

# ECOGRAPHY

## Research article

# Global biogeographical regions reveal a signal of past human impacts

Marta Rueda<sup>1,2</sup>, Manuela González-Suárez<sup>3</sup> and Eloy Revilla<sup>2</sup>

<sup>1</sup>Departamento de Biología Vegetal y Ecología, Universidad de Sevilla, Sevilla, Spain

<sup>2</sup>Departamento de Biología de la Conservación, Estación Biológica de Doñana (EBD-CSIC), Sevilla, Spain

<sup>3</sup>Ecology and Evolutionary Biology, School of Biological Sciences, University of Reading, Reading, UK

Correspondence: Marta Rueda ([mruedag@us.es](mailto:mruedag@us.es))

## Ecography

2023: e06762

doi: [10.1111/ecog.06762](https://doi.org/10.1111/ecog.06762)

Subject Editor:

Manuel J. Steinbauer

Editor-in-Chief: Miguel Araújo

Accepted 15 October 2023



Ecologists have long documented that the world's biota is spatially organised in regions with boundaries shaped by processes acting on geological and evolutionary timescales. Although growing evidence suggests that historical human impact has been key in how biodiversity is currently assembled, its role as a driver of the geographical organisation of biodiversity remains unclear. Using non-volant terrestrial mammals we set up a bioregionalization procedure focused on two datasets, one describing the current ranges of terrestrial mammals, and another describing their potential natural ranges in absence of historic anthropogenic land use. We then quantified the relative importance of anthropogenic land use (5000 and 2000 years ago, and present time) to predict the current and natural biogeographical regions across the Earth. In general, past and present human land use were important predictors of current bioregions but did not largely contribute to predict natural bioregions. Past anthropogenic land use seems to have left an imprint on the taxonomic differentiation of some of the largest biogeographical realms, whereas land use at present stands out as a driver of the taxonomic differences between medium-sized subregions, i.e. within and among continents. Our findings suggest that anthropogenic actions during the last millennia have had a far-reaching effect on the spatial organisation of the Earth's non-volant mammals.

Keywords: bioregionalization, historic anthropogenic land use, terrestrial mammals

## Introduction

How is the world's biodiversity organized, and why do large-scale patterns of taxonomic diversity change through natural geographic regions? These questions have attracted the attention of naturalists since the early 19th century (von Humboldt 1806, Wallace 1876, Ricklefs 2004, Daru et al. 2017, Dinerstein et al. 2017). A key step in understanding the organisation of biodiversity is to identify the assemblage of regions based on their shared organisms (Wiens 2011). Wallace was among the first to propose that the world's fauna is organized hierarchically in broad regions



[www.ecography.org](http://www.ecography.org)

© 2023 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

shaped by geographic and climatic factors (Wallace 1876). About 150 years later, the development of multivariate analytical techniques together with an increase in the availability of global species distribution databases have led to reevaluations of Wallace's proposal (Kreft and Jetz 2010, Procheş and Ramdhani 2012, Holt et al. 2013, Rueda et al. 2013) and the improvement of our understanding about the extrinsic determinants driving the main dissimilarities among bioregions (Ficetola et al. 2017). Broadly, what we know is that the interplay of multiple factors has jointly contributed to the formation of bioregions, but their prominence varies across the globe (Riddle and Hafner 2010, Ficetola et al. 2017, Mazel et al. 2017). Processes acting deeper in the past, like plate tectonics, are the most important in explaining the separation between strongly divergent biogeographical realms. Factors representing present-day ecological barriers, like climatic heterogeneity, determine the separation between less dissimilar regions, i.e. within and among continents; while mountains – the third factor in importance – operate at all levels of biogeographical differentiation (Ficetola et al. 2017).

The potential role of anthropogenic impact in shaping bioregions has been largely ignored because extant biogeographical regions have traditionally been assumed to reflect the natural organisation of biodiversity that resulted from ecological, historical, and evolutionary processes acting over millions of years (Lomolino et al. 2010). Yet, rising evidence showing that human-mediated species introductions are already affecting extant biogeographic regions (Capinha et al. 2015, Bernardo-Madrid et al. 2019) challenges this view. Furthermore, accumulating evidence shows that Quaternary human activities induced shifts in the plant and animal communities we see today (Lyons et al. 2016), and there is little doubt that modern humans have been a major driver in the extinction of large mammals during the late Pleistocene and early Holocene (Sandom et al. 2014, Smith et al. 2018). Moreover, archaeological, palaeoecological and genetic data suggest that cumulative human transformation of ecosystems over millennia has resulted in dramatic changes in composition, community structure, richness and genetic diversity of a diverse array of organisms across taxonomic groups (Ellis et al. 2013, 2021, Boivin 2016, Mottl et al. 2021). This body of evidence raises the question of whether historical anthropogenic pressures may have been large enough to leave an imprint detectable today on biogeographical assemblages globally.

To answer this question, we set up a bioregionalization and analytical procedure focused on two datasets; one describing the current distribution ranges of global terrestrial mammals, and another set inferring their present natural ranges, which represents estimates of where species would potentially live without anthropogenic pressures (Faurby and Svenning 2015, Faurby et al. 2018). Using both types of distribution ranges we build hierarchical bioregionalizations and compute models to identify the variables (plate tectonics, climatic heterogeneity, mountains, and anthropogenic land use over the Holocene, specifically 5000 and 2000 years ago as well as present time) that best predict the assemblage of species in bioregions. We hypothesize that if anthropogenic land use has had a notable

effect on the biogeographical assemblages and boundaries as we know them, then 1) the biogeographic configurations built using current distribution ranges will differ from those based on natural ranges; and 2) human land use should predict the biogeographic patterns obtained using current ranges, but not patterns obtained using natural ranges.

## Material and methods

### Generating presence–absence matrices using range distributions

Current and natural distribution ranges of terrestrial mammals were obtained from PHYLACINE 1.2 (Faurby et al. 2018), which contains range maps for all 5831 known mammal species that lived since the last interglacial (~ 130 000 years ago until present). Ranges for current extant species contained in PHYLACINE are IUCN (2016) distribution maps, whereas present natural ranges represent estimates of where species would potentially live if they had never experienced anthropogenic pressures, and – in the case of extinct species – had not gone extinct (Faurby and Svenning 2015 for details). Note that Faurby and Svenning (2015) did not use present or past human land use variables to estimate PHYLACINE natural ranges, thus, natural ranges were not defined considering the predictors we test here avoiding circularity. Both current and natural ranges in PHYLACINE are projected to Behrmann cylindrical equal area rasters with a cell size of 96.5 by 96.5 km. Using these data, we generated two presence–absence matrices, one for the current and one for the natural ranges, in which every row represents a grid cell and every column a species. From these matrices we excluded 1) *Homo* species; 2) bats, because their high dispersal capacity and their almost global distribution make it difficult to establish clear biogeographical boundaries (Wallace 1876); 3) marine mammals, including polar bears, and other fully aquatic species (<https://marinemammalscience.org/science-and-publications/list-marine-mammal-species-sub-species> for a full list of marine species not included); 4) cells containing less than 50% of land area to approximate equal-size samples; and 5) cells containing fewer than five species to reduce potential distortions caused by having few taxa (Kreft and Jetz 2010). These exclusion criteria rendered a total of 14 087 cells representing 3960 terrestrial mammal species for the current matrix, and 14 151 cells representing 4306 for the natural matrix.

Given that differences between the biogeographical configurations of current and natural bioregions could be driven by the huge distribution ranges typical of extinct megafauna, we also built presence–absence matrices using only the extant current mammals (i.e. not including fossil ranges) that have had range contractions (i.e. have and estimated natural distribution range in PHYLACINE different from the IUCN current range). Overall, we obtained 536 extant mammals whose distribution ranges have changed. Note that a few extant species became extinct on some continents as part

of the megafauna extinction event but survived on others. Consequently, we adjusted distribution ranges for those continentally extirpated species. To do that, first we did a literature review (Barnosky et al. 2004, Koch and Barnosky 2006) to find those species that had become extinct prehistorically on some continents but survived on others (Eurasia was considered a single continent) (Supporting information for species selected). Then, we excluded any portion of the species' range found on that continent.

### Building biogeographical regionalizations

To each of the presence–absence matrices we applied a machine-learning algorithm referred to as affinity propagation (AP hereafter) (Frey and Dueck 2007) to build biogeographical regionalizations at different biogeographical resolution. We chose AP because it can compress massive datasets very efficiently (i.e. with lower error in comparison with other clustering methods, Frey and Dueck (2007) for details), and its good performance in hierarchical bioregionalization procedures (from few grid cells to large realms) has already been demonstrated in a previous work (Rueda et al. 2013). The AP algorithm works by detecting special data points called exemplars, and by a message-passing procedure it iteratively connects every data point to the exemplar that best represents it until an optimal set of exemplars and clusters emerges. In contrast to algorithms in which exemplars are found by randomly choosing an initial subset of data points, AP takes as input measures of 'similarities' between pairs of data points (grid cells here) and simultaneously considers all the points as potential exemplars. The optimal set of exemplars is the one for which the sum of similarities of each point to its exemplar is maximized.

