Climate legacies drive the distribution and future restoration potential of dryland forests

Article *in* Nature Plants · July 2022

DOI: 10.1038/s41477-022-01198-8

6 authors, including:

Emilio Guirado
University of Alicante
83 PUBLICATIONS **1,855** CITATIONS
SEE PROFILE

Manuel Delgado-Baquerizo
Spanish National Research Council
402 PUBLICATIONS **22,911** CITATIONS
SEE PROFILE

Jaime Martínez-Valderrama
Spanish National Research Council
89 PUBLICATIONS **1,057** CITATIONS
SEE PROFILE

Siham Tabik
University of Granada
100 PUBLICATIONS **8,682** CITATIONS
SEE PROFILE

nature plants

LETTERS

https://doi.org/10.1038/s41477-022-01198-8

Check for updates

Climate legacies drive the distribution and future restoration potential of dryland forests

Emilio Guirado[1]✉, Manuel Delgado-Baquerizo[2,3], Jaime Martínez-Valderrama[1], Siham Tabik[4], Domingo Alcaraz-Segura[5,6,7] and Fernando T. Maestre[1,8]

Knowing the extent and environmental drivers of forests is key to successfully restore degraded ecosystems, and to mitigate climate change and desertification impacts using tree planting. Water availability is the main limiting factor for the development of forests in drylands, yet the importance of groundwater resources and palaeoclimate as drivers of their current distribution has been neglected. Here we report that mid-Holocene climates and aquifer trends are key predictors of the distribution of dryland forests worldwide. We also updated the global extent of dryland forests to 1,283 million hectares and showed that failing to consider past climates and aquifers has resulted in ignoring or misplacing up to 130 million hectares of forests in drylands. Our findings highlight the importance of a wetter past and well-preserved aquifers to explain the current distribution of dryland forests, and can guide restoration actions by avoiding unsuitable areas for tree establishment in a drier world.

Dryland forests are essential for the survival of the poorest human populations on our planet, which strongly rely on woody vegetation for obtaining fuel, shelter and food[1]. Reductions in soil water availability[2] associated with forecasted increases in aridity[3] and in the number and duration of droughts[4] are expected to reduce the area of drylands capable of supporting forest ecosystems[5]. Despite this, many dryland regions, especially in China[6] and Africa[7], are important candidates for global ecosystem restoration initiatives associated with the UN Decade of Ecosystem Restoration 2021–2030[8]. Trees in drylands typically use more water than grasses or shrubs[9], hence afforestation may reduce water availability for essential human activities such as agriculture[10], resulting in water shortages that could lead to local and regional conflicts[11,12]. Because of this, identifying those drylands capable of supporting forests is essential to guide current and future restoration efforts so that they can maximize the multiple benefits of forests while minimizing the risk of water scarcity in these areas.

Despite important recent advances in understanding the distribution of dryland forests[13], their definitive extent is far from being fully understood. For example, a recent study[14] found an unexpected number of trees in the Sahel, where numerous efforts involving the restoration of complex ecosystems are underway under the umbrella of Africa's Great Green Wall[15]. We posit that current discrepancies in forest extent are based on the lack of consideration of key factors influencing the development of trees in drylands. In contrast to forests located in humid areas, the distribution dryland forests is highly constrained by their typically low water availability and high evapotranspiration rates[16]. However, not all current drylands have experienced the same dry climate over millennia, and many of them come from a wetter past[17]. Palaeoclimatic conditions are known to influence the current structure and functioning of terrestrial ecosystems[18,19] and may have influenced the establishment of dryland forests over millennia[20]. Locations with wetter past conditions might thus have allowed the establishment of dryland forests, which otherwise might not exist under today's drier climates. However, empirical evidence supporting this is lacking. Similarly, the presence of local shallow aquifers—many of those located in drylands are relics from a wetter past—influences 22–32% of the global land surface[21] and has been found to influence the distribution of forests in particular drylands[22]. Remarkably, the role of past climates and groundwater resources as predictors of the current distribution of dryland forests at the global scale is poorly understood and has not been evaluated yet.

Here we combine a unique very high-resolution (<1 m per pixel) imagery dataset of 94,352 dryland plots (0.5 ha), with information on climate[23], aquifer trends[24], soil properties[25], environmental factors, land use maps[26,27] and vegetation height[28] to: (1) quantify the relative importance of current and past (mid-Holocene; 6,000 years before the present) climate, aquifers and other key environmental predictors (Supplementary Table 1) associated with the current distribution of forests across global drylands; (2) provide an accurate and updated distribution of dryland forests worldwide; (3) compare the current extent of dryland forests with maps of tree restoration potential[29]; and (4) forecast the future (2081–2100) extent of dryland forests according to multiple socio-economic and climate change scenarios. These are fundamental steps to advance our knowledge about the extent and predictors of forest ecosystems across global drylands, which cover ~41% of the Earth's land surface[30], to maximize the socio-economic and ecological benefits of afforestation efforts, and to inform policies to mitigate climate change and desertification.

In general, mid-Holocene precipitation and temperature (Fig. 1) and climatic legacies (differences in precipitation and temperature between past and current climates) predicted a unique and significant proportion of the variation in the distribution of current forests

[1]Instituto Multidisciplinar para el Estudio del Medio 'Ramón Margalef', Universidad de Alicante, Alicante, Spain. [2]Laboratorio de Biodiversidad y Funcionamiento Ecosistémico, Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNAS), CSIC, Sevilla, Spain. [3]Unidad Asociada CSIC-UPO (BioFun), Universidad Pablo de Olavide, Sevilla, Spain. [4]Department of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, Granada, Spain. [5]iecolab. Inter-University Institute for Earth System Research, University of Granada, Granada, Spain. [6]Department of Botany, Faculty of Science, University of Granada, Granada, Spain. [7]Andalusian Center for the Assessment and Monitoring of Global Change -CAESCG-, University of Almeria, Almeria, Spain. [8]Departamento de Ecología, Universidad de Alicante, Alicante, Spain. ✉e-mail: emilio.guirado@ua.es

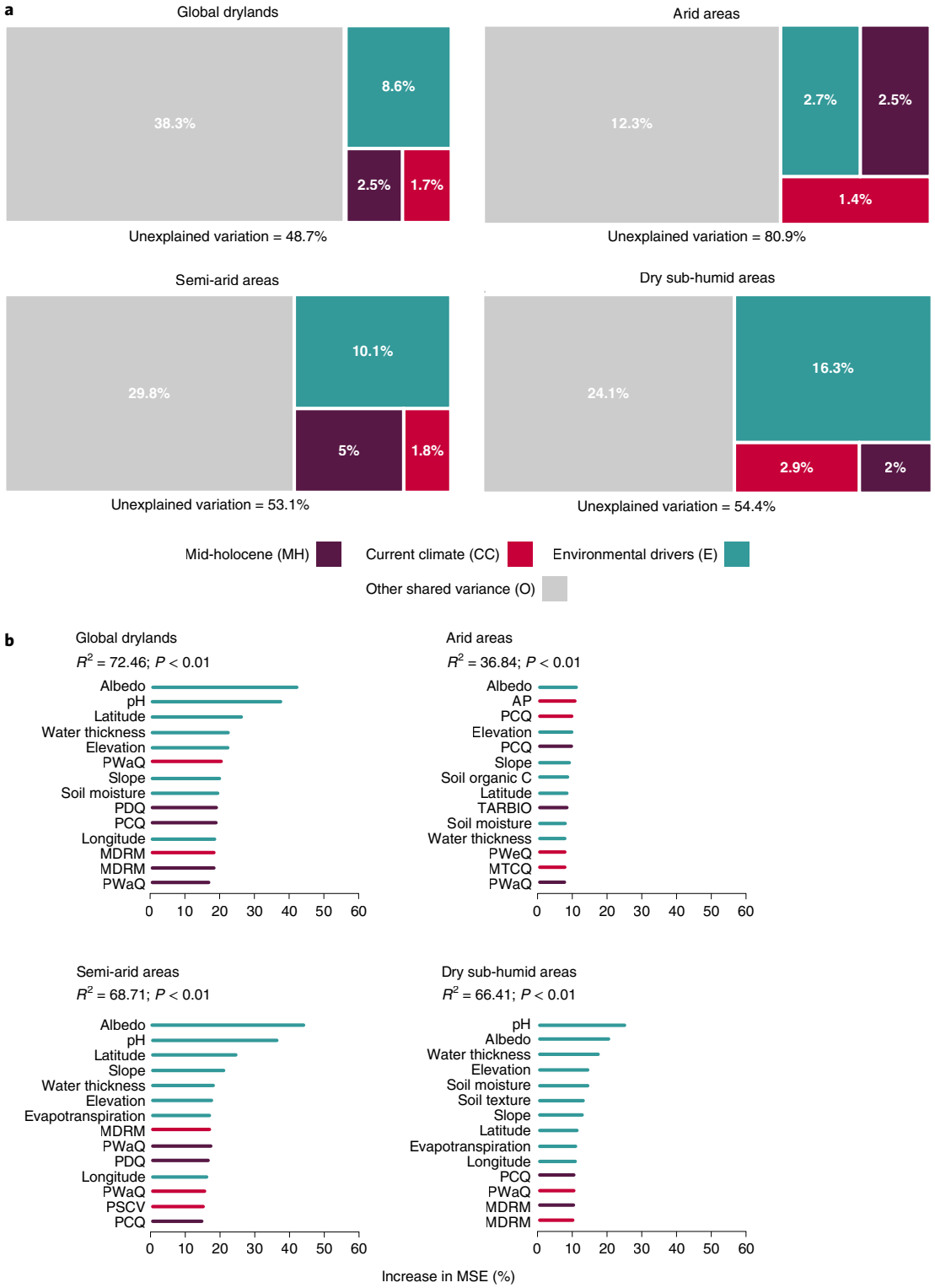


Fig. 1 | Relative contribution of past climate (mid-Holocene, 6,000 years before the present), current climate and other environmental variables as drivers of the current distribution of dryland forests worldwide. a, Results from variation partitioning modelling aiming to identify the percentage of the variance of forest distribution explained by each predictor. **b**, Results from random forest analysis aiming to identify the top 14 significant ($P < 0.01$; for one-sided testing) variables regulating forest distribution across global drylands. An increase in the percentage of mean squared error (MSE) in variables equals more importance. Acronyms are available in Supplementary Table 1.

