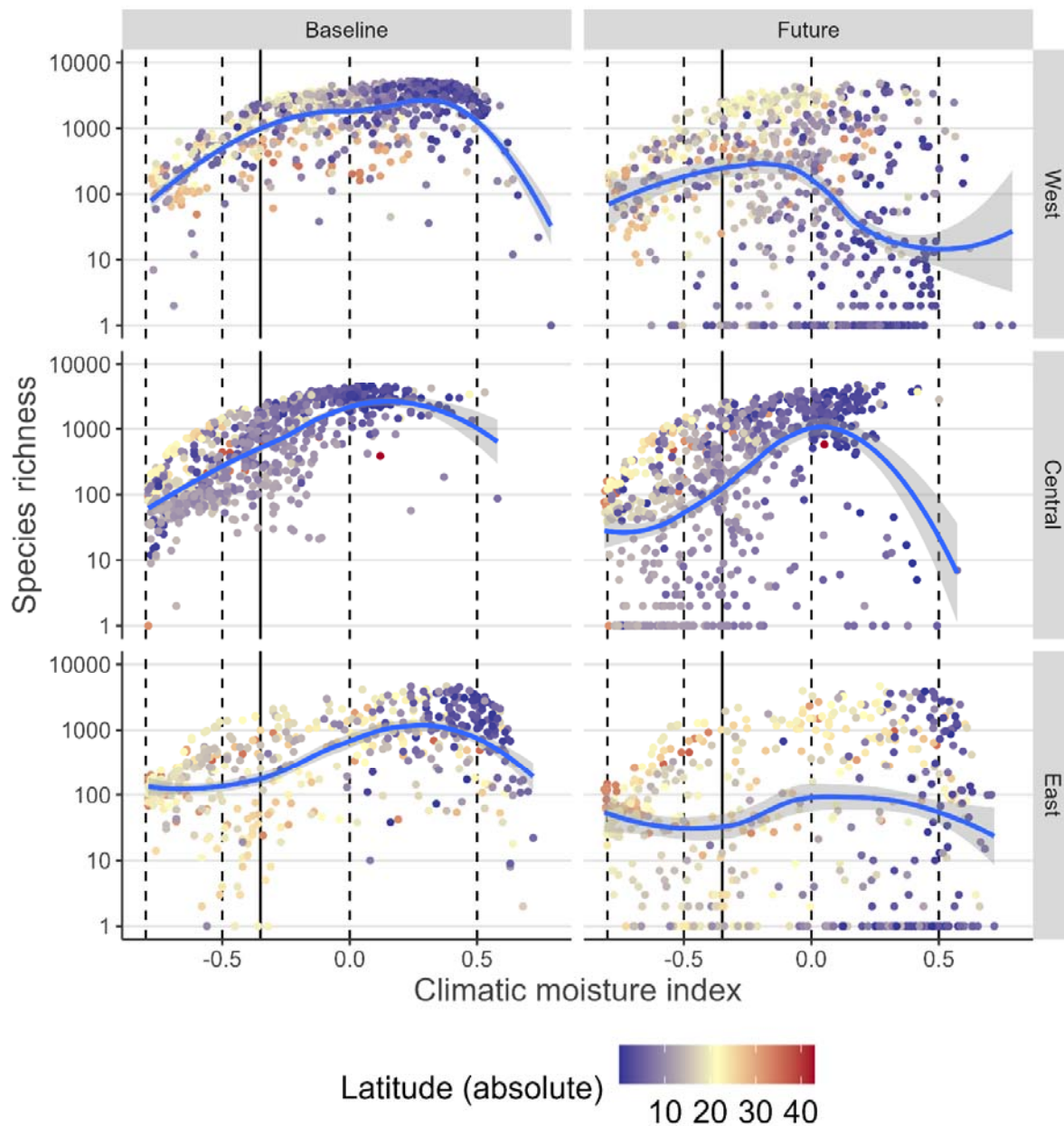


532

533 **FIGURE 4**

534 Predicted species richness (SR) of suitable tree species in the (a) baseline and (b) future climate for 2000  
535 randomly selected locations in tropical areas that exclude (hyper-)arid zones ( $T_{mo10} = 12$  and  $CMI > -0.8$ ;  
536 Data S6 in the Supporting Information). Panel (c) shows the future SR as a proportion of the baseline SR,  
537 with 20 locations with proportions above 200% depicted as 200%. Dryland locations were defined by a CMI  
538  $< -0.35$  conform global dryland definitions. Map lines obtained from a Natural Earth 1:110m vector layer  
539 (see methods) delineate study areas and do not necessarily depict accepted national boundaries. The maps  
540 were created in R with Equal Earth projection.

541



542

543 **FIGURE 5**

544 Predicted species richness of suitable tree species in the baseline and future climate for three longitudinal  
545 zones for 2000 randomly selected locations in tropical areas that exclude (hyper-)arid zones (see Figure 2  
546 and Figure S3 in the Supporting Information; the Central zone has longitudes between -30 and 65). Shown  
547 richness values were transformed by adding 1, hence values of 1 correspond to a predicted species richness  
548 of 0. Smoothed regression curves were added via `ggplot2::geom_smooth` (version 3.3.6) with the  
549 `loess` method. Full vertical reference lines correspond to the CMI = -0.35 upper boundary of dryland  
550 zones, with dashed lines delimiting other CMI zones from Table 3.