We first used the presence–absence matrices to calculate pairwise similarities between pairs of cells, and selected Hellinger distance as a similarity index. The Hellinger distance is a modification of the Euclidean distance (Legendre and Gallagher 2001) and is used to avoid the 'double-zero' problem, i.e. when two sites that have no species in common are assigned the same distance as two sites that share the same species; and the 'species–abundance paradox', which frequently occurs when two sites share only a small fraction of all the species in the same regional pool. This is expected to be a particular problem at the margins of biogeographical regions where sites may be quite different from one another rather than in the centre of a region where sites are likely to be very similar in their species assemblages (Legendre and Legendre 1998, Gagné and Proulx 2009). We then used a protocol based on a successive application of AP (Rueda et al. 2013) to obtain a biogeographical upscaling from the smallest possible bioregions (i.e. the highest biogeographical resolution) to the largest ones. We performed an initial AP analysis involving all grid cells of the similarity matrix. This first AP run generates the optimal solution of the highest resolution bioregions, while also identifying its exemplars. We obtained 1128 clusters and exemplars for the current distributions and 1053 for the natural distributions. Then, using the exemplars

as the new units of analysis, we again conducted an AP, i.e. we calculated a new similarity matrix and re-run a new AP. This process was repeated until a small and coherent number of large clusters emerged; specifically, four runs were needed to get nine (with the current distributions) and eight (with the natural distributions) large bioregions (Supporting information). Finally, to obtain maps of each clustering result, we classified each grid cell (row) of every presence–absence matrix according to the cluster to which they were assigned in its corresponding AP analysis. AP analyses were performed using the 'APCluster' R package ([www.r-project.org](http://www.r-project.org), Bodenhofer et al. 2011).

### Assessing differences in biogeographical configurations

The degree of similarity between clusters (i.e. bioregions) of current and natural bioregions was estimated using the Jaccard index. For that, we used the *cluster\_similarity* function of the 'clusteval' R package ([www.r-project.org](http://www.r-project.org), Ramey 2012). This function computes the similarity between two clusterings of the same data. Jaccard index ranges between 0 (no similarity) to 1 (perfect match).

### Predictors

To each of the cells of the presence–absence matrices we assigned a mean value of several predictors. We considered variables previously tested as determinants of biogeographical boundaries related to climatic heterogeneity, orographic barriers, tectonic movements, and instability of past climate (Ficetola et al. 2017) plus variables related to anthropogenic land use over the Holocene. In particular, we tested four climatic variables: annual total precipitation, mean annual temperature, seasonality in temperature, and seasonality in precipitation. All climatic variables were extracted from the WorldClim dataset (Fick and Hijmans 2017) up-scaled at a 96.5 km resolution. These variables represent both average conditions and variability within years and have been shown to be determinants of vertebrate distributions (Sexton et al. 2009). Climatic conditions have strongly shifted over the Quaternary period, and have been shown to play a role in the present-day species distributions, endemism, and assemblages (Araújo et al. 2011). To test for the potential effect of past climate change or stability, we calculated the average velocity of climate change since the Last Glacial Maximum (LGM; ~ 22 000 years ago) and since the Mid-Holocene (MH; ~ 6000 years ago) (Sandel et al. 2011). For that, we used the mean annual temperature and annual total precipitation for the MH and the LGM calculated by means of the model for interdisciplinary research on climate (MIROC-ESM) (Watanabe et al. 2011).

Plate tectonics are responsible of the long-term isolation of the biotas, whereas mountain ranges represent major barriers to dispersal for most mammals (Lomolino et al. 2010, Ficetola et al. 2017). Plate tectonics were obtained from Bird (2003). Each grid cell was assigned the tectonic plate to which

it belongs. We removed those cells that were represented by more than one tectonic plate to avoid potential wrong 'co-occurrence' patterns, i.e. mixing faunistic elements separated by different plates even though the species do not actually co-occur. We used the GTOPO30 to calculate the difference between minimum and maximum elevation per grid cell to obtain a metric of elevational variance aimed to capture the complexity of topographic changes.

Finally, past anthropogenic land use was obtained from Ellis et al. (2013). This dataset consists of two global spatially explicit reconstructions two main models of land use across the Holocene, HYDE and KK<sub>10</sub> models. We chose the more realistic KK<sub>10</sub> model, which assumes that humans use land more intensively when population density is high and land scarce (Kaplan et al. 2011). In counterpoint, the HYDE model omits land-use intensification and predicts that except for the developed regions of Europe, human use of land was insignificant in every biome and region before 1750 (start of the industrial revolution). The KK<sub>10</sub> model dataset contains raster information for human land use for 10 different time periods; 8000, 5000, 3000, and 2000 years ago, and 1000, 1500, 1750, 1900, 1950, and 2000 of the present era. We first selected four different time spans: 8000, 5000, and 2000 years ago, and current (year 2000 of the present time), but the descriptor of land use 8000 years ago was not included in the random forest models due to its low variability at global level (Supporting information).

### Evaluating potential drivers of the biogeographical configurations

We used random forest classification models (Breiman 2001) to assess the factors that may predict the classification of cells bioregions and to estimate the relative importance of the predictors. Note that random forest models were only carried out for large and medium-sized bioregions or subregions, as some of the predictors (e.g. plate tectonics) do not make ecological sense to explain smaller bioregions. Random forest is a machine learning method based on a combination of a large set of decision trees. Each tree is trained by selecting a random set of variables and a random sample from the training dataset (i.e. the calibration dataset). The accuracy of the models is given by the out-of-bag (OOB), an estimate of the misclassification rate that represents an unbiased error rate of the model that is calculated by counting how many cases in the training set are misclassified and dividing the number by the total number of observations.

First, we split every dataset into training set (70% data, used to fit the model) and test set (30% data, used to evaluate the performance of the model). Second, with the training set we performed cross-validation to evaluate different hyper parameters of the model using grid random search (Supporting information). Random forests main hyper parameters include the number of decision trees in the forest (*n*tree) and the number of features considered by each tree when splitting a node (*m*try). Values for *n*tree ranged between 7505 to 4604, while values for *m*try ranged between 3 and

7 (Supporting information). Third, we run the Random Forest classification models using the training dataset and, finally, we estimated the accuracy of our models by checking the predictions against the actual values in the test dataset. Random forest models were computed using the R package 'RandomForest' (Liaw and Wiener 2002). Additionally, we also defined classification models using XGBoost (R package 'xgboost', [www.r-project.org](http://www.r-project.org), Chen et al. 2023) to compare algorithm performance in terms of accuracy. Here, we report results from random forest as it had better overall accuracy.

Random forests are able to disentangle interacting effects and identify non-linear and scale-dependent relationships that often occur at the scale of the analysis performed here among multiple correlated predictors (Cutler et al. 2007). Evidence from genomic studies suggests that variable importance measures may show a bias towards correlated predictor variables (Nicodemus et al. 2010), hence we eliminated those predictors showing a moderate to high Pearson's correlation ( $r > 0.50$ , Supporting information). As a result, our final models included seasonality in temperature and precipitation as representatives of climate heterogeneity; velocity of climate change since the LGM and the MH to the present as representatives of past climate change; plate tectonics; variation in elevation representing mountains; and human land use at the present, 5000, and 2000 years ago. Human land use 5000 and 2000 years ago showed a large correlation ( $r=0.87$ ), so we decided to perform separate models for each time period. Globally the correlation between human land use at present and 2000 years ago is relatively low ( $r=0.29$ ); however, we also decided to perform independent models for both predictors because Pearson's coefficient using modified t-test (below) showed that the correlation between both variables is influenced by spatial autocorrelation.

Since our models include spatial data, the selection of predictors with random forest could be partly influenced by spatial autocorrelation. To assess this point, we performed correlations between predictors using the method of t-test modification (Clifford et al. 1989, Dutilleul 1993) to incorporate spatial structure in the Pearson's correlations. This is a hypothesis testing procedure based on a modification version of the correlation coefficient that takes into account the spatial structure of the predictors. This analysis was computed using the function *modified.ttest* of the R 'SpatialPack' package ([www.r-project.org](http://www.r-project.org), Vallejos et al. 2020). The results showed evidence of spatial autocorrelation in many of the correlations, even for low Pearson's correlation values (Supporting information). Therefore, to account for spatial autocorrelation we also run random forest classification models incorporating spatial structure as a possible predictor. To do that we first generated Moran's Eigenvector Maps (MEMs) (Dray et al. 2006), and ranked and reordered them based on their Moran's I. These eigenvectors can be included in model as spatial predictors. Here we include the first three eigenvectors. MEMs were calculated using the *mem* function from the 'spatialRF' package in R ([www.r-project.org](http://www.r-project.org), Benito 2021).