in drylands (variation partitioning analyses; Methods and Fig. 1a). These findings were particularly evident in semi-arid regions, which had a wetter climate in the mid-Holocene than today[31] (Extended Data Fig. 1). Similar results were found when using an alternative machine-learning approach (random forest modelling; Methods) to quantify the importance of the variables studied (Fig. 1b). Our

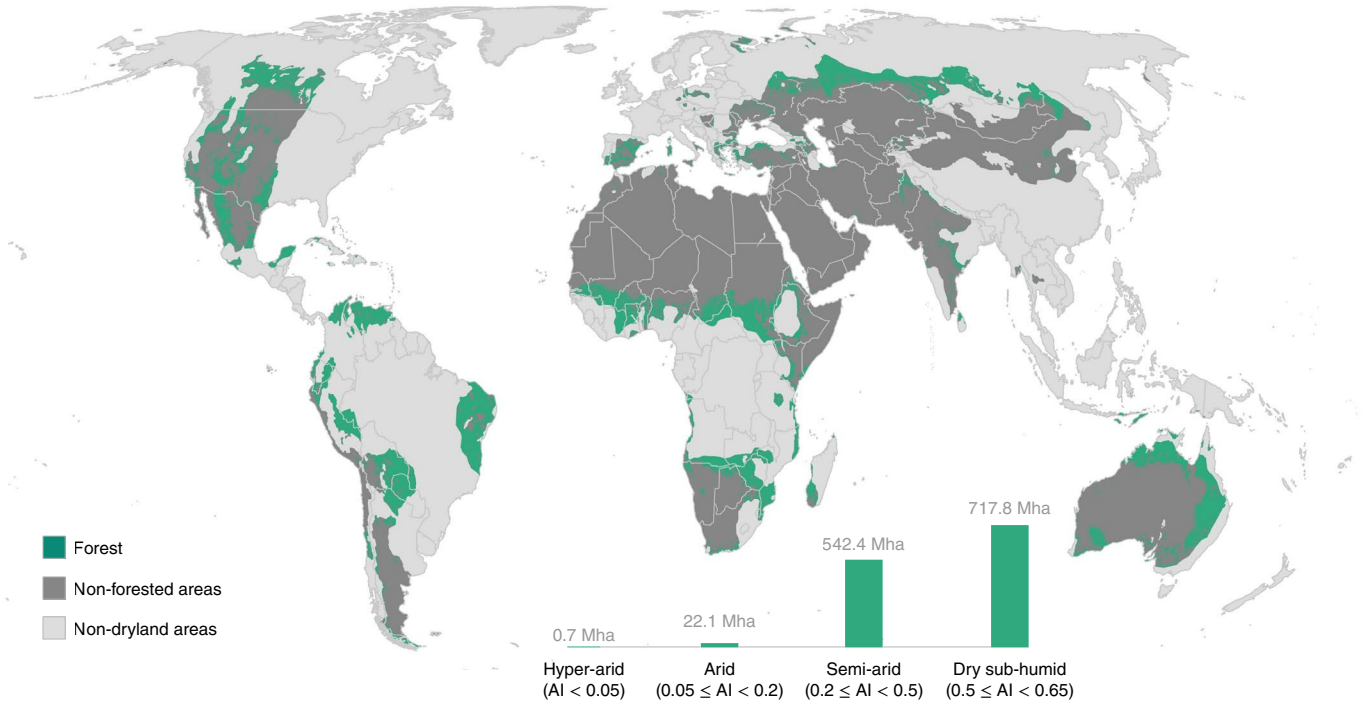


Fig. 2 | The distribution of forests across global drylands considering climate legacies (changes in climate over the past 6,000 years) and aquifer trends. Forest areas are those with a probability >50% according to the Random Forest approach used (Methods).

analyses show that locations with higher increases in precipitation since the mid-Holocene, particularly in the warmest, driest and coldest quarters, hold more forests now than should be expected according to their current climatic conditions. Moreover, areas that have suffered from larger increases in temperature over the past 6,000 years support fewer forests today. Our results thus provide new evidence on the importance of past climates as a major predictor of the current distribution of dryland forests worldwide.

The presence and evolution of aquifers (measured by water thickness over the period 2002–2017; Methods) was also a key factor explaining the current distribution of dryland forests (Fig. 1b). Almost half of the forests in drylands, 613 million hectares (Mha), are growing over aquifers where the piezometric level has declined[21] (Extended Data Fig. 2). These include forests found mainly in zones of eastern Brazil, central Canada, northern Mexico, southeastern Russia and the southwestern United States of America. These areas may thus not be able to support forests in the future, given the enhanced aridity and the increased duration and intensity of drought periods expected for many of them[5].

We found that drylands are covered by 1,283 ± 15 Mha of forests, excluding tree plantations (Fig. 2). Forests occupy 717.8, 542.4, 22.1 and 0.7 Mha of dry sub-humid, semi-arid, arid and hyper-arid areas, respectively, and represent ~19% of the surface of global drylands. The confidence of these estimates was highest in areas with a high probability of finding forest (that is, the models agreed on areas of bare soil or very dense forests, such as the Sahara Desert or the forests in South America). In contrast, areas with low confidence were those with low or medium probability (less than 50%, the threshold used to classify forest/non-forest; Extended Data Fig. 3). Compared with the most recent estimates available[13,32], our estimates increase the global forest area in drylands by ~200 Mha. Even more importantly, our forest map changed the location of ~33% of dryland forest area compared with the most recent global estimates[13,32], suggesting that 1/3 of forests were misplaced by the most recent estimates (Extended Data Fig. 4). The main reasons for these differences are probably related to the modelling approach we used, which combines the explicit consideration of past climate and groundwater trends with high-resolution imagery and discarded areas covered by large shrubs (Methods).

To further illustrate the importance of past climates and aquifers as drivers of the current distribution of forests in drylands, we located the extent of forests in drylands with or without consideration of past climate and aquifer trends. The consensus between the forest maps considering or excluding climatic legacies and aquifer trends was 1,225 Mha (Fig. 3). The discrepancies in these maps added up to ~130 Mha, equivalent to 3.7 billion trees[14] or the equivalent of the total area of France, Italy, the United Kingdom, Switzerland, the Netherlands and Belgium together. These discrepancies were mainly located in areas where palaeoclimatic legacies were especially important, which include the Sahel, south Australia, Mexico and the southwestern United States (Fig. 3). The presence of many of these forests cannot simply be explained without considering palaeoclimates and aquifer trends. The importance of past climates as drivers of the distribution of current forests was more noticeable in semi-arid regions across the globe (Fig. 3), with over 81 Mha of forests being neglected or misplaced by models based only on current climatic conditions. Over 31 Mha of these forests are moreover located in the transition between arid and semi-arid regions (aridity index (AI) = 0.2; AI, precipitation/potential evapotranspiration), which has recently been reported to be a threshold driving abrupt changes[33] in multiple ecosystem attributes, including a decline in vegetation cover and richness, across global drylands[34]. Planting trees in these transitional regions should be done with extreme caution and even avoided where it could accelerate the depletion of groundwater resources and jeopardize water availability for other organisms and human uses, as is already occurring in areas where extensive afforestation programmes have been established over the past decades, such as in the Chinese Loess Plateau[10].

The existence of dryland forests may be explained by different causes. They can be relict forests where recruiting is being hampered by current conditions, but their survival is not. If so, the trees in these areas would be old and simply a very long transient state

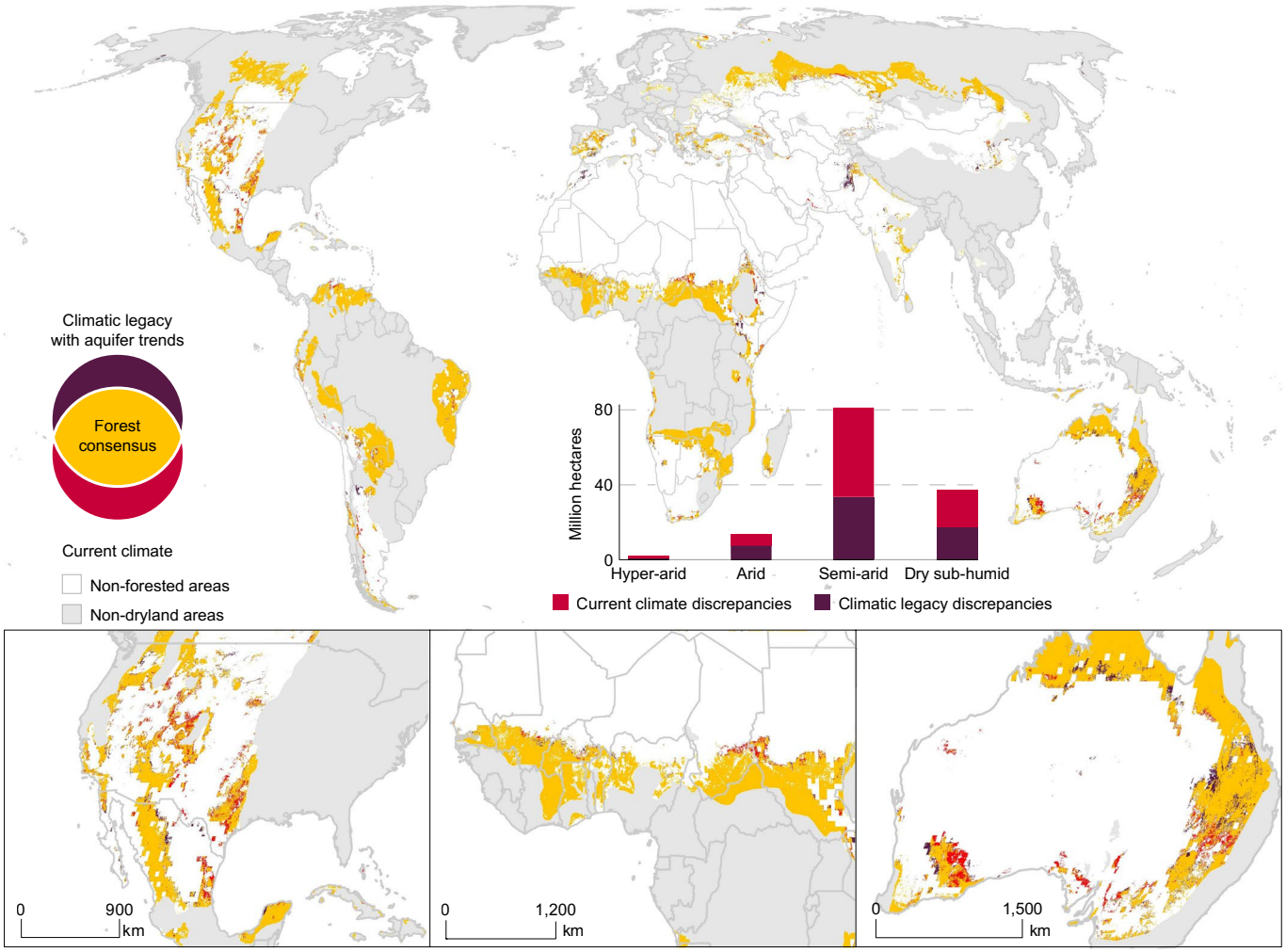


Fig. 3 | Comparison between the extent of forests calculated with and without considering the effects of climatic legacies and aquifer trends. Consensus areas are shown in yellow. Discrepancies in forest extent when considering climatic legacies (changes in climate over the past 6,000 years) and aquifer trends are shown in purple. Discrepancies in forest extent when considering current climate only are shown in red.