551

## 552 5 FUTURE DEVELOPMENTS

553

554 Looking ahead, when sufficient new observations become available in GBIF and particularly for species that  
555 are not yet covered, had low numbers of observations or with remaining large biases in geographical  
556 coverage (Meyer et al., [2016](#)), TreeGOER could be expanded with this new information. With new plants  
557 that continue to be described (Raven et al., [2020](#)) and estimates that the total number of tree species may  
558 be larger than 73,000 (Cazzolla Gatti et al., [2022](#)), it is likely that coverage of TreeGOER could be  
559 expanded substantially. Potentially new observations would become available from locations of novel  
560 climate conditions (Williams and Jackson; [2007](#)), what would benefit estimations of future geographical  
561 ranges. When updating TreeGOER for the new records, taxonomical changes should also be considered.

562 Another reason to update TreeGOER would be when new versions of WorldClim or SoilGrids were  
563 released, as differences between previous versions are known to result in differences in climate change  
564 predictions (Cerasoli et al., [2022](#)). One of the reasons that WorldClim might need a future revision is that it  
565 contains some anomalies for bioclimatic variables that combine temperature and precipitation data (Booth,  
566 [2022](#)).

567 It is likely that I may expand TreeGOER with bioclimatic ranges inferred from CHELSA (Climatologies at  
568 high resolution for the earth's land surface areas) as this data set contains several unique bioclimatic  
569 variables and also used alternative methods of statistical downscaling compared to the ones used by  
570 WorldClim, such as estimating orographic and wind effects on precipitation (Karger et al., [2017](#);  
571 <https://chelsa-climate.org/bioclim/>). What could also be an important addition are observation data from a  
572 proposed global tree trial database (Booth [2023](#)), especially if it was not straightforward to extract these  
573 data directly from such database. As fine-scale gradients in soil conditions have been shown to affect  
574 species distributions (e.g. Bartholomew et al., [2022](#)), it may be wise to compare quantiles of soil variables  
575 observed in TreeGOER with those obtained from high resolution soil surveys – provided databases are  
576 available that document the full range of suitable soil conditions inferred by detailed sampling.

577

578

## 579 AUTHORS' CONTRIBUTIONS

580

581 R.K. conceived the idea and compiled the TreeGOER database by the processes documented here. He also  
582 created the different drafts of the manuscript, including all its analyses and figures.

583

## 584 ACKNOWLEDGEMENTS

585

586 I am indebted to the countless individuals and organizations that were involved in collecting and availing  
587 species occurrence data available via the GBIF occurrence download (GBIF.org 2021;  
588 <https://doi.org/10.15468/dl.77gcvq>). I greatly appreciate the technical feedback provided by Ian Dawson,  
589 Lars Graudal, David Bartholomew, Jens-Peter Lillesø, Fabio Pedercini and Rhett Harrison which helped me  
590 improve the paper, the writing assistance provided by Ian Dawson and for the acquisition of funds and  
591 other support provided by Ramni Jamnadass and Paul Smith. R.K. was supported by the Darwin Initiative to  
592 project DAREX001 of *Developing a Global Biodiversity Standard certification for tree-planting and*  
593 *restoration*, by Norway's International Climate and Forest Initiative through the Royal Norwegian Embassy  
594 in Ethiopia to the *Provision of Adequate Tree Seed Portfolio* project in Ethiopia and by the Green Climate  
595 Fund through the IUCN-led *Transforming the Eastern Province of Rwanda through Adaptation* project.

596

## 597 DATA AVAILABILITY STATEMENT

598

599 The version of the TreeGOER database described in this paper is permanently archived on Zenodo  
600 (<https://doi.org/10.5281/zenodo.7922928>). This repository contains a metadata file describing each field of  
601 the different files of the database. R code and ancillary data needed to replicate the case studies presented  
602 in this paper is available in the Supporting Information.