The algorithm used in the process of regionalization only uses species presence-absences, which means that

humans can only have a role if present and if causing local extinctions, thereby changing the amount of shared species between grid cells. At the largest biogeographical scale two bioregions, Madagascar and Australia, appear characterised by an extremely low historical human land use (Supporting information). Note that in both cases human pressures may be poorly captured by the input layers of the  $KK_{10}$  model used to define the variables of past anthropogenic land use, as the role of pre-historic human impact in the extinction of the megafauna of both Australia and Madagascar is well known (Miller et al. 2005, Hansford et al. 2018). In any case, the ‘false’ absence of human impact could incorrectly give support for historical land use as a cause. Besides, both regions have a 100% endemism for non-flying mammals, meaning that humans cannot change their clustering without introducing species. Therefore, to assess to what extent these factors might be affecting the results, we also ran the models excluding Australia and Madagascar. Finally, we note that our models are unbalanced regarding the number of climate versus human land use variables, which overemphasize the importance of human impact. To evaluate this possibility, we also fitted models without climate variables.

### Calculating global and local predictor importance

We measured variable importance using the mean decrease in accuracy, which is obtained by permuting randomly each variable and assessing the decrease in classification accuracy of the model (Liaw and Wiener 2002). Thus, the greater the loss of accuracy due to the permutation (or exclusion) of a single variable, the more important that variable is considered, i.e. variables with a large mean decrease in accuracy are more important for classification of the data.

Random forest also calculates the local variable importance, which defines the importance of a variable in the classification of a single sample (grid cell here) and therefore shows a direct link between variables and samples (Touw et al. 2012). The local importance score is derived from all trees for which the sample was not used to train the tree. For each sample, the percentage of votes for the correct class in the variable permuted OOB data is subtracted from the percentage of votes for the correct sample in the original OOB data (i.e. the untouched

OOB data) to assign a local importance score. The score reflects the impact of a variable on correct classification of a given sample: negative (i.e. the variable does not positively contribute to the classification), 0 (the contribution is neutral) and positive (the variable positively contributes to the classification). Given that local importance values are noisier than global importance ones we run the same classification five times (five per biogeographical scale) and averaged the local importance scores to obtain a robust estimation of local importance values (Touw et al. 2012). We present maps of mean local importance values for present and past human land use. In these, it is important to note that high local importance can reflect both areas where human land use was high and areas where it was negligible, i.e. both are equally valuable for the classification analysis (Supporting information). To facilitate interpretation, in the main text we present maps of local importance only for areas where land used was  $> 10\%$ .

All analyses were performed in the R ver. 4.3.0 ([www.r-project.org](http://www.r-project.org)).

## Results

### Bioregionalizations

Both for the current and natural bioregionalizations we identified a hierarchical system of biogeographic regions with four levels (Fig. 1). For current mammalian ranges, the broadest delineation of nine large bioregions was mostly consistent with the original maps of Wallace’s realms and subregions (Fig. 1a). For natural mammalian ranges, the eight bioregions obtained showed a different biogeographic arrangement with respect to the nine large current ones, especially in the Northern Hemisphere (Supporting information). Thus, while the natural broadest bioregions show a single Holarctic region, in the current ones this is separated into a Nearctic and a Palearctic region. Differences between current and natural bioregions extended to the subregions or medium-sized bioregions and smaller bioregions (Supporting information)

and were confirmed for the bioregionalizations carried out with the 536 species with different natural and current ranges (Supporting information).

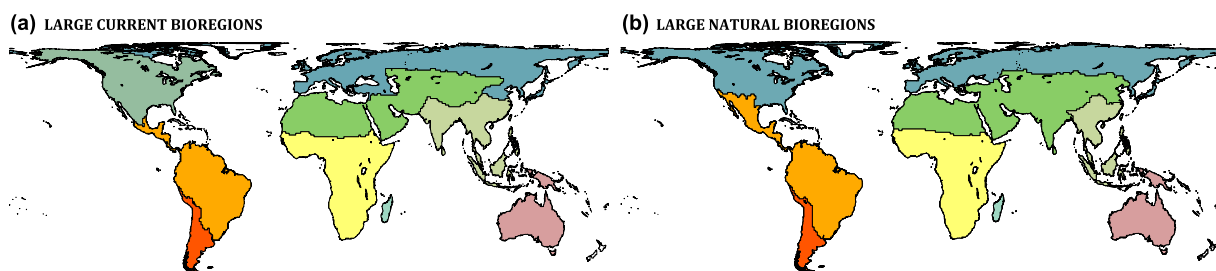


Figure 1. Current and natural broader regions or realms. Only results for the broader regions are shown here: nine current bioregions (a), and eight natural bioregions (b). The greatest differences between both biogeographical configurations are more relevant in the Northern Hemisphere, where, for the natural bioregions, the Palearctic and the Nearctic are united forming a great extension, the typical biogeographical boundary between the Nearctic and the Neotropical realms disappears and extends towards the north, and India is no longer part of the Asian region and is included instead within the ‘Saharo-Arabian’ region.

## Global importance of predictors of the biogeographical patterns for regions and subregions

For the broader bioregions, human land use at present had a moderate relevance in the classification models (Fig. 2a). In contrast, past human land use (5000 and 2000 years ago) was the third most important predictor of the biogeographic structure of current realms behind plate tectonics and seasonality in temperature (Fig. 2c, e), with a moderate, non-negligible predictive value (~ 22 and 21% of cells classified correctly when the model was run just with human land use 2000 and 5000 years ago, respectively, Fig. 3). In broad accordance with our prediction, past and present human land use were poorer predictors of natural bioregions (Fig. 2b, d, f). Results were consistent also if we excluded Australia and Madagascar (Supporting information). The misclassification rate (out-of-bag, OOB hereafter) for these set of models was quite low, never exceeding 3.6%, and the model accuracy ranged around 0.97. For medium-sized bioregions or subregions, past and present human land use were the second predictors in importance for the current bioregions (Fig. 4a, c, d; OOB values in general < 8.6%, model accuracy 0.92–0.94). As hypothesized, the importance of land use at present was smaller for the natural bioregions; although, unexpectedly, past land use remained a strong driver (Fig. 4b, d, e).

Results for models accounting for the spatial autocorrelation (Supporting information) showed a similar tendency, especially for the broader bioregions and past human land use (OOB values in general < 2.5%, model accuracy 0.98–0.99). In fact, for these models the differences for past human impact are sharpened when current broader bioregions are compared with the natural ones. On the contrary, the importance of human land use at present fades in favor of other predictors when the spatial structure is taken into account for both current and natural broader bioregions. For medium-sized bioregions, the importance of plate tectonics decreases and it is no longer the main predictor (OOB values in general < 3.8%, model accuracy 0.96–0.97). When accounting for spatial autocorrelation the importance of human land use at present becomes greater for current bioregions and decreases for natural bioregions. However, the past human land use has a completely opposite effect to that expected since it becomes the main predictor of medium-sized natural bioregions, implying that spurious factors might be acting on this scale.

Models without climatic variables had lower accuracy as expected but did not alter our main findings with similar variable importance rankings that support the role of human land use (Supporting information).

## Local importance of predictors of the biogeographical patterns

Maps showing local importance scores revealed differences in the location of grid cells where past and present human land use have been analytically important for the classification of

the broader current bioregions (Fig. 5). The cells where human land use at present was most relevant are located in southern Africa, the easternmost boundary between the Palearctic and the Asian regions, and part of the northern boundary between the Palearctic and 'Saharo-Arabian' regions (Fig. 5a). The cells where human land use 2000 years ago was more important are located mainly throughout Central America and the Andes mountain range, Europe except for the northernmost part, most of the Asian region, and the boundary between the Palearctic and the Asian regions (Fig. 5b). The cells where human land use 5000 years ago was more important mostly coincided with those of 2000 years ago, but with minor scores of importance (Fig. 5c). Some of these areas coincided for the broader natural bioregions, yet the importance of Central America and the Andes for the classification of natural bioregions weakened (Fig. 5e–f) whereas the importance of the area of southern India increased (Fig. 5d–f).