towards conversion into a non-forested vegetation[35]. However, since the trees in these areas would be at least 6,000 years old and these ages are uncommon for many tree species, this cause can probably be overruled. A more plausible explanation is that forests in these areas are stable configurations that can survive and recruit under current conditions, even if they do so at the edge of their optimal environmental suitability. To check whether trees existed in the mid-Holocene in the locations where forests exist today, we explored data from LegacyPollen v.1[36], a harmonization of palaeo-pollen databases. We found pollen samples (n = 1,121) of tree species aged 5,000–7,000 years at 119 sites located in current dryland forests from six large regions (Africa, Asia, eastern North America, Europe, South America and western North America; Extended Data Fig. 5a). The median of the percentage of pollen from tree species was more than 40% in these regions (Extended Data Fig. 5b), and samples with more than 5% tree species pollen were found in 93% of the cases (Asia 100%, Europe 100%, North America 100%, Africa 67% and South America 96%). We observed a significant negative relationship between the percentage of tree species present in the pollen samples and aridity across Europe, eastern North America and western North America (Extended Data Fig. 5c). These findings suggest that most dryland areas sustaining forests today already had a prevalence of forest vegetation in the mid-Holocene and that the dominance of tree species decreased as aridity increased back then (as it happens today). Although these findings cannot provide a definitive proof, they suggest the presence of hysteretic behaviour in dryland forests (that is, forests inherited from the past do not revert to non-forest states even when environmental conditions suggest that this would be the most reasonable). This type of hysteretic behaviour has already been found in tropical forests[37]. If true for dryland forests, this would suggest the existence of some type of forest-specific stabilizing mechanism that allows forests to thrive in drylands. Potential stabilizing mechanisms include hydroclimatic feedbacks[37] and the modification of the surrounding physical environment by trees[38], which create suitable conditions for tree development.

By identifying climatic legacies and aquifers as important factors explaining the current distribution of forests across global drylands, our findings have important implications for their restoration. The comparison of the current forest extent in drylands, as revealed by our results, with tree restoration potential maps provided by Bastin et al.[29] results in a consensus of 464 Mha of dryland forests with suitable conditions for tree planting (Fig. 4). These areas include, for example, the state of Texas (USA), northern Argentina, Paraguay and Bolivia, areas of central and western Sahel, north-eastern Australia, southern Angola, northern Namibia, Botswana, Zimbabwe, Mozambique and southern Madagascar. In contrast, 819 Mha of dryland forests revealed by our results are not included in

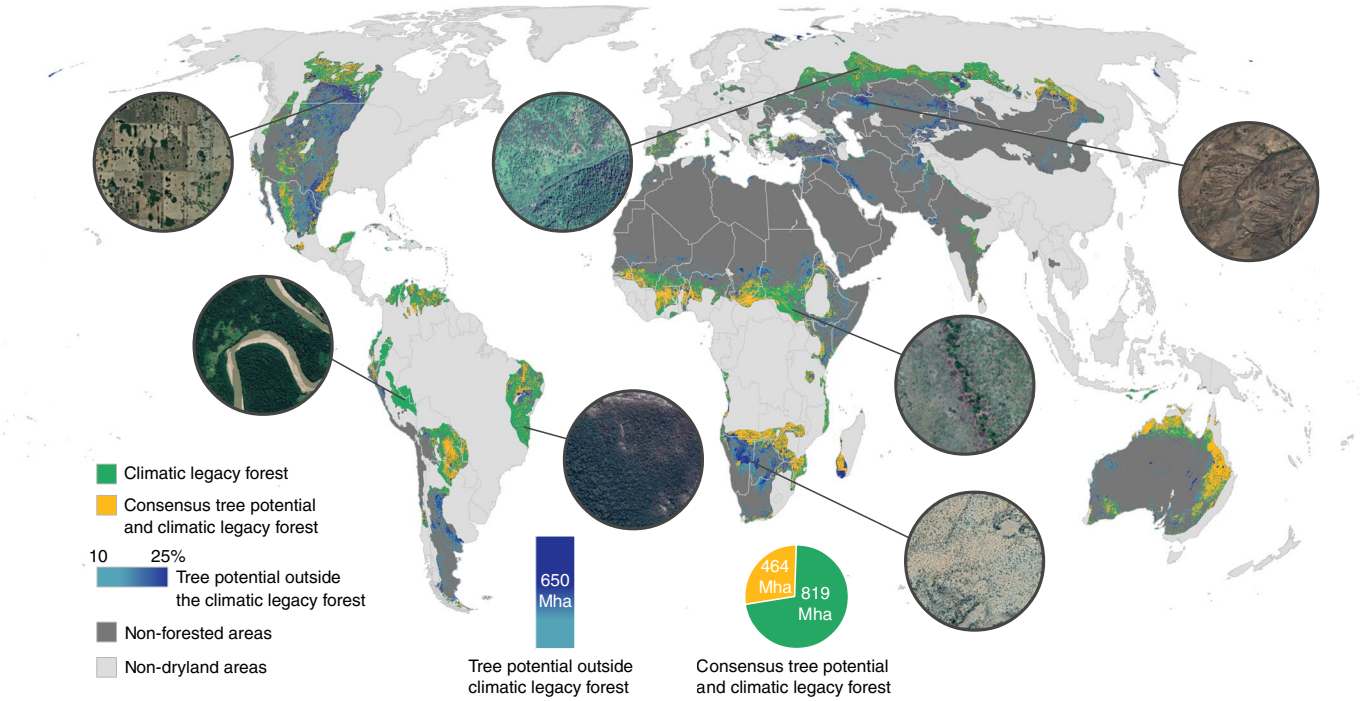


Fig. 4 | Comparison between the calculated forest extent considering climate legacies and aquifers and the potential areas for restoring trees. The comparison is based on results shown in Fig. 2 and tree restoration potential identified by Bastin et al.[29]. Forest areas matching those with tree restoration potential are in yellow. Areas with a blue gradient show the potential for tree cover restoration in non-forested areas (between 10% and 25%). Areas with no restoration potential and already containing forests are in green. Examples of areas with different restoration potential are shown in the inset circles. Map data: Google, Maxar Technologies.

the tree restoration potential maps. This is reasonable, as in most of these areas the capacity to accommodate more trees is very limited[29]. These include areas in northwest Mexico, Venezuela, Peru, Bolivia, Paraguay, eastern Brazil, southern Russia and Sudan. In addition, our results revealed 650 Mha of non-forest areas with potential for tree restoration[29]. These areas, which include southern Canada, northeastern Mexico, central Namibia, Argentina and Kazakhstan, need to be carefully studied locally to verify whether they are forestless due to unfavourable environmental conditions for tree development (hence not suitable for afforestation programmes) or because of deforestation (hence suitable for restoration using trees).

In addition to considering where forests are found today, restoration actions involving tree planting must also consider future climatic conditions, as they will largely determine suitable areas for tree development in drylands, and thus suitable areas for these actions[39]. To gain insights into the potential future distribution of dryland forests, we rerun our model for forest extent considering both climatic legacies and projected climate data[23] from the MIROC6 global climate model[40]. We used a combination of three shared socio-economic pathways (SSP) and representative concentration pathways (RCP) scenarios (1–2.6, 3–7.0 and 5–8.5) for these forecasts (Methods). To obtain a more realistic forecast, we also used projections of the extent of crop and urban land uses in the same three scenarios using the MIROC global climate model[27] (Methods). Our results indicate that about 11% of the current extent of dryland forests (~180 Mha) will be lost in the period 2081–2100 under the SSP5-8.5 projection (Extended Data Fig. 6). At the same time, a total of 309 Mha (20% of the current extent of dryland forest) of new forests will appear in places not holding forests now (Extended Data Fig. 6). Comparing current and potential future forest extent, the areas most affected by forest loss are in eastern Brazil, central Canada and eastern Australia. The areas projected to have the greatest potential for new forests are in Africa.

In summary, our work provides novel evidence that past climates and aquifer trends are fundamental in explaining the current distribution of forests in global drylands and highlights those dryland regions that could foresee future losses and gains of forests under realistic socio-economic and climatic scenarios. The updated estimations of the current and future extension and location of dryland forests provided can be used to improve the management and conservation of forests across drylands worldwide. Our results can also help to prioritize target areas for the establishment of forests in drylands, and to select alternative species (for example, grasses or shrubs) in areas where future climatic conditions and/or depleted aquifers may not allow the establishment of trees.

Methods

Definition of forest. We followed the definition of the Food and Agriculture Organization of the United Nations (FAO), which considers forests as ecosystems with >10% tree cover, >5 m in height and over 0.5 ha located in areas not predominantly used for agriculture or urban land use[41].