603

604

## 605 REFERENCES

606

- 607 Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). Sphyn: An r package for  
608 spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38(5), 541–545.  
609 <https://doi.org/10.1111/ecog.01132>
- 610 Bartholomew, D. C., Banin, L. F., Bittencourt, P. R. L., Suis, M. A. F., Mercado, L. M., Nilus, R., Burslem, D. F. R. P., &  
611 Rowland, L. (2022). Differential nutrient limitation and tree height control leaf physiology, supporting niche  
612 partitioning in tropical dipterocarp forests. *Functional Ecology*, 36(8), 2084–2103.  
613 <https://doi.org/10.1111/1365-2435.14094>
- 614 Beech, E., Rivers, M., Oldfield, S., & Smith, P. P. (2017). GlobalTreeSearch: The first complete global database of tree  
615 species and country distributions. *Journal of Sustainable Forestry*, 36(5), 454–489.  
616 <https://doi.org/10.1080/10549811.2017.1310049>
- 617 Belda, M., Holtanová, E., Halenka, T., & Kalvová, J. (2014). Climate classification revisited: From Köppen to Trewartha.  
618 *Climate Research*, 59(1), 1–13. <https://doi.org/10.3354/cr01204>
- 619 BGCI. (2021). *State of the world's trees*. Botanic Gardens Conservation International.
- 620 BGCI. (2022). *GlobalTreeSearch online database (Version 1. 6)* [Data set]. Botanic Gardens Conservation International.  
621 <https://doi.org/10.13140/RG.2.2.34206.61761>
- 622 Boisvert-Marsh, L., & Blois, S. (2021). Unravelling potential northward migration pathways for tree species under  
623 climate change. *Journal of Biogeography*, 48(5), 1088–1100. <https://doi.org/10.1111/jbi.14060>
- 624 Booth, T. H. (2016). Estimating potential range and hence climatic adaptability in selected tree species. *Forest Ecology  
625 and Management*, 366, 175–183. <https://doi.org/10.1016/j.foreco.2016.02.009>
- 626 Booth, T. H. (2018a). Species distribution modelling tools and databases to assist managing forests under climate  
627 change. *Forest Ecology and Management*, 430, 196–203. <https://doi.org/10.1016/j.foreco.2018.08.019>
- 628 Booth, T. H. (2018b). Why understanding the pioneering and continuing contributions of BIOCLIM to species  
629 distribution modelling is important. *Austral Ecology*, 43(8), 852–860. <https://doi.org/10.1111/aec.12628>
- 630 Booth, T. H. (2022). Checking bioclimatic variables that combine temperature and precipitation data before their use  
631 in species distribution models. *Austral Ecology*, 47(7), 1506–1514. <https://doi.org/10.1111/aec.13234>
- 632 Booth, T. H. (2023). The need for a global tree trial database. *New Forests*, 54(1), 1–7.  
633 <https://doi.org/10.1007/s11056-022-09940-y>
- 634 Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). BIOCLIM: The first species distribution modelling  
635 package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions*,  
636 20(1), 1–9. <https://doi.org/10.1111/ddi.12144>
- 637 Borsch, T., Berendsohn, W., Dalcin, E., Delmas, M., Demissew, S., Elliott, A., Fritsch, P., Fuchs, A., Geltman, D., Güner,  
638 A., Haevermans, T., Knapp, S., Roux, M. M., Loizeau, P., Miller, C., Miller, J., Miller, J. T., Palese, R., Paton, A.,  
639 ... Zamora, N. (2020). World Flora Online: Placing taxonomists at the heart of a definitive and comprehensive  
640 global resource on the world's plants. *TAXON*, 69(6), 1311–1341. <https://doi.org/10.1002/tax.12373>
- 641 Braunisch, V., Coppes, J., Arlettaz, R., Suchant, R., Schmid, H., & Bollmann, K. (2013). Selecting from correlated climate  
642 variables: A major source of uncertainty for predicting species distributions under climate change. *Ecography*,  
643 36(9), 971–983. <https://doi.org/10.1111/j.1600-0587.2013.00138.x>
- 644 Brummitt, N., Araújo, A. C., & Harris, T. (2021). Areas of plant diversity—What do we know? *PLANTS, PEOPLE, PLANET*,  
645 3(1), 33–44. <https://doi.org/10.1002/ppp3.10110>
- 646 Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., & Smets, B. (2020). Copernicus global land cover  
647 layers—Collection 2. *Remote Sensing*, 12(6), 1044. <https://doi.org/10.3390/rs12061044>
- 648 Cai, L., Kreft, H., Taylor, A., Denelle, P., Schrader, J., Essl, F., Van Kleunen, M., Pergl, J., Pyšek, P., Stein, A., Winter, M.,  
649 Barcelona, J. F., Fuentes, N., Inderjit, Karger, D. N., Kartesz, J., Kuprijanov, A., Nishino, M., Nickrent, D., ...  
650 Weigelt, P. (2023). Global models and predictions of plant diversity based on advanced machine learning  
651 techniques. *New Phytologist*, 237(4), 1432–1445. <https://doi.org/10.1111/nph.18533>
- 652 Cazzolla Gatti, R., Reich, P. B., Gamarra, J. G. P., Crowther, T., Hui, C., Morera, A., Bastin, J.-F., de-Miguel, S., Nabuurs,