For the medium current bioregions, the grid cells where the human land use at present was most relevant for the classification of bioregions included those highlighted for the broader bioregions plus most part of Europe, southern India, eastern Australia, and a large area of the Nearctic (Fig. 6a). When focusing on the human land use 2000 years ago, the areas affected by human land use were similar to those of the broader bioregions plus a Sub-Saharan area crossing Africa from west to east (Fig. 6b). Again, the cells where human land use 5000 years ago was more important mostly coincided with those of 2000 years ago, but it is noteworthy that the importance of the Sub-Saharan area disappears (Fig. 6c). The location of cells affected by human land use for the medium natural bioregions quite coincided with that for current bioregions but with less importance scores generally in Central America and the Sub-Saharan area (Fig. 6e–f).

## Discussion

Our study shows that human activity during the last few millennia has left its mark on the spatial organisation of the Earth's biodiversity. First, we show that biogeographical configurations based on current and natural mammal distributions differ, particularly in the Northern Hemisphere. These differences were not solely the result of including extinct megafauna – with huge distribution ranges – in the natural biogeographic regions, as biogeographical patterns for the 536 extant terrestrial mammals with different current and natural ranges confirmed our findings. Second, we find that human land use is an important predictor of current bioregions but does not largely contribute to predict natural bioregions. For example, for the broader regions or realms, past human land use was the third most important predictor behind plate tectonics and temperature seasonality, but was of minor importance on natural regions. The difference between current and natural bioregions for past human land use was even more exacerbated when spatial autocorrelation was considered, such that past human land use became the second

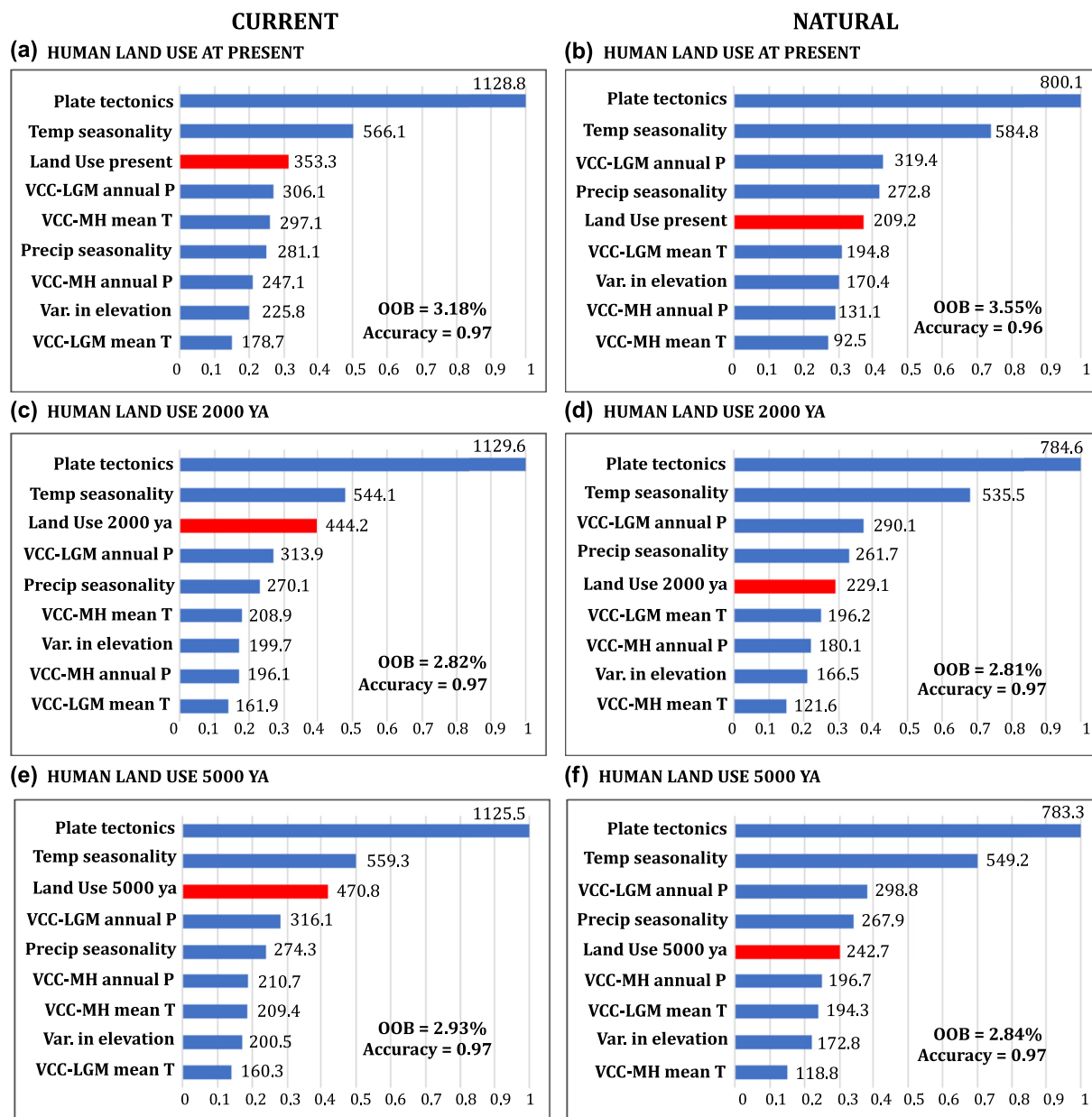


Figure 2. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural broader bioregions. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2000 years ago for (c) current and (d) natural bioregions, and human land use 5000 years ago for (e) current and (f) natural bioregions. The numerical values associated with each horizontal bar are mean decreased in accuracy values (i.e. how much the accuracy decreases when the variable is excluded from the model) and represents variable importance. Importance was measured by the drop-in classification accuracy after predictor randomization in random forests. Consequently, higher values of mean decreased in accuracy indicate variables that are more important to the classification. Values from 0 to 1 in the x-axis represents relative importance values (i.e. division of each value of mean decrease in accuracy by the largest value) and show a ranking of variable importance. OOB (out-of-bag) represents the percentage of cells misclassified in each model. VCC = velocity of climate change; MH = Mid-Holocene; LGM = last glacial maximum; annual P = annual precipitation; mean T = mean annual temperature; ya = years ago; Var. in elevation = variation in elevation.

most important factor after plate tectonics in predicting the configuration of the current broader bioregions. This implies that past anthropogenic actions have seemingly left a perceptible biogeographical imprint on the assembly of the global realms we recognize today. Moreover, for the medium-sized

subregions, human land use at present was among the first most important predictors, suggesting that recent human activity is already causing changes in the biogeographical assemblages at the subregional scale, i.e. within and among continents. Again, the importance of human land use at

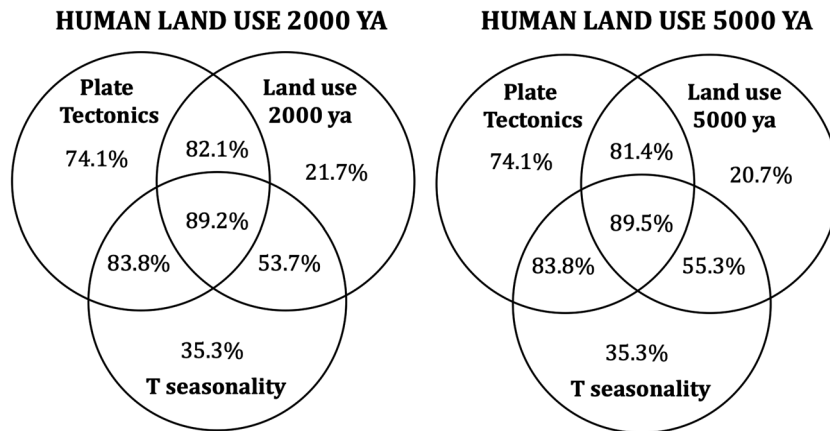


Figure 3. Diagrams describing the accuracy in classification of the three most important drivers of the current broader bioregions for models including human land use 2000 and 5000 years ago. Values indicate the percentage of cells correctly classified (1-OOB error) in their bioregion, and are obtained from running random forest classification models for each individual variable, by pairs and for the three most important factors together. Note that the number of variables modifies the parameters of the models and causes there to be a variation in the out-of-bag (OOB) when compared with models containing all variables.

present increased as a predictor of current medium-sized subregions when the spatial autocorrelation was considered.