Datasets for climatic and other environmental drivers. To delimit the extent of global drylands, we followed the United Nations Convention to Combat Desertification and the Convention on Biological Diversity[42], which defined drylands as those areas with AI < 0.65. According to this definition, drylands can be divided into four areas: hyper-arid (AI < 0.05), arid (0.05 ≤ AI < 0.2), semi-arid (0.2 ≤ AI < 0.5) and dry sub-humid (0.5 ≤ AI < 0.65). Here we used the aridity zone map created by the United Nations Environment Programme (UNEP)[43].

We used a total of 19 standardized climatic variables (Supplementary Table 1), which were obtained for all the sites surveyed from Worldclim Global Climate Data[23]. We used data at a 2.5 min resolution (~4.5 km at the equator) because they are available for both the current and past climate of the mid-Holocene (~6,000 years ago) so that comparisons of bioclimatic data at different periods can be done. We focused on mid-Holocene (that is, 6,000 years before the present) climates because many current drylands were wetter during that period[31]. The overall accuracy of the Worldclim climatic models was very high for temperature variables[23]. These variables had an overall correlation coefficient (between estimated and observed values) of ≥0.99 and an average root-mean-square error between 1.1 °C and 1.4 °C. Precipitation in drylands can be highly variable in time

and space, with some regions showing abrupt changes across spatial scales. In general, the prediction error increased with station elevation and distance to the nearest neighbouring station (in the training set) for all variables. Generalized additive models of cross-validation errors showed that higher elevations tended to be associated with lower interpolation accuracy, even after accounting for the effects of isolation and spatial variation in errors, although this effect differed between variables.

We also gathered soil moisture data from TerraClimate[44], which is composed of data from Worldclim[23] and the Japanese 55-year Reanalysis (JRA55)[45]; soil organic carbon stock, texture (sand content) and pH from Soilgrids[25]; albedo from MODIS/Terra MCD43A3 Version 6 at 500 m pixel resolution[46]; elevation and slope from the Advanced Land Observation Satellite (ALOS)[47]; evapotranspiration from MODIS/Terra MOD16A2 Version 6 at 500 m pixel resolution[48]; and equivalent liquid water thickness of aquifers by measuring monthly changes in gravity from the Gravity Recovery and Climate Experiment (GRACE)[49].

Automatic identification of forest plots. To avoid the low accuracy of classifications using coarse-resolution images[13,14] and the subjectivity of using multiple human operators[50], we gathered a new high-resolution image dataset and used an automatic classification system to distinguish between forest and non-forest plots on the basis of the Convolutional Neural Network (CNN) model, a type of artificial intelligence method inspired by the human brain[51]. This model can minimize the risk of uncertainty in results but not eliminate it, as the deep learning-based model has been trained with user-provided labels expert in identifying forests in satellite images but is not completely infallible.

We first compiled an updated, globally consistent and accurate dataset of precise locations of forest and non-forest plots. On the basis of the valuable information from the 213,795 0.5 ha plots provided by the Global Drylands Assessment[13], where 239 operators and FAO staff participated in the task, we selected 94,352 plots where very high-resolution Google Earth images (less than 1 m per pixel, that is, eye altitude of ~150 m and zoom level 19) were available between 1 and 13 December 2017. Then, the images of these selected plots were automatically classified into forest and non-forest using the CNN-based model.

To train the CNN-based model to differentiate forest from non-forest images, we built a new auxiliary training dataset by regrouping the 45 categories of the NWPU-RESISC45 benchmark database[52] (for example, farmland, forest, mountain, beach, island, lake, river, airport, bridge, church, chaparral and ship) into two classes: forest and non-forest. Our forest class was obtained by grouping together two categories of the NWPU-RESISC45 database: forest (mainly close forest images) and chaparral (mainly open forest images). To fulfill FAO's definition of forest in these classes, we manually removed all images with less than 10% of canopy cover and those including some portion of evident human activity (for example, infrastructures, tree and non-tree croplands, urban settlements, urban forests and so on). As a result, our new forest class contained 681 images that fulfilled FAO's criteria, and our non-forest class contained 30,100 images from all remaining 43 categories of the NWPU-RESISC45 database (for example, 700 images per category; Supplementary Data 1). We used the Inception v.3 architecture[53], one of the most accurate CNN models being used nowadays, and two optimization techniques: (1) data augmentation and (2) transfer learning. Data augmentation involves artificially increasing the number of independent samples in the training dataset by applying specific transformations to the images (for example, flipping 180°, margin cropping 10%, scaling up the size of images by 10%, brightening pixel level up to 50% and darkening pixel level up to 50%). Transfer learning involves using the knowledge acquired from a previous problem to solve a new problem. Instead of starting the learning from scratch, with transfer learning, a pre-trained model is selected and re-trained on the new problem. We used a pre-trained CNN-based model using the ImageNet database[54], which contains 1,000 image categories including fauna, flora, artificial elements and other features, with a learning rate of 0.001 and a decay factor of 16 every 30 epochs. As an optimization algorithm, we used RMSProp with momentum and decay of 0.9, and epsilon of 0.1.

Once the CNN-based model was trained, we used it to classify the 94,352 0.5 ha plots described above as forest/non-forest. The CNN-based model analyses each image and outputs two probability values (with their respective confidence intervals), one for the forest class and the other for the non-forest class. To ensure the highest accuracy of the classification (assessed as described below), we removed all plots that were classified by the CNN model with a probability lower than 99%, as well as all plots that had a vegetation height lower than 5 m in the Global Vegetation Height dataset[28]. By doing so, plots with large shrubs would not be counted as forests. Hence, the final dataset used for further analyses contained 16,739 plots with a wide representation along the global drylands (hyper-arid $n=702$, arid $n=2{,}883$, semi-arid $n=8{,}166$ and dry sub-humid areas $n=4{,}988$) (Supplementary Data 2 and Fig. 1).

Assessing the accuracy of the CNN-based model. To assess the accuracy of the CNN-based model classification of the 94,352 0.5 ha plots of our dataset into forests and non-forests, we randomly selected 705 of these plots and manually checked whether the model properly classified them using Google Earth. These validation plots were selected in a stratified way across the global drylands considering: (1) four aridity levels: hyper-arid ($AI<0.05$), arid ($0.05\leq AI<0.2$), semi-arid ($0.2\leq AI<0.5$) and dry sub-humid ($0.5\leq AI<0.65$); (2) four tree cover levels[55]: non-forest (<10%), open forest (10–40%), closed forest (41–65%) and dense forest (66–100%); and (3) 12 regions[13]: Northern Africa, Horn of Africa, the Sahel, Southern Africa, North America, South America (east), South America (west), Central Asia, Southwest Asia, Europe and Russia, Middle East, and Oceania (Supplementary Table 2). For each combination of aridity level, cover level and region, we randomly selected five images so that the potential number of images would be $4\times4\times12\times5=960$; however, some strata combinations did not exist. Thus, the total number of images used was 705. The accuracy metrics used were Precision, Recall and the *F*1-measure (Supplementary Table 3). The *F*1-measure is an overall score that considers both Precision and Recall, and is preferable to simpler methods (*F*1-measure values close to 1 are better).

Assessing the drivers of current forest distribution. We used variation partitioning[56] to quantify the relative importance of bioclimatic variables at different periods and environmental drivers (Supplementary Table 4) as predictors of the 16,739 forest/non-forest plots previously classified using the CNN model. This method is specifically recommended for dealing with multicollinearity because it partitions the variance in a given response variable that is attributed to a particular group of predictors from that variance shared among all predictors. In particular, this analysis provides insights into whether climatic variables from current and mid-Holocene periods can explain a unique portion of the variance that is not explained by climate in other periods[56]. Variation partitioning analyses were conducted with the R package 'Vegan'[57].

Importance of the drivers of current forest distribution. We conducted a random forest permutation analysis[58] to identify the main predictors of the 16,739 forest/non-forest plots previously classified using the CNN model. Contrary to the variation partitioning model described above, random forest analysis allowed us to identify the most important drivers of forest distribution among 19 bioclimatic variables[23] from the different climatic periods studied and other environmental drivers. This method is a novel machine-learning algorithm that extends standard classification and regression tree methods by creating a collection of classification trees with binary divisions. The importance of each predictor variable is determined by evaluating the decrease in prediction accuracy when the data for that predictor are randomly permuted. This decrease is averaged over all trees to produce the final measure of importance. This accuracy importance measure was computed for each tree and averaged over the forest (999 trees). Unlike multimodel inference using linear regressions or regression tree analyses, random forest analysis alleviates multicollinearity problems in multivariate analyses by building bagged tree ensembles and including a random subset of features for each tree (999 trees).

Predicting the extent and distribution of dryland forests. To quantify the global extent and current distribution of dryland forests, we used a random forest regression analysis[59] and coupled information on past climates and aquifer trends with other key environmental predictors (for example, albedo, pH, water thickness, elevation, precipitation of warmest quarter in the current climate, slope, soil moisture, precipitation of driest quarter in the mid-Holocene, precipitation of coldest quarter in the mid-Holocene, mean diurnal range in the current climate; Supplementary Tables 1 and 4) that were most important in the permutation analysis[58], identifying the main predictors of dryland forest at the 16,739 locations identified by the CNN-based model. This model was built by finding the set of covariate combinations that most robustly predict the training samples and 999 trees. The quality of the classification was tested and validated using a *k*-fold cross-validation method[60], where *k* models ($k=5$) were trained from *k* subsets of the original data and tested on *k* subsets of the remaining independent data (total number of plots divided by *k*). By combining the *k* iterations, we compared the original full dataset with the full set of the remaining independent data. The modelling approach was then validated by returning the predicted values (*x* axis) vs the observed values (*y* axis), following ref. [61]. The model had a high predictive power ($R^2=0.89$; Supplementary Fig. 2) and the validation of the *k*-fold cross revealed that our model explained 71% of the variation in forest extent without bias.

To obtain the extent of forests across global drylands, we calculated a map of forest/non-forest areas considering forest as those areas with a probability of being a forest of >50% (as provided by the random forest regression analysis). To provide realistic numbers, we eliminated all areas with croplands in at least 60% of the surface and small-scale cultivation mosaics, as well as urban and built-up lands with at least 30% of the surface being impervious (including buildings and asphalt) as identified in the global Land Cover Type by Annual International Geosphere-Biosphere Programme (IGBP)[26] classification from MOD12Q1 v.6 by MODIS/Terra satellite sensor[62].