- 653 G.-J., Svenning, J.-C., Serra-Diaz, J. M., Merow, C., Enquist, B., Kamenetsky, M., Lee, J., Zhu, J., Fang, J., Jacobs,  
654 D. F., Pijanowski, B., ... Liang, J. (2022). The number of tree species on Earth. *Proceedings of the National*  
655 *Academy of Sciences*, 119(6), e2115329119. <https://doi.org/10.1073/pnas.2115329119>
- 656 Cerasoli, F., D'Alessandro, P., & Biondi, M. (2022). Worldclim 2.1 versus Worldclim 1.4: Climatic niche and grid  
657 resolution affect between-version mismatches in Habitat Suitability Models predictions across Europe.  
658 *Ecology and Evolution*, 12(2). <https://doi.org/10.1002/ece3.8430>
- 659 Chapman, M., Walker, W. S., Cook-Patton, S. C., Ellis, P. W., Farina, M., Griscom, B. W., & Baccini, A. (2020). Large  
660 climate mitigation potential from adding trees to agricultural lands. *Global Change Biology*, 26(8), 4357–4365.  
661 <https://doi.org/10.1111/gcb.15121>
- 662 Charney, N. D., Record, S., Gerstner, B. E., Merow, C., Zarnetske, P. L., & Enquist, B. J. (2021). A test of species  
663 distribution model transferability across environmental and geographic space for 108 western north  
664 american tree species. *Frontiers in Ecology and Evolution*, 9.  
665 <https://www.frontiersin.org/articles/10.3389/fevo.2021.689295>
- 666 Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M. S., Delitti, W. B. C., Duque, A., Eid, T., Fearnside,  
667 P. M., Goodman, R. C., Henry, M., Martínez-Yrizar, A., Mugasha, W. A., Muller-Landau, H. C., Mencuccini, M.,  
668 Nelson, B. W., Ngomanda, A., Nogueira, E. M., Ortiz-Malavassi, E., ... Vieilledent, G. (2014). Improved  
669 allometric models to estimate the aboveground biomass of tropical trees. *Global Change Biology*, 20(10),  
670 3177–3190. <https://doi.org/10.1111/gcb.12629>
- 671 Couvreur, T. L. P. (2015). Odd man out: Why are there fewer plant species in African rain forests? *Plant Systematics*  
672 *and Evolution*, 301(5), 1299–1313. <https://doi.org/10.1007/s00606-014-1180-z>
- 673 de Sousa, K., van Zonneveld, M., Holmgren, M., Kindt, R., & Ordoñez, J. C. (2019). The future of coffee and cocoa  
674 agroforestry in a warmer Mesoamerica. *Scientific Reports*, 9(1), 8828. <https://doi.org/10.1038/s41598-019-45491-7>
- 676 Di Sacco, A., Hardwick, K. A., Blakesley, D., Brancalion, P. H. S., Breman, E., Cecilio Rebola, L., Chomba, S., Dixon, K.,  
677 Elliott, S., Ruyonga, G., Shaw, K., Smith, P., Smith, R. J., & Antonelli, A. (2021). Ten golden rules for  
678 reforestation to optimize carbon sequestration, biodiversity recovery and livelihood benefits. *Global Change*  
679 *Biology*, 27(7), 1328–1348. <https://doi.org/10.1111/gcb.15498>
- 680 Duguma, L. A., Minang, P. A., Watson, C., Nath, A. J., Muthee, K. W., van Noordwijk, M., Mutune, J. M., & Sileshi, G. W.  
681 (2023). Agroforestry as a key intervention to achieve nationally determined contribution (Ndc) targets. In J. C.  
682 Dagar, S. R. Gupta, & G. W. Sileshi (Eds.), *Agroforestry for Sustainable Intensification of Agriculture in Asia and*  
683 *Africa* (pp. 641–664). Springer Nature. [https://doi.org/10.1007/978-981-19-4602-8\\_19](https://doi.org/10.1007/978-981-19-4602-8_19)
- 684 Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J.,  
685 Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC.  
686 M. Overton, J., Townsend Peterson, A., ... E. Zimmermann, N. (2006). Novel methods improve prediction of  
687 species' distributions from occurrence data. *Ecography*, 29(2), 129–151.  
688 <https://doi.org/10.1111/j.2006.0906-7590.04596.x>
- 689 Erickson, K. D., & Smith, A. B. (2023). Modeling the rarest of the rare: A comparison between multi-species  
690 distribution models, ensembles of small models, and single-species models at extremely low sample sizes.  
691 *Ecography*, e06500. <https://doi.org/10.1111/ecog.06500>
- 692 Esperon-Rodriguez, M., Tjoelker, M. G., Lenoir, J., Baumgartner, J. B., Beaumont, L. J., Nipperess, D. A., Power, S. A.,  
693 Richard, B., Rymer, P. D., & Gallagher, R. V. (2022). Climate change increases global risk to urban forests.  
694 *Nature Climate Change*, 12(10), 950–955. <https://doi.org/10.1038/s41558-022-01465-8>
- 695 Etherington, T. R. (2019). Mahalanobis distances and ecological niche modelling: Correcting a chi-squared probability  
696 error. *PeerJ*, 7, e6678. <https://doi.org/10.7717/peerj.6678>
- 697 Feeley, K. J., & Silman, M. R. (2011). Keep collecting: Accurate species distribution modelling requires more collections  
698 than previously thought: Temporal autocorrelated biases necessitate more collections. *Diversity and*  
699 *Distributions*, 17(6), 1132–1140. <https://doi.org/10.1111/j.1472-4642.2011.00813.x>
- 700 Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas.  
701 *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- 702 Fremout, T., Thomas, E., Gaisberger, H., Van Meerbeek, K., Muenchow, J., Briers, S., Gutierrez-Miranda, C. E.,  
703 Marcelo-Peña, J. L., Kindt, R., Atkinson, R., Cabrera, O., Espinosa, C. I., Aguirre-Mendoza, Z., & Muys, B.