The mammalian distributions that we know today are not only a reflection of the most recent human actions but also those exerted during the last few millennia (Faurby and Svenning 2015, Polaina et al. 2020). Recent studies had already shown that human actions in the present, in terms of human-mediated species introductions, can be evident at biogeographical scales (Capinha et al. 2015, Bernardo-Madrid et al. 2019), but here we also show evidence of an effect of past human land use. What has happened from 5000 years ago that has left a noticeable footprint in the Earth's biogeographical regions? This period coincides with the emergence of an 'intense' agricultural land expansion and the settlement of densely populated areas in the Mediterranean, southwest Asia, south Asia, and eastern China (Kirch 2005, Ellis et al. 2013, Stephens et al. 2019). Long-term impacts to enhance agricultural productivity, including forest clearing, increased fire frequencies, species invasions, and soil erosion, were already apparent in some regions at this time (Ellis et al. 2013, Boivin et al. 2016), and may have transformed vegetation structure and species composition across many regions, from the Mediterranean to the Tropics (Stephens et al. 2019, Zheng et al. 2021). In fact, reconstructions suggest that > 20% of Earth's temperate woodlands had been already impacted by humans by the late Holocene (Ellis et al. 2013). Indeed, our results on local importance show the match between some of the biogeographical boundaries and regions of the New World and the distribution – and presumed impact – of ancient Mesoamerican and South American pre-European civilizations (Richard 1997, Dunning et al. 2012). In the Old World, we note the prominence of past human land use covering most of the Asian realm territory, associated to the rise of rice cultivation in eastern, southern, and southeast Asia over the Holocene (Silva et al. 2015, Zheng et al. 2021). Actually, cumulative archaeological data show that the scale of agriculture and land use in these regions was

significant; expansion of land area used for livestock and rice *Oryza sativa* paddy agriculture in southern Asia was sufficient to increase atmospheric methane emissions between 4000 and 1000 years ago (Fuller et al. 2011). Moreover, the atmospheric CO<sub>2</sub> decline registered between 1570 and 1620 has been attributed to the uptake by vegetation following the agriculture abandonment caused by the population crash after the arrival of Europeans in America (Lewis and Maslin 2015). Past human land use can also partly explain differences between the southern Palearctic boundary and the Afrotropic below the Sahara. Both can be related to the transformation of European and Near Eastern landscapes during the Roman period (Butzer 2005), and the expansion of sorghum *Sorghum bicolor* cultivation 3000 to 2000 years ago in Sub-Saharan Africa (Boivin et al. 2016).

While we found differences in importance of human land use 5000 and 2000 years ago between current and natural medium-sized bioregions, we were surprised that past human land use was an important determinant of natural bioregions, which were based on distributions of mammals in the absence of human impact. These results could reflect an overlap between human land use and biogeography. If the distribution of humans during the middle and late Holocene reacted to the same drivers as the distributions of terrestrial mammals, human land use could appear as a predictor even if it had no causal effect. In fact, the importance of this variable appears to be a consequence of how human land use was distributed globally 5000 and 2000 years ago, or more accurately where human activity was largely absent. Note that while current human land use is distributed over wide areas of the planet, that of 5000 and 2000 years ago was centralized in specific areas coinciding with the distribution of ancient civilizations. This means that, on a subregional scale, many bioregions can be well predicted by the absence of human land use. Indeed, an additional analysis that excluded subregions with no reported human land use 2000 years ago showed very little importance of past human land



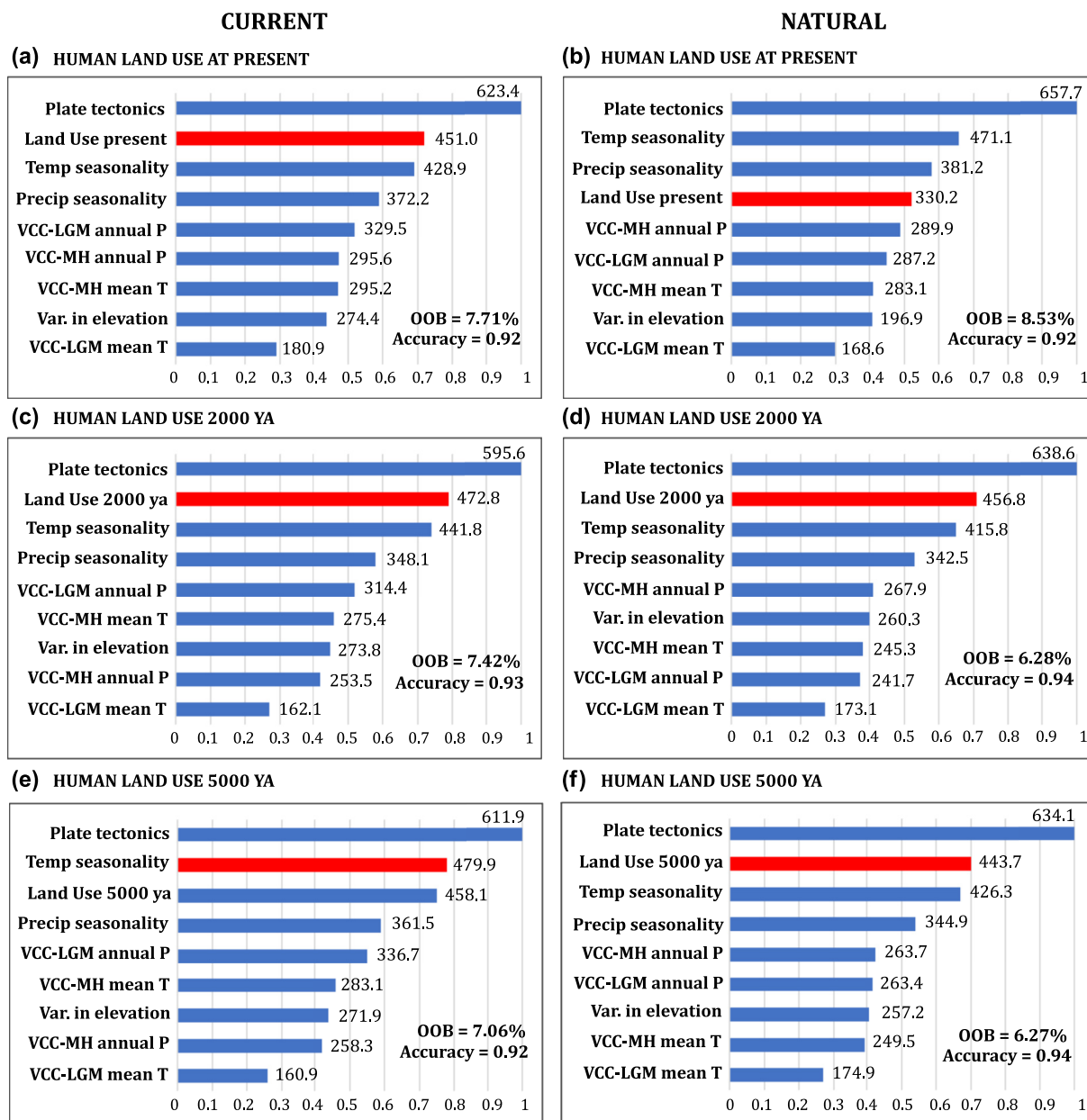


Figure 4. Ranking of importance values for the drivers of taxonomic differentiation for the current and natural medium-sized bioregions or subregions. Above panels show importance values from models with human land use at present for (a) current and (b) natural bioregions, while the panels below show importance values from models with human land use 2000 years ago for (c) current and (d) natural bioregions, and human land use 5000 years ago for (e) current and (f) natural bioregions. Specifications as in Fig. 2.

use for both current and natural medium-sized bioregions (Supporting information).

While our results show perceptible impacts of past human land use on current biogeographical patterns for terrestrial mammals, there are some limitations in the available data that could potentially bias our results. First, IUCN distribution ranges do not represent true areas of habitat, meaning that there may be a fair amount of unsuitable habitat in the ranges resulting in errors of commission (Brooks et al. 2019). Yet, our approach (assigning presences to grid cells) does not assume a given species is found in all areas of a cell, but rather

that it could be found in some areas, which for cells of the size considered here ( $96.5 \times 96.5$  km) should greatly reduce commission errors. Second, IUCN distribution maps are the most comprehensive global source of range data for terrestrial mammals; however, we are aware that sampling effort, that influences range definition, varies across the world. Unfortunately, that spatial bias is also present in other available sources, like GBIF, in which nation-wide differences in funding and data sharing have led to huge differences in data contributions (Yesson et al. 2007, Beck et al. 2014). Third, while in PHYLACINE (Faurby et al. 2018) it is claimed