Future projections of forest extent in drylands. To calculate future estimates of dryland forest extent globally, we reran our original model, keeping the variables elevation, slope, precipitation of driest quarter in mid-Holocene and precipitation of coldest quarter in mid-Holocene while updating the bioclimatic variables precipitation of warmest quarter and mean diurnal range using the estimates

provided by the MIROC6 global climate model[63]. To generate more reasonable future scenarios[63], we chose the combination of three SSP and RCP scenarios: 1–2.6, 3–7.0 and 5–8.5 from the Coupled Model Intercomparison Project Phase 5 (CMIP5) over the period 2081–2100. In addition, and to obtain a more realistic extent of dryland forests, we forecasted the extent of agricultural and urban land uses in the same three scenarios using estimations from the MIROC global climate model[27]. SSP1 is a sustainability scenario with low population growth, high economic growth, high levels of education, good governance, a globalized society, international cooperation, technological development and environmental awareness. Under these assumptions, this scenario represents low levels of mitigation and adaptation challenges. SSP3 is a fragmentation scenario with high population growth and low economic development, lower levels of education and a regionalized society with low environmental awareness, thus representing a high level of adaptation and mitigation challenges. SSP5 scenario assumes a very high dependence on fossil fuels, low population growth, high economic growth and high human development, thus representing a high level of mitigation challenge. The RCP2.6, RCP7.0 and RCP8.5 scenarios consider lower, medium and high greenhouse gas emission rates (see ref. [64] for details).

Uncertainty map. To represent the uncertainty of our predictions of dryland forest extent (Fig. 2), we used the standard deviation of the predictions obtained in the 5-fold cross-validation[60] (see 'Predicting the extent and distribution of dryland forests' above). By stacking the predictions of forest extent, we obtained the mean of the probabilities to determine the extent of dryland forest and its standard deviation among the five predictive models used. In summary, the standard deviation per pixel represents the confidence of the model in space (Extended Data Fig. 3). The higher values of this metric, the higher the uncertainty and vice versa.

Consensus and discrepancies between forests areas. We used a simple approach to quantify and locate the consensus between forest maps created by considering climate legacies and aquifer trends and the forest map created by considering only the current climate. The forest consensus is calculated by area A_i, which identifies the i-th pixels of the forest map considering climate legacies and aquifer trends (Forest A), and area B_j, which identifies the j-th pixels of the forest map without considering climate legacies and aquifer trends (Forest B). Then we have the sets of forest areas $A = (a_i: i = 1, 2, \ldots, m)$ and $B = (b_j: j = 1, 2, \ldots, n)$. Here, the subscripts i and j are sequential numbers for the pixels of Forest A and Forest B, respectively. m and n indicate the total numbers of the pixels of both forest maps. Finally, the corresponding sets of areas of forest A and forest B intersect, obtaining a consensus C between Forest A and Forest B (equation 1).

$$C_{ij} = \mathrm{area}A_i \cap \mathrm{area}B_j \quad (1)$$

where C_{ij} is the area intersected between the area of map Forest A ($\mathrm{area}A_i$) and the area of map Forest B ($\mathrm{area}B_j$).

This approach generates three types of areas—the intersection of both (Forest A and B) and the exclusive area for each forest map (Fig. 3).

Palaeo-pollen analysis. To check whether trees existed in the mid-Holocene in the locations where dryland forests exist today, we explored data from those sites that are forests today included in the LegacyPollen v.1[36] database, a harmonization of palaeo-pollen databases including a total of 1,002 harmonized taxon names. It integrates the Neotoma palaeoecology database (https://www.neotomadb.org/), with additional data[65,66]. Age data were obtained according to the newly established LegacyAge 1.0 framework[67] that includes the mid-Holocene period (5,000–7,000 BP). First, we identified all locations from the LegacyPollen v.1 database that are in areas of current dryland forests. Second, from this subset of dryland locations we identified those pollen samples coming from trees using the GlobalTreeSearch v.1.5 database[68], which contains the names of 60,000 tree species. Finally, we summed the percentages of each of the tree pollen samples ($n = 1{,}121$ at 119 sites) and calculated how many samples have 5% or more tree pollen in the mid-Holocene period (Extended Data Fig. 5).

Groundwater balance in forest areas. To identify the extent of forest areas growing over declining aquifers, we calculated the balance (2002–2017) of the accumulated annual equivalent water height (water thickness expressed in cm yr^{-1}) estimated from GRACE[49], which has been successfully used to monitor the evolution of the piezometric level of large aquifers on the basis of microgravimetric differences[24] (Extended Data Fig. 2). The result of summing monthly values from 2002 to 2017 was classified into three categories: declining (aquifers showing a decrease in their piezometric levels), stable equilibrium (aquifers with unchanged piezometric levels) and increasing (aquifers showing an increase in their piezometric levels).

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

Data availability

All data generated or analysed during this study, which support the maps within this paper and other findings of this study, are available from Figshare at https://doi.org/10.6084/m9.figshare.13635212.

Code availability

The CNN-based code for the classification of forest/non-forest described in the Methods is freely available at https://github.com/EGuirado/CNN-Forest-Drylands

Received: 15 March 2021; Accepted: 15 June 2022;
Published online: 25 July 2022

References

1. Middleton, N., Stringer, L., Goudie, A., & Thomas, D. *The Forgotten Billion: MDG Achievement in the Drylands* (UNDP United Nations Convention to Combat Desertification, 2011).
2. Soong, J. L., Phillips, C. L., Ledna, C., Koven, C. D. & Torn, M. S. CMIP5 models predict rapid and deep soil warming over the 21st century. *J. Geophys. Res.* **125**, e2019JG005266 (2020).
3. Huang, J., Yu, H., Guan, X., Wang, G. & Guo, R. Accelerated dryland expansion under climate change. *Nat. Clim. Change* **6**, 166–171 (2016).
4. Williams, A. P. et al. Large contribution from anthropogenic warming to an emerging North American megadrought. *Science* **368**, 314–318 (2020).
5. Schlaepfer, D. et al. Climate change reduces extent of temperate drylands and intensifies drought in deep soils. *Nat. Commun.* **8**, 14196 (2017).
6. Jiang, H. in *The End of Desertification?* (eds Behnke, R. & Mortimore, M.) 513–536 (Springer, 2016).
7. Gadzama, N. M. Attenuation of the effects of desertification through sustainable development of Great Green Wall in the Sahel of Africa. *World J. Sci. Technol. Sustain. Dev.* **14**, 279–289 (2017).
8. *United Nations Decade on Restoration* (accessed January 2021); https://www.decadeonrestoration.org/
9. Ellison, D. et al. Trees, forests and water: cool insights for a hot world. *Glob. Environ. Change* **43**, 51–61 (2017).
10. Feng, X. et al. Revegetation in China's Loess Plateau is approaching sustainable water resource limits. *Nat. Clim. Change* **6**, 1019–1022 (2016).
11. Megdal, S. B. Transboundary groundwater resources: sustainable management and conflict resolution. *Groundwater* **55**, 701–702 (2017).
12. Jarvis, W.T. in *Advances in Groundwater Governance* (eds Villholth, K. G. et al.) 177–192 (CRC Press, 2017).
13. Bastin, J.-F. et al. The extent of forest in dryland biomes. *Science* **356**, 635–638 (2017).
14. Brandt, M. et al. An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature* **587**, 78–82 (2020).
15. Mbow, C. The Great Green Wall in the Sahel. *Oxf. Res. Encycl. Clim. Sci.* https://doi.org/10.1093/acrefore/9780190228620.013.559 (2017).
16. Petrie, M. D. et al. Climate change may restrict dryland forest regeneration in the 21st century. *Ecology* **98**, 1548–1559 (2017).
17. Liu, S., Jiang, D. & Lang, X. Mid-Holocene drylands: a multi-model analysis using Paleoclimate Modelling Intercomparison Project Phase III (PMIP3) simulations. *Holocene* **29**, 1425–1438 (2019).
18. Delgado-Baquerizo, M. et al. Palaeoclimate explains a unique proportion of the global variation in soil bacterial communities. *Nat. Ecol. Evol.* **1**, 1339–1347 (2017).
19. Delgado-Baquerizo, M. et al. Effects of climate legacies on above- and belowground community assembly. *Glob. Change Biol.* **24**, 4330–4339 (2018).
20. Hoelzmann, P. et al. Mid-Holocene land-surface conditions in northern Africa and the Arabian Peninsula: a data set for the analysis of biogeophysical feedbacks in the climate system. *Glob. Biogeochem. Cycles* **12**, 35–51 (1998).
21. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. *Science* **339**, 940–943 (2013).
22. Smettem, K. R. J., Waring, R. H., Callow, J. N., Wilson, M. & Mu, Q. Satellite-derived estimates of forest leaf area index in southwest Western Australia are not tightly coupled to interannual variations in rainfall: implications for groundwater decline in a drying climate. *Glob. Change Biol.* **19**, 2401–2412 (2013).
23. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
24. Schmidt, R. et al. GRACE observations of changes in continental water storage. *Glob. Planet. Change* **50**, 112–126 (2006).
25. Hengl, T. et al. SoilGrids250m: global gridded soil information based on machine learning. *PLoS ONE* **12**, e0169748 (2017).
26. Friedl, M. A. et al. *ISLSCP II MODIS (Collection 4) IGBP Land Cover, 2000–2001* (ORNL DAAC, Oak Ridge, TN, USA, 2010); https://doi.org/10.3334/ORNLDAAC/968
27. Chen, M. et al. Global land use for 2015–2100 at 0.05° resolution under diverse socioeconomic and climate scenarios. *Sci. Data* **7**, 320 (2020).
28. National Centre for Earth Observation & Los, S.O. *Global Vegetation Height Frequency Distributions from the ICESAT GLAS instrument produced as part of the National Centre for Earth Observation (NCEO)* (NERC Earth Observation Data Centre, accessed 10 December 2020); http://catalogue.ceda.ac.uk/uuid/85e7d70a74244c73b71446940e05cde6