- 704 (2020). Mapping tree species vulnerability to multiple threats as a guide to restoration and conservation of  
705 tropical dry forests. *Global Change Biology*, 26(6), 3552–3568. <https://doi.org/10.1111/gcb.15028>
- 706 Gann, G. D., McDonald, T., Walder, B., Aronson, J., Nelson, C. R., Jonson, J., Hallett, J. G., Eisenberg, C., Guariguata, M.  
707 R., Liu, J., Hua, F., Echeverría, C., Gonzales, E., Shaw, N., Decler, K., & Dixon, K. W. (2019). International  
708 principles and standards for the practice of ecological restoration. Second edition. *Restoration Ecology*,  
709 27(S1). <https://doi.org/10.1111/rec.13035>
- 710 GBIF.org (16 March 2021) GBIF Occurrence Download <https://doi.org/10.15468/dl.77gcvq>
- 711 Geraci M (2022). Qtools: Utilities for quantiles. R package version 1.5.6. Available at [https://cran.r-](https://cran.r-project.org/package=Qtools)  
712 [project.org/package=Qtools](https://cran.r-project.org/package=Qtools).
- 713 Geraci, M. (2016). The r journal: Qtools: a collection of models and tools for quantile inference. *The R Journal*, 8(2),  
714 117–138. <https://doi.org/10.32614/RJ-2016-037>
- 715 Govaerts, R., Nic Lughadha, E., Black, N., Turner, R., & Paton, A. (2021). The World Checklist of Vascular Plants, a  
716 continuously updated resource for exploring global plant diversity. *Scientific Data*, 8(1), 215.  
717 <https://doi.org/10.1038/s41597-021-00997-6>
- 718 Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*,  
719 135(2), 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- 720 Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in r*  
721 (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781139028271>
- 722 Hagen, O., Skeels, A., Onstein, R. E., Jetz, W., & Pellissier, L. (2021). Earth history events shaped the evolution of  
723 uneven biodiversity across tropical moist forests. *Proceedings of the National Academy of Sciences*, 118(40),  
724 e2026347118. <https://doi.org/10.1073/pnas.2026347118>
- 725 Hannah, L., Roehrdanz, P. R., Marquet, P. A., Enquist, B. J., Midgley, G., Foden, W., Lovett, J. C., Corlett, R. T., Corcoran,  
726 D., Butchart, S. H. M., Boyle, B., Feng, X., Maitner, B., Fajardo, J., McGill, B. J., Merow, C., Morueta-Holme, N.,  
727 Newman, E. A., Park, D. S., ... Svenning, J. (2020). 30% land conservation and climate action reduces tropical  
728 extinction risk by more than 50%. *Ecography*, 43(7), 943–953. <https://doi.org/10.1111/ecog.05166>
- 729 Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.  
730 N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J.  
731 G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information  
732 based on machine learning. *PLOS ONE*, 12(2), e0169748. <https://doi.org/10.1371/journal.pone.0169748>
- 733 Hijmans, R. J., & Graham, C. H. (2006). The ability of climate envelope models to predict the effect of climate change  
734 on species distributions: COMPARING CLIMATE ENVELOPE AND MECHANISTIC MODELS. *Global Change*  
735 *Biology*, 12(12), 2272–2281. <https://doi.org/10.1111/j.1365-2486.2006.01256.x>
- 736 Höhl, M., Ahimbisibwe, V., Stanturf, J. A., Elsasser, P., Kleine, M., & Bolte, A. (2020). Forest landscape restoration—  
737 What generates failure and success? *Forests*, 11(9), 938. <https://doi.org/10.3390/f11090938>
- 738 Jucker, T., Fischer, F. J., Chave, J., Coomes, D. A., Caspersen, J., Ali, A., Loubota Panzou, G. J., Feldpausch, T. R., Falster,  
739 D., Usoltsev, V. A., Adu-Bredu, S., Alves, L. F., Aminpour, M., Angoboy, I. B., Anten, N. P. R., Antin, C., Askari,  
740 Y., Muñoz, R., Ayyappan, N., ... Zavala, M. A. (2022). Tallo: A global tree allometry and crown architecture  
741 database. *Global Change Biology*, 28(17), 5254–5268. <https://doi.org/10.1111/gcb.16302>
- 742 Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P., &  
743 Kessler, M. (2017). Climatologies at high resolution for the earth's land surface areas. *Scientific Data*, 4(1),  
744 170122. <https://doi.org/10.1038/sdata.2017.122>
- 745 Keil, P., & Chase, J. M. (2019). Global patterns and drivers of tree diversity integrated across a continuum of spatial  
746 grains. *Nature Ecology & Evolution*, 3(3), 390–399. <https://doi.org/10.1038/s41559-019-0799-0>
- 747 Keppel, G., Craven, D., Weigelt, P., Smith, S. A., Van Der Sande, M. T., Sandel, B., Levin, S. C., Kreft, H., & Knight, T. M.  
748 (2021). Synthesizing tree biodiversity data to understand global patterns and processes of vegetation. *Journal*  
749 *of Vegetation Science*, 32(3). <https://doi.org/10.1111/jvs.13021>
- 750 Kindt, R. (2018). Ensemble species distribution modelling with transformed suitability values. *Environmental Modelling*  
751 *& Software*, 100, 136–145. <https://doi.org/10.1016/j.envsoft.2017.11.009>
- 752 Kindt, R. (2020). WorldFlora: An R package for exact and fuzzy matching of plant names against the World Flora Online