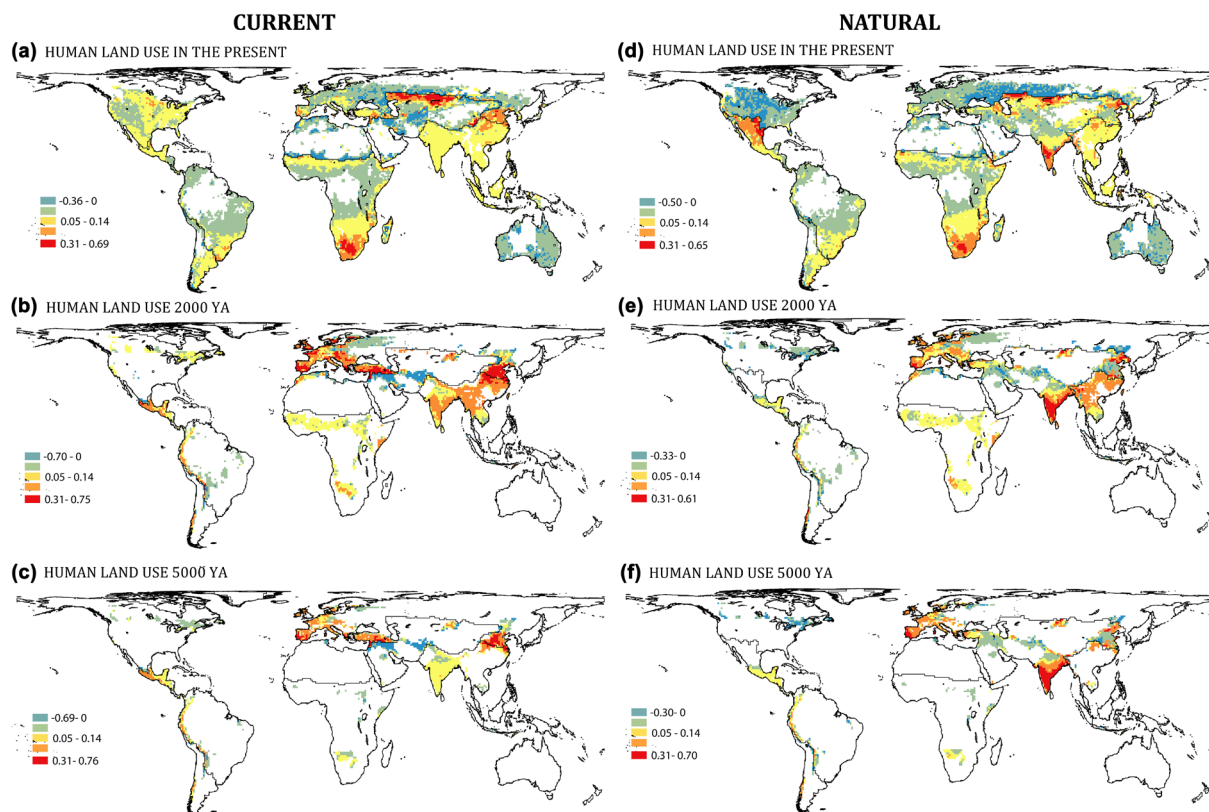


Figure 5. Maps of the local importance of present and past human land use for the current and natural broader bioregions when percentage of human land use was  $> 10\%$  in each grid cell. The local importance shows the importance of the variable (human land use here) for the classification of a single sample (grid cells). The score of the legends shows the impact on correct classification of grid cells, with blue colors indicating from negative to 0 values (incorrect to neutral classification), and yellow, orange and red indicating positive values (correct classification). In these maps, white cells imply human land use values  $< 10\%$ .

that ‘both current ranges and present natural ranges can be compared to estimate macroecological relationships with and without anthropogenic pressures’, there is uncertainty in the reconstructions of present natural ranges ranging from high certainty for species with recent, documented human-induced range expansions or reductions to more uncertain inferences based on fossil co-occurrences. Uncertainty may be higher among extinct megafauna which could drive the configuration of natural bioregions. Importantly, our study does not claim that the delineated natural bioregions are the real bioregions that should exist in the absence of human impacts, instead this is a counterfactual scenario for comparison.

Finally, we acknowledge that although our work is framed in two specific moments (5000 and 2000 years ago), the effects shown do not necessarily have to have occurred 5000 years ago, but possibly earlier. We have not been able to demonstrate this point since information available for 8000 years ago has low variability at global level. Existing models of long-term changes in global land use, as the  $KK_{10}$  model used here, differ in their representation of these early transformations as they are dependent on limited disparate empirical data from archaeology and paleoecology (Stephens et al. 2019). Efforts to map land-cover change over the past 10 000 years using pollen data have recently increased (Stephens et al. 2019) and

possibly in the next few years we will already be able to use high-quality global reconstructions combining both land-use and land-cover change that will help us to understand better the impact of early civilizations on the present-day biodiversity patterns.

In summary, our results show the value of considering impacts of past human actions to understand the current organisation of biodiversity globally. Previous studies have documented lasting effects of human land use changes during the last millennia on current biodiversity patterns (Dambrine et al. 2007, Haberle 2007, van der Sande et al. 2019), but this is the first time the signal has been recognized on the taxonomic differentiation of the largest realms. Some of the biogeographical boundaries proposed by Wallace 140 years ago (Wallace 1876) were likely already showing the influence of anthropogenic actions. The human transformation of ecosystems that occurred in large terrestrial extensions coinciding with the distribution of ancient civilizations, makes past human land use important to discern taxonomic differences among the large biogeographical zones of the world. It is certainly difficult to disentangle direct causal effects from correlative relationships in non-experimental large-scale studies such as this. We know the same geographic and environmental characteristics that influenced

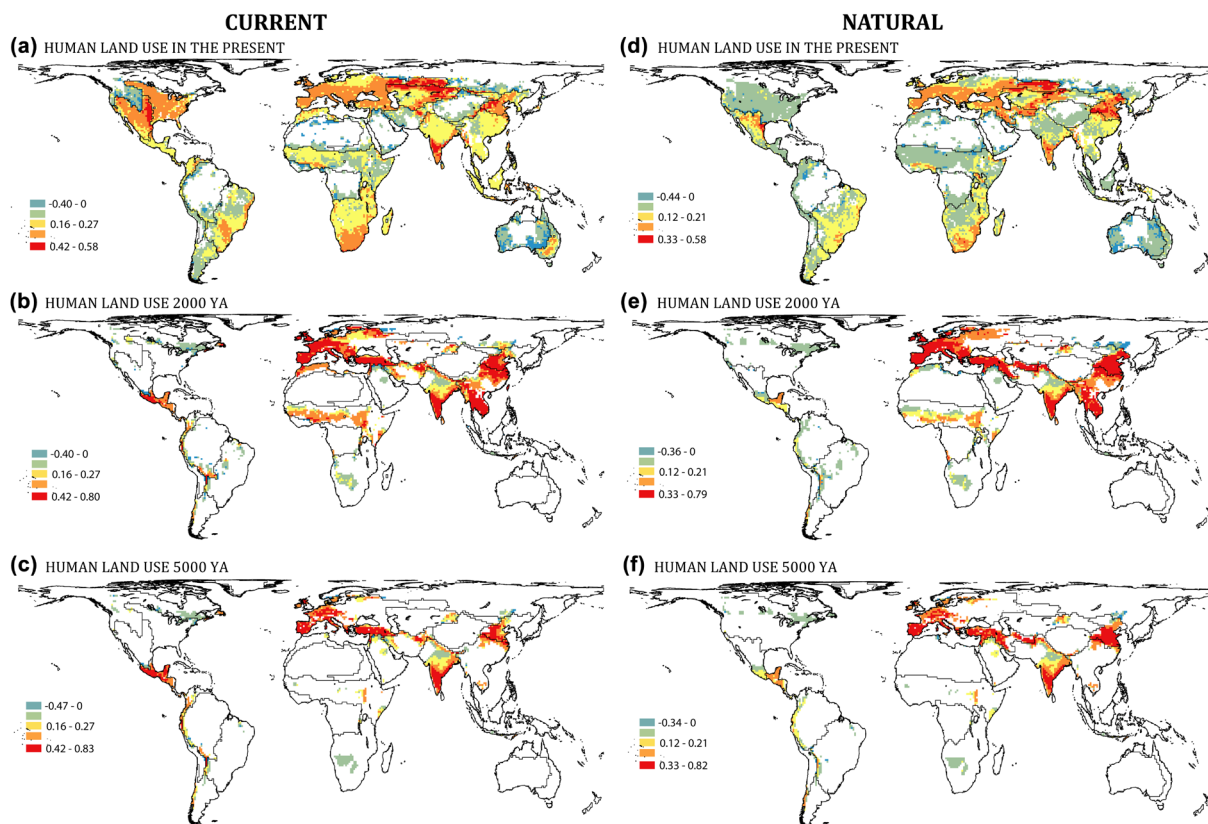


Figure 6. Maps of the local importance of present and past human land use for the medium current and natural bioregions when percentage of human land use was > 10% in each grid cell. Specifications as in Fig. 5.

the distribution across space and time of mammals likely also influenced the distribution of our ancestors (Lomolino, 2018). However, we considered several of those predictors and still found a global imprint of past human actions on the patterns of distribution of global diversity. Our results force us to reflect on how the much more widespread and severe changes that have occurred since the beginning of the industrial revolution will affect the organisation of biodiversity in the future. We show here the effects of recent human actions are already detectable at the subregional scale. Although it may seem like an unlikely or distant future, the possibility that human activities become the main driver that shapes Earth's biodiversity, above historical geomorphological or climatic processes, cannot be ignored.