29. Bastin, J.-F. et al. The global tree restoration potential. *Science* **365**, 76–79 (2019).
30. Cherlet, M. et al. *World Atlas of Desertification: Rethinking Land Degradation and Sustainable Land Management* (Publications Office of the European Union, 2018).
31. Ganopolski, A., Kubatzki, C., Claussen, M., Brovkin, V. & Petoukhov, V. The influence of vegetation-atmosphere-ocean interaction on climate during the mid-holocene. *Science* **280**, 1916–1919 (1998).
32. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
33. Scheffer, M. *Tipping Points* (Princeton Univ. Press, 2009).
34. Berdugo, M. et al. Global ecosystem thresholds driven by aridity. *Science* **367**, 787–790 (2020).
35. Runyan, C. W. & D'Odorico, P. *Global Deforestation* (Cambridge Univ. Press, 2016).
36. Herzschuh, U. et al. Global taxonomically harmonized pollen data set for Late Quaternary with revised chronologies (LegacyPollen 1.0). *PANGAEA* https://doi.org/10.1594/PANGAEA.929773 (2021).
37. Staal, A. et al. Hysteresis of tropical forests in the 21st century. *Nat. Commun.* **11**, 4978 (2020).
38. Belsky, A. J. et al. The effects of trees on their physical, chemical and biological environments in a semi-arid savanna in Kenya. *J. Appl. Ecol.* **26**, 1005–1024 (1989).
39. Li, C. et al. Drivers and impacts of changes in China's drylands. *Nat. Rev. Earth Environ.* **2**, 858–873 (2021).
40. Tatebe, H. et al. Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geosci. Model Dev.* **12**, 2727–2765 (2019).
41. *Trees, Forests and Land Use in Drylands: the First Global Assessment. Full Report* (FAO, 2019).
42. Diallo, H. A. in *The Future of Drylands* (eds Lee, C. & Schaaf, T.) 13–16 (Springer, 2008).
43. *A Spatial Analysis Approach to the Global Delineation of Dryland Areas of Relevance to the CBD Programme of Work on Dry and Subhumid Lands* (UNEP-WCMC, 2014).
44. Abatzoglou, J. et al. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **5**, 170191 (2018).
45. Tachikawa, T., Hato, M., Kaku, M. & Iwasaki, A. Characteristics of ASTER GDEM version 2. *IEEE Int. Geosci. Remote Sens. Symp. Proc.* https://doi.org/10.1109/igarss.2011.6050017 (2011).
46. Alibakhshi, S., Crowther, T. W. & Naimi, B. Land surface black-sky albedo at a fixed solar zenith angle and its relation to forest structure during peak growing season based on remote sensing data. *Data Brief.* **31**, 105720 (2020).
47. Hamazaki, T. Advanced land observation satellite (ALOS). 5 Outline of ALOS satellite system. *J. Jpn Soc. Photogramm. Remote Sens.* **38**, 25–26 (1999).
48. Mu, Q., Zhao, M., & Running, S. W. Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sens. Environ.* **115**, 1781–1800 (2011).https://doi.org/10.1016/j.rse.2011.02.019
49. Zlotnicki, V., Bettadpur, S., Landerer, F. W. & Watkins, M. M. in *Encyclopedia of Sustainability Science and Technology* (ed. Meyers, R. A.) 4563–4584 (Springer, 2012).https://doi.org/10.1007/978-1-4419-0851-3_745
50. Schepaschenko, D. et al. Comment on 'The extent of forest in dryland biomes'. *Science* **358**, 6362 (2017).
51. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
52. Cheng, G., Han, J. & Lu, X. Remote sensing image scene classification: benchmark and state of the art. *Proc. IEEE* **105**, 1865–1883 (2017).
53. Xia, X., Xu, C. & Nan, B. Inception-v3 for flower classification. In *Proc. 2nd International Conference on Image, Vision and Computing (ICIVC)* 783–787 (IEEE, 2017).
54. Fei-Fei, L., Deng, J. & Li, K. ImageNet: constructing a large-scale image database. *J. Vis.* **9**, 1037 (2010).
55. Guirado, E. et al. Tree cover estimation in global drylands from space using deep learning. *Remote Sens.* **12**, 343 (2020).
56. Legendre, P., Borcard, D. & Roberts, D. W. Variation partitioning involving orthogonal spatial eigenfunction submodels. *Ecology* **93**, 1234–1240 (2012).
57. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
58. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
59. Lahouar, A. & Slama, J. B. H. Day-ahead load forecast using random forest and expert input selection. *Energy Convers. Manage.* **103**, 1040–1051 (2015).
60. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI* **14**, 1137–1145 (1995).
61. Piñeiro, G., Perelman, S., Guerschman, J. P. & Paruelo, J. M. How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecol. Model.* **216**, 316–322 (2008).
62. Friedl, M. & Sulla-Menashe, D. *MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006* (NASA EOSDIS Land Processes DAAC, 2019); https://doi.org/10.5067/MODIS/MCD12Q1.006
63. The CMIP6 landscape. *Nat. Clim. Change* **9**, 727 (2019).
64. Meinshausen, M. et al. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109.1**, 213–241 (2011).
65. Cao, X. et al. A taxonomically harmonized and temporally standardized fossil pollen dataset from Siberia covering the last 40 kyr. *Earth Syst. Sci. Data* **12**, 119–135 (2020).
66. Cao, X. et al. A late Quaternary pollen dataset from eastern continental Asia for vegetation and climate reconstructions: set up and evaluation. *Rev. Palaeobot. Palynol.* **194**, 21–37 (2013).
67. Li, C. et al. Harmonized chronologies of a global late Quaternary pollen dataset (LegacyAge 1.0). *PANGAEA* https://doi.org/10.1594/PANGAEA.933132 (2021).
68. *GlobalTreeSearch Online Database* (Botanic Gardens Conservation International, UK, accessed 20 January 2022); https://tools.bgci.org/global_tree_search.php

Acknowledgements

We thank M. Berdugo for advice on the alternative stable states and hysteresis section, and B. M. Benito for advise on the biogeographical analysis of the paleopollen databases. This research was funded by the European Research Council (ERC Grant agreement 647038 (BIODESERT)) and Generalitat Valenciana (CIDEGENT/2018/041). E.G. was supported by the Consellería de Educación, Cultura y Deporte de la Generalitat Valenciana, and the European Social Fund (APOSTD/2021/188). D.A-S. was partially supported by DETECTOR (grant no. A-RNM-256-UGR18, Universidad de Granada/FEDER), LifeWatch SmartEcoMountains (grant no. LifeWatch-2019-10-UGR-01, Ministerio de Ciencia e Innovación/Universidad de Granada/FEDER), RESISTE (grant no. P18-RT-1927, Consejería de Economía, Conocimiento y Universidad from the Junta de Andalucía/FEDER) and EBV–ScaleUp project (funded by Google Earth Engine and the Group on Earth Observations). M.D-B. acknowledges support from the Spanish Ministry of Science and Innovation (for the I+D+i project PID2020-115813RA-I00 funded by MCIN/AEI/10.13039/501100011033), and from a project of the Fondo Europeo de Desarrollo Regional (FEDER) and the Consejería de Transformación Económica, Industria, Conocimiento y Universidades of the Junta de Andalucía (FEDER Andalucía 2014-2020 Objetivo temático '01 - Refuerzo de la investigación, el desarrollo tecnológico y la innovación') associated with the research project P20_00879 (ANDABIOMA). S.T. was supported by DeepL-ISCO (grant no. A-TIC-458-UGR18, Ministerio de Ciencia e Innovación/FEDER), BigDDL-CET (grant no. P18-FR-4961, Proyectos I+D+i Junta de Andalucia 2018) and LifeWatch SmartEcoMountains.

Author contributions

E.G., M.D.-B. and F.T.M. developed the original idea of the analyses presented in the manuscript. E.G and D.A.-S. developed the global survey. Artificial intelligence and remote sensing analyses were done by E.G. and S.T. Statistical modelling, mapping and data interpretations were done by E.G. and M.D-B. The manuscript was written by E.G., F.T.M. and M.D.-B., with contributions from all co-authors.

Competing interests

The authors declare no competing interests

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41477-022-01198-8.

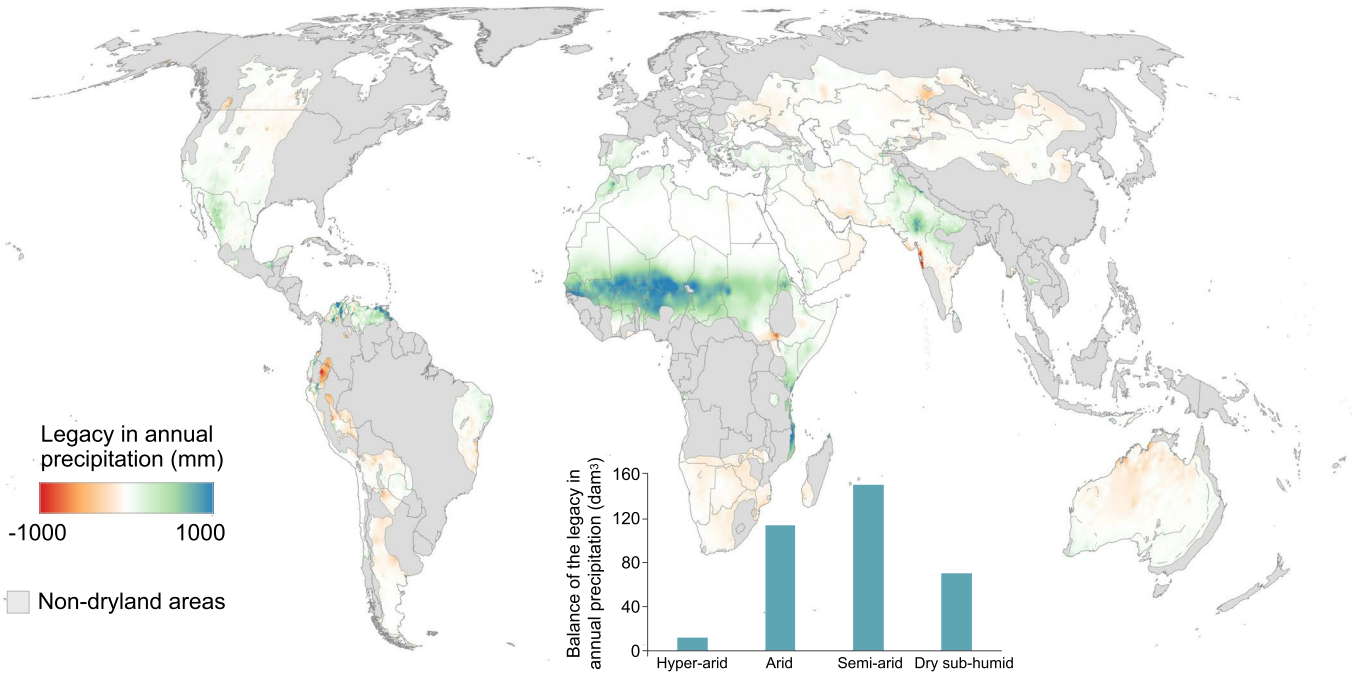
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41477-022-01198-8.