- 753 taxonomic backbone data. *Applications in Plant Sciences*, 8(9). <https://doi.org/10.1002/aps3.11388>
- 754 Kindt, R., & Coe, R. (2005). *Tree diversity analysis: A manual and software for common statistical methods for*  
755 *ecological and biodiversity studies*. World Agroforestry Centre.
- 756 Kindt, R., Graudal, L., Lillesø, J.-P., Pedercini, F., Smith, P., & Jamnadass, R. (2022). *GlobalUsefulNativeTrees, a*  
757 *database of 14,014 tree species and their uses, supports synergies between biodiversity recovery and local*  
758 *livelihoods in landscape restoration*. bioRxiv. <https://doi.org/10.1101/2022.11.25.517923>
- 759 Kindt, R., Innocent, J., Dawson, I. K., Graudal, L., Lillesø, J.-P., Ordonez, J., & Jamnadass, L. (2022). *Agroforestry Species*  
760 *Switchboard: A synthesis of information sources to support tree research and development activities. Version*  
761 *3.0*. World Agroforestry | Transforming Lives and Landscapes with Trees; World Agroforestry.  
762 <https://www.worldagroforestry.org/output/agroforestry-species-switchboard-30>
- 763 Kinlock, N. L., Prowant, L., Herstoff, E. M., Foley, C. M., Akin-Fajiyi, M., Bender, N., Umarani, M., Ryu, H. Y., Şen, B.,  
764 Gurevitch, J., & Pither, J. (2018). Explaining global variation in the latitudinal diversity gradient: Meta-analysis  
765 confirms known patterns and uncovers new ones. *Global Ecology and Biogeography*, 27(1), 125–141.  
766 <https://doi.org/10.1111/geb.12665>
- 767 Liang, J., Gamarra, J. G. P., Picard, N., Zhou, M., Pijanowski, B., Jacobs, D. F., Reich, P. B., Crowther, T. W., Nabuurs, G.-  
768 J., de-Miguel, S., Fang, J., Woodall, C. W., Svenning, J.-C., Jucker, T., Bastin, J.-F., Wiser, S. K., Slik, F., Hérault,  
769 B., Alberti, G., ... Hui, C. (2022). Co-limitation towards lower latitudes shapes global forest diversity gradients.  
770 *Nature Ecology & Evolution*, 6(10), 1423–1437. <https://doi.org/10.1038/s41559-022-01831-x>
- 771 Lima, V. P., de Lima, R. A. F., Joner, F., Siddique, I., Raes, N., & ter Steege, H. (2022). Climate change threatens native  
772 potential agroforestry plant species in Brazil. *Scientific Reports*, 12(1), 2267. <https://doi.org/10.1038/s41598-022-06234-3>
- 773
- 774 Lyon, C., Saupe, E. E., Smith, C. J., Hill, D. J., Beckerman, A. P., Stringer, L. C., Marchant, R., McKay, J., Burke, A.,  
775 O'Higgins, P., Dunhill, A. M., Allen, B. J., Riel-Salvatore, J., & Aze, T. (2022). Climate change research and  
776 action must look beyond 2100. *Global Change Biology*, 28(2), 349–361. <https://doi.org/10.1111/gcb.15871>
- 777 Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in  
778 predictive species distribution modelling. *Diversity and Distributions*, 15(1), 59–69.  
779 <https://doi.org/10.1111/j.1472-4642.2008.00491.x>
- 780 Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A.,  
781 Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P. B., Luderer, G., Meinshausen, N., Montzka, S. A.,  
782 Rayner, P. J., Reimann, S., ... Wang, R. H. J. (2020). The shared socio-economic pathway (Ssp) greenhouse gas  
783 concentrations and their extensions to 2500. *Geoscientific Model Development*, 13(8), 3571–3605.  
784 <https://doi.org/10.5194/gmd-13-3571-2020>
- 785 Meybeck, A., C., L. M., & V., G. (2021). *Adaptation to climate change with forests, trees and agroforestry*. Center for  
786 International Forestry Research (CIFOR). <https://doi.org/10.17528/cifor/008222>
- 787 Meyer, C., Weigelt, P., & Kreft, H. (2016). Multidimensional biases, gaps and uncertainties in global plant occurrence  
788 information. *Ecology Letters*, 19(8), 992–1006. <https://doi.org/10.1111/ele.12624>
- 789 *Natural Earth*» 1:110m Cultural Vectors—Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales.  
790 (2022). <https://www.naturalearthdata.com/downloads/110m-cultural-vectors/>
- 791 Nogués-Bravo, D., Pulido, F., Araújo, M. B., Diniz-Filho, J. A. F., García-Valdés, R., Kollmann, J., Svenning, J.-C.,  
792 Valladares, F., & Zavala, M. A. (2014). Phenotypic correlates of potential range size and range filling in  
793 European trees. *Perspectives in Plant Ecology, Evolution and Systematics*, 16(5), 219–227.  
794 <https://doi.org/10.1016/j.ppees.2014.05.005>
- 795 Ossola, A., Hoepfner, M. J., Burley, H. M., Gallagher, R. V., Beaumont, L. J., & Leishman, M. R. (2020). The Global  
796 Urban Tree Inventory: A database of the diverse tree flora that inhabits the world's cities. *Global Ecology and*  
797 *Biogeography*, 29(11), 1907–1914. <https://doi.org/10.1111/geb.13169>
- 798 Pebesma, E. (2018). Simple features for r: Standardized support for spatial vector data. *The R Journal*, 10(1), 439.  
799 <https://doi.org/10.32614/RJ-2018-009>
- 800 Pironon, S., Etherington, T. R., Borrell, J. S., Kühn, N., Macias-Fauria, M., Ondo, I., Tovar, C., Wilkin, P., & Willis, K. J.  
801 (2019). Potential adaptive strategies for 29 sub-Saharan crops under future climate change. *Nature Climate*  
802 *Change*, 9(10), 758–763. <https://doi.org/10.1038/s41558-019-0585-7>