*Acknowledgements* – Authors thank the two anonymous reviewers whose comments improved the manuscript significantly.

*Funding* – MR was funded by the project 2020/125-US/JUNTA/FEDER EU (Programa Operativo FEDER/Junta de Andalucía 2014-2020), and ER was supported by project CGL2017-83045-R AEI/FEDER EU (Agencia Estatal de Investigación from the Ministry of Economy, Industry and Competitiveness, Spain co-financed with FEDER).

### Author contributions

**Marta Rueda:** Conceptualization (equal); Data curation (lead); Formal analysis (lead); Funding acquisition (lead);

Investigation (equal); Methodology (lead); Writing – original draft (equal); Writing – review and editing (equal). **Manuela González-Suárez:** Conceptualization (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review and editing (equal). **Eloy Revilla:** Conceptualization (equal); Writing – original draft (equal); Writing – review and editing (equal).

### Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06762>.

### Data availability statement

The data supporting the results are all from publicly accessible sources. Data for current and natural distribution ranges of terrestrial mammals are from PHYLACINE 1.2 (<https://doi.org/10.5061/dryad.bp26v20>); climate data are from the WorldClim v.2 dataset (<https://worldclim.org/data/worldclim21.html>); elevation data are from GTOPO30 ([https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30?qt-science\\_center\\_objects=0#qt-science\\_center\\_objects](https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-30-arc-second-elevation-gtopo30?qt-science_center_objects=0#qt-science_center_objects)); plate tectonics data are from [http://peterbird.name/publications/2003\\_PB2002/2003\\_PB2002.htm](http://peterbird.name/publications/2003_PB2002/2003_PB2002.htm); and data from past human land use are from <https://ecotope.org/products/>

[datasets/used\\_planet](#). Sources are all described clearly in the manuscript.

Data are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.cfxpvnxcn> (Rueda et al. 2023).

## Supporting information

The Supporting information associated with this article is available with the online version.

## References

- Araújo, M. B., Nogués-Bravo, D., Haywood, A. M., Valdes, P. J. and Rahbek, C. 2008. Quaternary climate changes explain diversity among reptiles and amphibians. – *Ecography* 31: 8–15.
- Barnosky, A. D., Koch, P. L., Feranec, R. S., Wing, S. L. and Shabel, A. B. 2004. Assessing the causes of Late Pleistocene extinctions on the continents. – *Science* 306: 70–75.
- Beck, J., Böller, M., Erhardt, A. and Schwanghart, W. 2014. Spatial bias in the GBIF database and its effect on modelling species' geographic distributions. – *Ecol. Infor* 19: 10–15.
- Benito, B. M. 2021. R package spatialRF: easy spatial regression with Random Forest. – <https://blasbenito.github.io/spatialRF>.
- Bernardo-Madrid, R., Calatayud, J., González-Suárez, M., Rosvall, M., Lucas, P. M., Rueda, M., Antonelli, A. and Revilla, E. 2019. Human activity is altering the world's zoogeographical regions. – *Ecol. Lett.* 22: 1297–1305.
- Bird, P. 2003. An updated digital model of plate boundaries. – *Geochem. Geophys. Geosyst.* 4: 1027.
- Bodenhofer, U., Kothmeier, A. and Hochreiter, S. 2011. APCluster: a R package for affinity propagation clustering. – *Bioinformatics* 27: 2463–2464.
- Boivin, N. L., Zeder, M. A., Fuller, D. Q., Crowther, A., Larson, G., Erlandson, J. M., Denham, T. and Petraglia, M. D. 2016. Ecological consequences of human niche construction: examining long-term anthropogenic shaping of global species distributions. – *Proc. Natl Acad. Sci. USA* 113: 6388–6296.
- Breiman, L. 2001. Random forests. – *Mach. Learn.* 45: 5–32.
- Brooks, T. M., Pimm, S. L., Akçakaya, H. R., Buchanan, G. M., Butchart, S. H. M., Foden, W., Hilton-Taylor, C., Hoffmann, M., Jenkins, C. N., Joppa, L., Li, B. V., Menon, V., Ocampo-Peñuela, N. and Rondinini, C. 2019. Measuring terrestrial area of habitat (AOH) and its utility for the IUCN Red List. – *Trends Ecol. Evol.* 34: 977–986.
- Butzer, K. W. 2005. Environmental history in the Mediterranean world: cross-disciplinary investigation of cause-and-effect for degradation and soil erosion. – *J. Archaeol. Sci.* 32: 1773–1800.
- Capinha, C., Essl, F., Seebens, H., Moser, D. and Pereira, H. M. 2015. The dispersal of alien species redefines biogeography in the Anthropocene. – *Science* 348: 1248–1251.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y. and Yuan, J. 2023. xgboost: extreme gradient boosting. – R package, ver. 1.7.5.1, <https://CRAN.R-project.org/package=xgboost>.
- Clifford, P., Richardson, S. and Hémon, D. 1989. Assessing the significance of the correlation between two spatial processes. – *Biometrics* 45: 123–134.
- Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J. and Lawler, J. J. 2007. Random forests for classification in ecology. – *Ecology* 88: 2783–2792.
- Dambrine, E., Dupouey, J. L., Laüt, L., Humbert, L., Thion, M., Beaufigl, T. and Richard, H. 2007. Present forest biodiversity patterns in France related to former Roman agriculture. – *Ecology* 88: 1430–1439.
- Daru, B. H., Elliott, T. L., Park, D. S. and Davies, T. J. 2017. Understanding the processes underpinning patterns of phylogenetic regionalization. – *Trends Ecol. Evol.* 32: 845–860.
- Dinerstein, E. et al. 2017. An ecoregion-based approach to protecting half the terrestrial realm. – *BioScience* 67: 534–545.
- Dray, S., Legendre, P. and Peres-Neto, P. R. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). – *Ecol. Modell.* 196: 483–493.
- Dunning, N. P., Beach, T. P. and Luzzadder-Beach, S. 2012. Kax and kol: collapse and resilience in lowland Maya civilization. – *Proc. Natl Acad. Sci. USA* 109: 3652–3657.
- Dutilleul, P., Clifford, P., Richardson, S. and Hemon, D. 1993. Modifying the t test for assessing the correlation between two spatial processes. – *Biometrics* 49: 305–314.
- Ellis, E. C., Kaplan, J. O., Fuller, D. Q., Vavrus, S., Goldewijk, K. K. and Verburg, P. H. 2013. Used planet: a global history. – *Proc. Natl Acad. Sci. USA* 110: 7978–7985.
- Ellis, E. C., Gauthier, N., Klein Goldewijk, K., Bliege Bird, R., Díaz, S., Fuller, D. Q., Gill, J. L., Kaplan, J. O., Kingston, N., Locke, H., McMichael, C. N. H., Ranco, D., Rick, T. C., Shaw, M. R., Stephens, L., Svenning, J. C. and Watson, J. E. M. 2021. People have shaped most of terrestrial nature for at least 12 000 years. – *Proc. Natl Acad. Sci. USA* 118: e2023483118.
- Faurby, S. and Svenning, J.-C. 2015. Historic and prehistoric human-driven extinctions have reshaped global mammal diversity patterns. – *Divers. Distrib.* 21: 1155–1166.
- Faurby, S., Davis, M., Pedersen, R. O., Schowaneck, S. D., Antonelli, A. and Svenning, J. C. 2018. PHYLACINE 1.2: the phylogenetic atlas of mammal macroecology. – *Ecology* 99: 2626.
- Ficetola, G. F., Mazel, F. and Thuiller, W. 2017. Global determinants of zoogeographical boundaries. – *Nat. Ecol. Evol.* 1: 0089.
- Fick, S. E. and Hijmans, R. J. 2017. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. – *Int. J. Climatol.* 37: 4302–4315.
- Frey, B. J. and Dueck, D. 2007. Clustering by passing messages between data points. – *Science* 315: 972–976.
- Fuller, D. Q., van Erten, J., Manning, K., Castillo, C., Kingwell-Banham, E., Weisskopf, A., Qin, L., Sato, Y.-I. and Hijmans, R. J. 2011. The contribution of rice agriculture and livestock pastoralism to prehistoric methane levels. An archaeological assessment. – *Holocene* 21: 743–759.
- Gagné, S. and Proulx, R. 2009. Accurate delineation of biogeographical regions depends on the use of an appropriate distance measure. – *J. Biogeogr.* 36: 561–562.
- Haberle, S. G. 2007. Prehistoric human impact on rainforest biodiversity in highland New Guinea. – *Philos. Trans. R. Soc. B* 362: 219–228.
- Hansford, J., Wright, P. C., Rasoamiamanana, A., Pérez, V. R., Godfrey, L. R., Errickson, D., Thompson, T. and Turvey, S. T. 2018. Early Holocene human presence in Madagascar evidenced by exploitation of avian megafauna. – *Sci. Adv.* 4: eaat6925.
- Holt, B. G., Lessard, J. P., Borregaard, K., Fritz, S. A., Araújo, M. B., Dimitrov, D., Fabre, P. H., Graham, C. H., Graves, G. R., Jonsson, K. A., Nogués-Bravo, D., Wang, Z., Whittaker, R. J., Fjeldsä, J. and Rahbek, C. 2013. An update of Wallace's zoogeographic regions of the world. – *Science* 339: 74–78.