Correspondence and requests for materials should be addressed to Emilio Guirado.

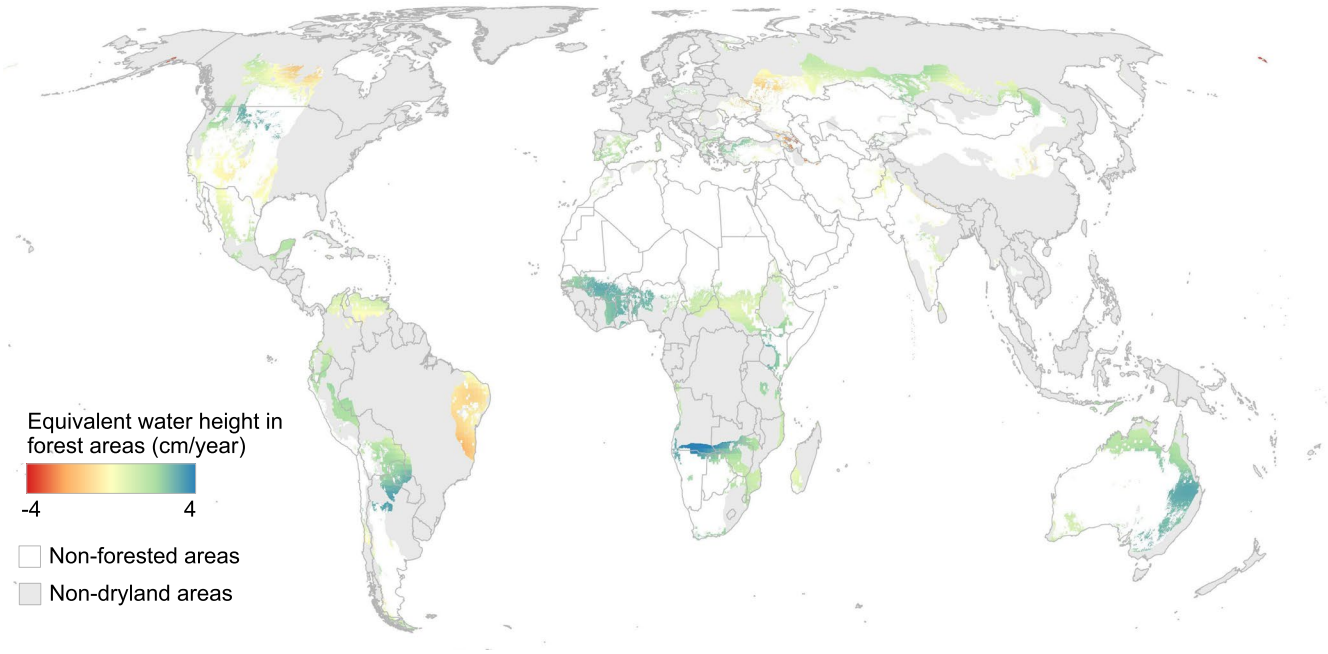
Peer review information *Nature Plants* thanks Jean-François Bastin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

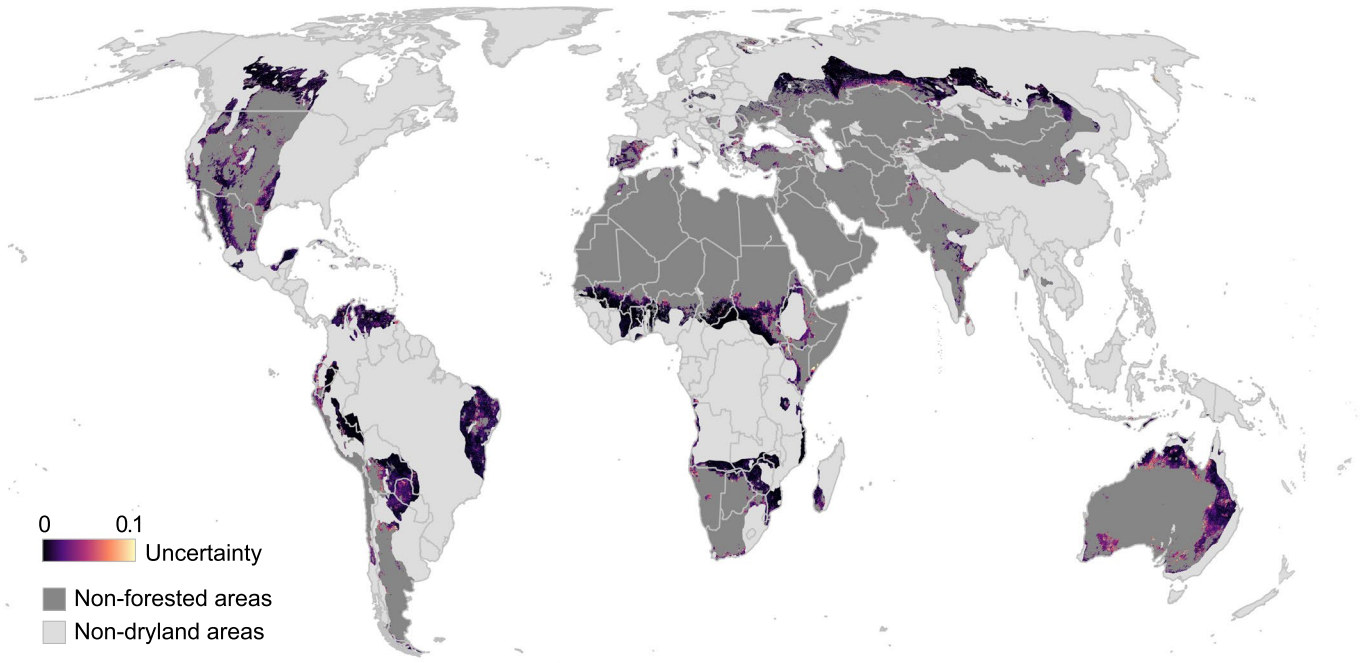
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



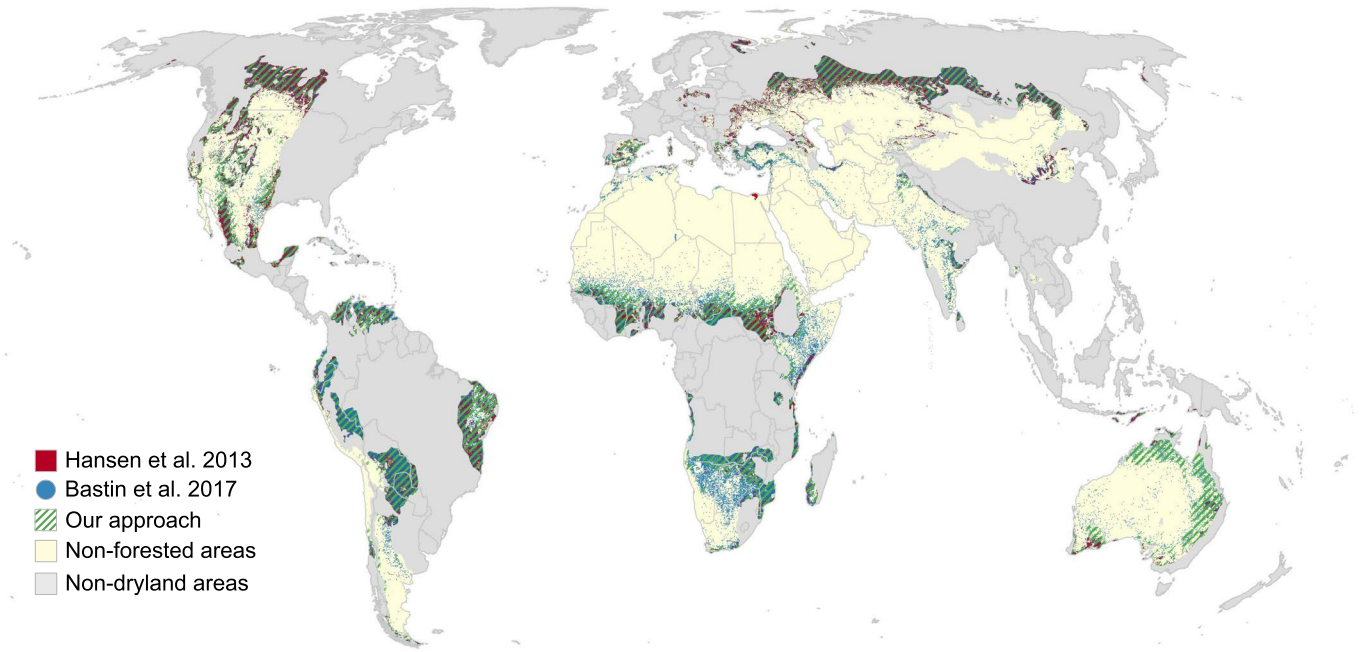
Extended Data Fig. 1 | Legacies in annual precipitation across global drylands. The bar graphs show the sum of the annual precipitation legacy (precipitation from 6000 years before present minus current precipitation) in dam^3 for each aridity level.