- 803 Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids  
804 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL*, *7*(1), 217–240.  
805 <https://doi.org/10.5194/soil-7-217-2021>
- 806 Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to  
807 develop accurate species distribution models. *Ecography*, *39*(6), 542–552.  
808 <https://doi.org/10.1111/ecog.01509>
- 809 QGIS Development Team. (2022). *Qgis geographic information system* (3.22.11). Open Source Geospatial Foundation  
810 Project. <http://qgis.osgeo.org>
- 811 R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing,  
812 Vienna, Austria. URL <https://www.R-project.org/>.
- 813 R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.  
814 <https://www.R-project.org/>
- 815 Ranjitkar, S., Kindt, R., Sujakhu, N. M., Hart, R., Guo, W., Yang, X., Shrestha, K. K., Xu, J., & Luedeling, E. (2014).  
816 Separation of the bioclimatic spaces of Himalayan tree rhododendron species predicted by ensemble  
817 suitability models. *Global Ecology and Conservation*, *1*, 2–12. <https://doi.org/10.1016/j.gecco.2014.07.001>
- 818 Raven, P. H., Gereau, R. E., Phillipson, P. B., Chatelain, C., Jenkins, C. N., & Ulloa Ulloa, C. (2020). The distribution of  
819 biodiversity richness in the tropics. *Science Advances*, *6*(37), eabc6228.  
820 <https://doi.org/10.1126/sciadv.abc6228>
- 821 Ripple, W. J., Wolf, C., Newsome, T. M., Galetti, M., Alamgir, M., Crist, E., Mahmoud, M. I., Laurance, W. F., & 15,364  
822 scientist signatories from 184 countries. (2017). World scientists' warning to humanity: A second notice.  
823 *BioScience*, *67*(12), 1026–1028. <https://doi.org/10.1093/biosci/bix125>
- 824 Rius, B. F., Filho, J. P. D., Fleischer, K., Hofhansl, F., Blanco, C. C., Rammig, A., Domingues, T. F., & Lapola, D. M. (2023).  
825 Higher functional diversity improves modeling of Amazon forest carbon storage. *Ecological Modelling*, *481*,  
826 110323. <https://doi.org/10.1016/j.ecolmodel.2023.110323>
- 827 Rivers, M., Newton, A. C., Oldfield, S., & Global Tree Assessment Contributors. (2022). Scientists' warning to humanity  
828 on tree extinctions. *PLANTS, PEOPLE, PLANET*, ppp3.10314. <https://doi.org/10.1002/ppp3.10314>
- 829 Rosenstock, T. S., Dawson, I. K., Aynekulu, E., Chomba, S., Degrande, A., Fornace, K., Jamnadass, R., Kimaro, A., Kindt,  
830 R., Lamanna, C., Malesu, M., Mausch, K., McMullin, S., Murage, P., Namoi, N., Njenga, M., Nyoka, I., Paez  
831 Valencia, A. M., Sola, P., ... Steward, P. (2019). A planetary health perspective on agroforestry in sub-saharan  
832 africa. *One Earth*, *1*(3), 330–344. <https://doi.org/10.1016/j.oneear.2019.10.017>
- 833 Santini, L., Benítez-López, A., Maiorano, L., Čengić, M., & Huijbregts, M. A. J. (2021). Assessing the reliability of species  
834 distribution projections in climate change research. *Diversity and Distributions*, *27*(6), 1035–1050.  
835 <https://doi.org/10.1111/ddi.13252>
- 836 Serra-Diaz, J. M., Enquist, B. J., Maitner, B., Merow, C., & Svenning, J.-C. (2018). Big data of tree species distributions:  
837 How big and how good? *Forest Ecosystems*, *4*(1), 30. <https://doi.org/10.1186/s40663-017-0120-0>
- 838 Silva De Miranda, P. L., Dexter, K. G., Swaine, M. D., De Oliveira-Filho, A. T., Hardy, O. J., & Fayolle, A. (2022).  
839 Dissecting the difference in tree species richness between Africa and South America. *Proceedings of the*  
840 *National Academy of Sciences*, *119*(14), e2112336119. <https://doi.org/10.1073/pnas.2112336119>
- 841 Tack, W., Engledow, H., Pereira, N. V., Amani, C., Bachman, S. P., Barberá, P., Beentje, H. J., Bouka, G. U. D., Cheek, M.,  
842 Cosiaux, A., Dauby, G., Block, P. D., Ewango, C. E. N., Fischer, E., Gereau, R. E., Hargreaves, S., Harvey-Brown,  
843 Y., Ikabanga, D. U., Ilunga, E. I. wa, ... Sosef, M. S. M. (2022). The ECAT dataset: Expert-validated distribution  
844 data of endemic and sub-endemic trees of Central Africa (Dem. Rep. Congo, rwanda, burundi). *PhytoKeys*,  
845 *206*, 137–151. <https://doi.org/10.3897/phytokeys.206.77379>
- 846 Title, P. O., & Bemmels, J. B. (2018). ENVIREM: An expanded set of bioclimatic and topographic variables increases  
847 flexibility and improves performance of ecological niche modeling. *Ecography*, *41*(2), 291–307.  
848 <https://doi.org/10.1111/ecog.02880>
- 849 Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160).
- 850 van Noordwijk, M., Coe, R., Sinclair, F. L., Luedeling, E., Bayala, J., Muthuri, C. W., Cooper, P., Kindt, R., Duguma, L.,  
851 Lamanna, C., & Minang, P. A. (2021). Climate change adaptation in and through agroforestry: Four decades of  
852 research initiated by Peter Huxley. *Mitigation and Adaptation Strategies for Global Change*, *26*(5), 18.