- IUCN. 2016. IUCN Red List of threatened species, ver. 2016-3. – <http://www.iucnredlist.org>.
- Kaplan, J. O., Krumhardt, K. M., Ellis, E. C., Ruddiman, W. F., Lemmen, C. and Goldewijk, K. K. 2011. Holocene carbon emissions as a result of anthropogenic land cover change. – *Holocene* 21: 775–791.
- Kirch, P. V. 2005. Archaeology and global change: the Holocene record. – *Annu. Rev. Environ. Resour.* 30: 409–440.
- Koch, P. L. and Barnosky, A. D. 2006. Late Quaternary extinctions: state of the debate. – *Annu. Rev. Ecol. Evol. Syst.* 37: 215–250.
- Kreft, H. and Jetz, W. A. 2010. A framework for delineating biogeographical regions based on species distributions. – *J. Biogeogr.* 37: 2029–2053.
- Legendre, P. and Gallagher, E. D. 2001. Ecological meaningful transformations for ordination of species data. – *Oecologia* 129: 271–280.
- Legendre, P. and Legendre, L. 1998. Numerical ecology. – Elsevier.
- Lewis, S. L. and Maslin, M. A. 2015. Defining the Anthropocene. – *Nature* 519: 171–180.
- Liaw, A. and Wiener, M. 2002. Classification and regression by RandomForest. – *R News* 2: 18–22.
- Lomolino, M. V. 2018. On teaching “...that grand subject, ...”. – *Front. Biogeogr.* 10: e37812.
- Lomolino, M. V., Riddle, B. R., Whittaker, R. J. and Brown, J. H. 2010. Biogeography. – Sinauer Associates.
- Lyons, S. K. et al. 2016. Holocene shifts in the assembly of plant and animal communities implicate human impacts. – *Nature* 529: 80–85.
- Mazel, F., Wüest, R. O., Lessard, J.-P., Renaud, J., Ficetola, G. F., Lavergne, S. and Thuiller, W. 2017. Global patterns of  $\beta$ -diversity along the phylogenetic time-scale: the role of climate and plate tectonics. – *Global Ecol. Biogeogr.* 26: 1211–1221.
- Miller, G. H., Fogel, M. L., Magee, J. W., Gagan, M. K., Clarke, S. J. and Johnson, B. J. 2005. Ecosystem collapse in Pleistocene Australia and a human role in megafaunal extinction. – *Science* 309: 287–290.
- Mottl, O., Flantua, S. G. A., Bhatta, K. P., Felde, V. A., Giesecke, T., Goring, S., Grimm, E. C., Haberle, S., Hooghiemstra, H., Ivory, S., Kuneš, P., Wolters, S., Seddon, A. W. R. and Williams, J. W. 2021. Global acceleration in rates of vegetation change over the past 18 000 years. – *Science* 372: 860–864.
- Nicodemus, K. K., Malley, J. D., Strobl, C. and Ziegler, A. 2010. The behaviour of random forest permutation-based variable importance measures under predictor correlation. – *BMC Bioinform.* 11: 110.
- Polaina, E., González-Suárez, M. and Revilla, E. 2020. The legacy of past human land use in current patterns of mammal distribution. – *Ecography* 42: 1623–1635.
- Procheş, Ş. and Ramdhani, S. 2012. The world’s zoogeographical regions confirmed by cross-taxon analyses. – *BioScience* 62: 260–270.
- Ramey, J. A. 2012. Clusteval: evaluation of clustering algorithms. – R package ver. 0.1, <https://CRAN-R-project.org/package=clusteval>.
- Richard, E. W. A. 1997. Ancient civilizations of the new world: essays in world history. – Westview Press.
- Ricklefs, R. E. 2004. A comprehensive framework for global patterns in biodiversity. – *Ecol. Lett.* 7: 1–15.
- Riddle, B. R. and Hafner, D. J. 2010. Integrating pattern with process at biogeographic boundaries: the legacy of Wallace. – *Ecography* 33: 321–325.
- Rueda, M., Rodríguez, M. Á. and Hawkins, B. A. 2013. Identifying global zoogeographical regions: lessons from Wallace. – *J. Biogeogr.* 40: 2215–2225.
- Rueda, M., González-Suárez, M. and Revilla, E. 2023. Data from: Global biogeographical regions reveal a signal of past human impacts. – Dryad Digital Repository, <https://doi.org/10.5061/dryad.cfxpvnxcn>.
- Sandel, B., Dalsgaard, B., Davies, R. G., Gaston, K. J., Sutherland, W. J. and Svenning, J. C. 2011. The influence of late Quaternary climate-change velocity on species endemism. – *Science* 334: 660–664.
- Sandom, C., Faurby, S., Sandel, B. and Svenning, J. C. 2014. Global late Quaternary megafauna extinctions linked to humans, not climate change. – *Proc. R. Soc. B* 281: 20133254.
- Sexton, J. P., McIntyre, P. J., Angert, A. L. and Rice, K. J. 2009. Evolution and ecology of species range limits. – *Annu. Rev. Ecol. Evol.* 40: 415–436.
- Silva, F., Stevens, C. J., Weisskopf, A., Castillo, C., Qin, L., Bevan, A. and Fuller, D. Q. 2015. Modelling the geographical origin of rice cultivation in Asia using the rice archaeological database. – *PLoS One* 10: e0137024.
- Smith, F., Smith, R. E. E., Lyons, S. K. and Payne, J. 2018. Body size downgrading of mammals over the Late Quaternary. – *Science* 360: 310–313.
- Stephens, L. et al. 2019. Archaeological assessment reveals Earth’s early transition through land use. – *Science* 365: 897–902.
- Touw, W. G., Bayjanov, J. R., Overmars, L., Backus, L., Boekhorst, J., Wels, M. and van Hijum, S. A. F. T. 2012. Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? – *Brief. Bioinform.* 14: 315–326.
- Vallejos, R., Osorio, F. and Bevilacqua, M. 2020. Spatial relationships between two georeferenced variables: with applications in R. – Springer.
- van der Sande, M. T., Gosling, W., Correa-Metrio, A., Prado-Junior, J., Poorter, L., Oliveira, R. S., Mazzei, L. and Bush, M. B. 2019. A 7000-year history of changing plant trait composition in an Amazonian landscape; the role of humans and climate. – *Ecol. Lett.* 22: 925–935.
- von Humboldt, A. 1806. *Essai sur la Géographie des plantes; accompagné d’un tableau physique des régions équinoxiales, accompagné d’un tableau physique des régions équinoctiales.* – Schoel & Co.
- Wallace, A. R. 1876. *The geographical distribution of animals.* – Harper & Brothers.
- Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S. and Kawamiya, M. 2011. Miroc-ESM 2010: model description and basic results of CMIP5-20c3m experiments. – *Geosci. Model Dev.* 4: 845–872.
- Wiens, J. J. 2011. The niche, biogeography and species interactions. – *Philos. Trans. R. Soc. B* 366: 2336–2350.
- Yesson, C., Brewer, P. W., Sutton, T., Caithness, N., Pahwa, J. S., Burgess, M., Gray, W. A., White, R. J., Jones, A. C., Bisby, F. A. and Culham, A. 2007. How global is the global biodiversity information facility? – *PLoS One* 2: e1124.
- Zheng, Z., Ma, T., Roberts, P., Li, Z., Yue, Y., Peng, H., Huang, K., Han, Z., Wan, Q., Zhang, Y., Zhang, X., Zheng, Y. and Satio, Y. 2021. Anthropogenic impacts on Late Holocene land-cover change and floristic biodiversity loss in tropical southeastern Asia. – *Proc. Natl Acad. Sci. USA* 118: e2022210118.