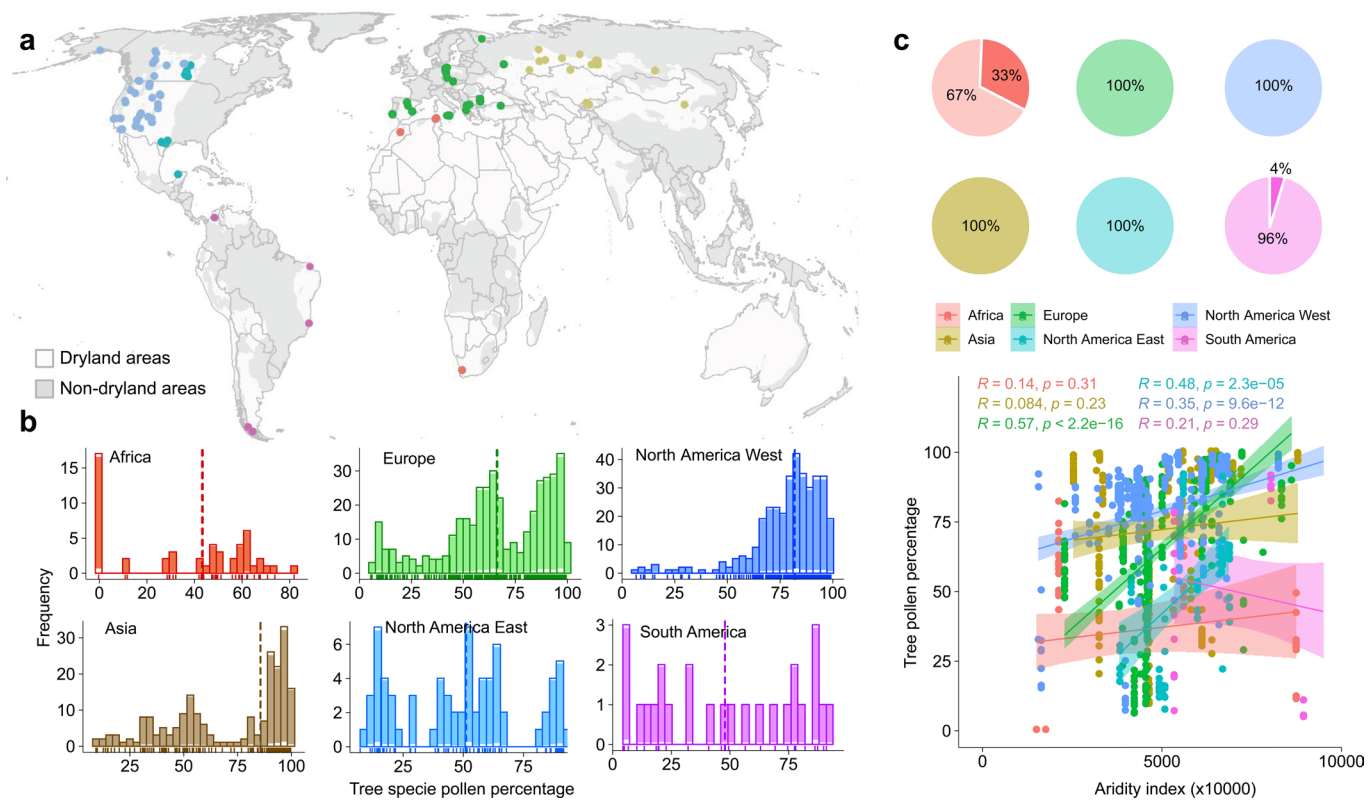
Extended Data Fig. 2 | Trends (2002-2017) of aquifers located beneath forest areas in drylands.



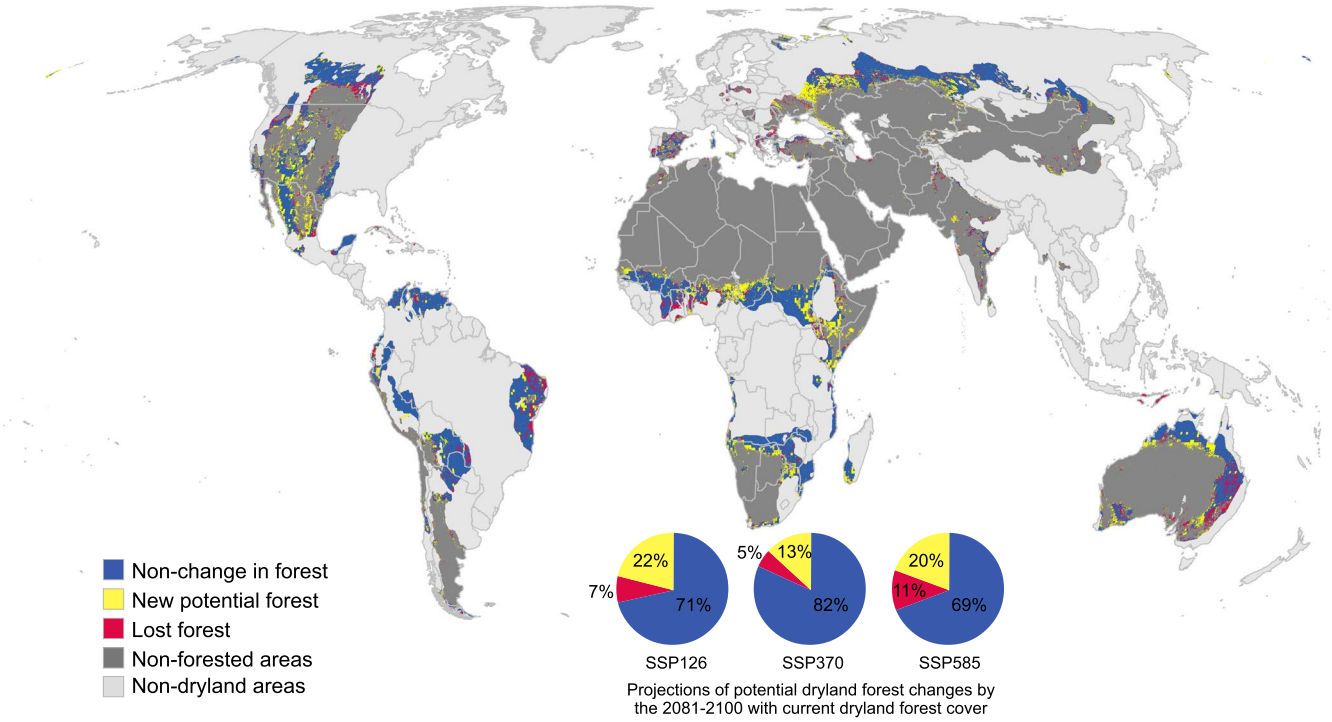
Extended Data Fig. 3 | Uncertainty map of the current extent of dryland forests. Data based on the standard deviation of predictions obtained from the 5-fold cross-validation (see Methods section).



Extended Data Fig. 4 | Comparison of our estimates of forest areas in drylands with recent estimates from the literature[15,36].



Extended Data Fig. 5 | Results of the analyses of pollen samples of mid-Holocene tree species from the LegacyPollen v.1 database[42]**.** a) Spatial distribution of the studied mid-Holocene pollen samples (n=1121 at 119 sites) overlapping with current dryland forest areas (Fig. 2). b) Frequency histogram of the percentage of tree species pollen and its median represented as a dashed vertical line found in the samples per zone. c) Percentage of arboreal pollen found in the samples above 5% in pie charts (top) and Spearman-based correlation of the percentage of arboreal pollen with the aridity index (bottom). Significant in Europe (p-value < 2.2e-16), North America East (p-value = 2.3e-5) and North America West (p-value = 9.6e-12). Shades surrounding the lines represent 95% confidence interval.



Extended Data Fig. 6 | Future changes (potential gains and losses) in dryland forest extent. Data for the 2081-2100 time period according to Shared Socio-economic Pathways (SSP) and representative concentration pathways (RCP) scenarios 1–2.6, 3–7.0 and 5–8.5. Results represented in the map are from the SSP5-RCP8.5 scenarios (see Methods for details).

nature portfolio

Corresponding author(s): Emilio Guirado

Last updated by author(s): May 27, 2022

Reporting Summary

Nature Portfolio 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 Portfolio 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	The results, calculated as described in the Methods, are based on the data from Worldclim v2, MODIS/Terra, ALOS, TerraClimate, Soilgrids, GRACE databases and LegacyPollen v.l and GlobalTreeSearch v.1.5 database, all of which are publicly available.
Data analysis	R version 3.6.0 - "Planting of a Tree", QGIS 3.14, Tensorflow 1.4 and python 3.6

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data generated or analysed during this study, which support the maps within this paper and other findings of this study, are available from Figshare, https://doi.org/10.6084/m9.figshare.13635212. The CNN-based code for classification forest/non-forest described in the Methods is freely available at https://github.com/EGuirado/CNN-Forest-Drylands

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

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

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

Ecological, evolutionary & environmental sciences study design

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

Study description	We first compiled an updated, globally consistent, and accurate dataset of precise locations of forest and non-forest plots. Based on the valuable information from the 213,795 0.5-ha plots provided by the Global Drylands Assessment. We selected 94,352 plots where very high-resolution Google Earth images. Then, the images of these selected plots were automatically classified into forest and non-forest using a CNN-based model. Second, we calculate the variables important for the distribution of forests in drylands. We then created a map of dryland forests for the present and projected future. Finally, with these maps we were able to obtain areas of special interest for the restoration of these ecosystems.
Research sample	Using remote sensing data and artificial intelligence (deep learning), we quantified and mapped the extent of forests in global drylands considering past (Mid-Holocene) and present climate and other environmental predictors (e.g. aquifer trend, albedo, evapotranspiration, slope, soil moisture, elevation, PH, soil organic carbon, soil texture, soil nitrogen). In addition, current and future considerations for forest restoration in drylands are offered.
Sampling strategy	A total of 94,352 plots were selected with very high resolution Google Earth imagery from a systematic grid of 213,795 0.5 ha plots provided by the Global Dryland Assessment.
Data collection	The data were extracted from freely available databases. The new forest and non-forest plot classification data are available from their respective publicly available in the supplementary data.
Timing and spatial scale	Time scales included climatic and Mid-Holocene paleo-pollen data. Climatic, environmental and soil data from the present day, and climatic and land use data from the year 2081-2100.
Data exclusions	We do not exclude any data.
Reproducibility	All results can be reproduced, as all data associated with the results are freely available.
Randomization	The quality of the classification was tested and validated using a k-fold cross validation method, where k models (k=5) were trained from k subsets of the original data and tested on k subsets of the remaining independent data (total number of plots divided by k).
Blinding	Blinding was not used in our analyses. Not applicable.

Did the study involve field work? ☐ Yes ☒ No

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 & experimental systems

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

Methods

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