- 853 <https://doi.org/10.1007/s11027-021-09954-5>
- 854 Van Noordwijk, M., Gitz, V., Minang, P. A., Dewi, S., Leimona, B., Duguma, L., Pingault, N., & Meybeck, A. (2020).  
855 People-centric nature-based land restoration through agroforestry: A typology. *Land*, *9*(8), 251.  
856 <https://doi.org/10.3390/land9080251>
- 857 Van Zonneveld, M., Kindt, R., McMullin, S., Achigan-Dako, E. G., N'Danikou, S., Hsieh, W., Lin, Y., & Dawson, I. K.  
858 (2023). Forgotten food crops in sub-Saharan Africa for healthy diets in a changing climate. *Proceedings of the*  
859 *National Academy of Sciences*, *120*(14), e2205794120. <https://doi.org/10.1073/pnas.2205794120>
- 860 Van Zonneveld, M., Thomas, E., Castañeda-Álvarez, N. P., Van Damme, V., Alcazar, C., Loo, J., & Scheldeman, X. (2018).  
861 Tree genetic resources at risk in South America: A spatial threat assessment to prioritize populations for  
862 conservation. *Diversity and Distributions*, *24*(6), 718–729. <https://doi.org/10.1111/ddi.12724>
- 863 Varela, S., Anderson, R. P., García-Valdés, R., & Fernández-González, F. (2014). Environmental filters reduce the effects  
864 of sampling bias and improve predictions of ecological niche models. *Ecography*, no-no.  
865 <https://doi.org/10.1111/j.1600-0587.2013.00441.x>
- 866 Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer.
- 867 Williams, J. W., & Jackson, S. T. (2007). Novel climates, no-analog communities, and ecological surprises. *Frontiers in*  
868 *Ecology and the Environment*, *5*(9), 475–482. <https://doi.org/10.1890/070037>
- 869 Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & NCEAS Predicting Species Distributions  
870 Working Group†. (2008). Effects of sample size on the performance of species distribution models. *Diversity*  
871 *and Distributions*, *14*(5), 763–773. <https://doi.org/10.1111/j.1472-4642.2008.00482.x>
- 872 Zhao, L., Oleson, K., Bou-Zeid, E., Kravynhoff, E. S., Bray, A., Zhu, Q., Zheng, Z., Chen, C., & Oppenheimer, M. (2021).  
873 Global multi-model projections of local urban climates. *Nature Climate Change*, *11*(2), 152–157.  
874 <https://doi.org/10.1038/s41558-020-00958-8>
- 875 Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M.,  
876 Scharn, R., Svantesson, S., Wengström, N., Zizka, V., & Antonelli, A. (2019). COORDINATECLEANER: Standardized  
877 cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*, *10*(5),  
878 744–751. <https://doi.org/10.1111/2041-210X.13152>
- 879 Zuidema, P. A., Babst, F., Groenendijk, P., Trouet, V., Abiyu, A., Acuña-Soto, R., Adenesky-Filho, E., Alfaro-Sánchez, R.,  
880 Aragão, J. R. V., Assis-Pereira, G., Bai, X., Barbosa, A. C., Battipaglia, G., Beeckman, H., Botosso, P. C., Bradley,  
881 T., Bräuning, A., Brienen, R., Buckley, B. M., ... Zhou, Z.-K. (2022). Tropical tree growth driven by dry-season  
882 climate variability. *Nature Geoscience*, *15*(4), 269–276. <https://doi.org/10.1038/s41561-022-00911-8>
- 883