Science
AAAS

Supplementary Materials for

Global ecosystem thresholds driven by aridity

Miguel Berdugo, Manuel Delgado-Baquerizo, Santiago Soliveres, Rocío Hernández-Clemente, Yanchuang Zhao, Juan J. Gaitán, Nicolas Gross, Hugo Saiz, Vincent Maire, Anika Lehman, Matthias C. Rillig, Ricard V. Solé and Fernando T. Maestre

Correspondence to: mglberdugo@gmail.com

This PDF file includes:

Materials and Methods

Data collection

We used multiple global data sources (interpolated and remote sensing data, standardized field surveys and individual plant measurements obtained in the laboratory), which are described in detail below. The databases chosen comprise the widest observational database on global drylands to date (BIOCOM), complemented with observations from physiological functional traits (*20*, *31*) which are not well represented in BIOCOM. Also we retrieved map information to complement these variables (albedo, NDVI, VSI, shrubland occurrence and coefficient of variation of interannual precipitation) or to complete key information available in BIOCOM but that missed important dryland areas of the world (Vegetation Cover, SOC, soil nitrogen and sand content). The variables were selected based on their importance as surrogates of key ecosystem attributes/processes such as nutrient cycling, plant productivity, important biotic interactions, biodiversity and climatic patterns. Together, the variables selected characterize three major ecosystem components/processes (soils, plants and plant-soil interactions) that largely determine ecosystem functioning and its capacity to provide essential ecosystem services such as soil fertility and plant productivity. The variables selected are also often used to monitor soil health and land degradation (6, *32*, *33*), and are thus of particular relevance to scientists and stakeholders alike.

Interpolated and remote sensing data

We gathered interpolated and remote sensing information from freely available global maps (see list below and Table S1). Measurements from these maps were obtained by taking data from a point every 30 arc-minutes. This yielded a total of 500,000 sampling points, from which we discarded all of those not falling in terrestrial ecosystems classified as drylands (i.e., those scoring aridity index = Annual precipitation/annual potential evapotranspiration lower than 0.65 *34*). The aridity index has been proposed by FAO (*34*) and is widely used both in the current scientific literature (*8*, *35*, *36*) and in management/policy activities related with desertification and land use management (*6*). This metric is also commonly used to frame both the current (*2*) and future (*3*, *37*) extension of global drylands. The aridity index was obtained from the global maps of Zomer et al (*38*), which provides the averaged aridity index of the period 1970-2000, and has a spatial resolution of 30 arc-seconds. Apart from sites not classified as drylands, we also removed those points that had ever been classified as urban, cultivated lands or water bodies in the land use and land cover maps provided by FAO (*39*, also with a spatial resolution of 30 arc-seconds). We did so to avoid outliers coming from agricultural or urban lands, as changes in the variables recorded in these areas are more likely to be driven by the direct impact of human activities than by changes in aridity. A total of 51,013 points remained for our analyses; these encompass a regularly spaced grid of sites summarizing all possible environmental situations found in global drylands. From these points, we extracted the following information:

Soil variables- We extracted the following soil variables from ref. *40*: carbon content, which is strongly related to soil fertility and water retention (*41*), nitrogen content and silt and clay content. These measurements are extrapolations from soil classifications and are corrected by environmental and other local conditions (see ref. *40* for details).

Plant productivity- We used the Normalized Difference Vegetation Index (NDVI) as a surrogate of plant productivity. NDVI is an index that combines the spectral reflectance measurements acquired in the red (visible) and near-infrared bands, and provides a global measure of the "greenness" of vegetation across the Earth's landscapes (*42*, *43*). It is related to vegetation productivity because it indicates the photosynthetically active radiation absorbed by plant canopies (*44*, *45*). NDVI data for each plot were acquired from two sources: i) the MOD13Q1 product from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor (*46*), which provides 23 images per year (every 16 days) with a pixel size of 250 m × 250 m (https://modis.ornl.gov/data.html) and ii) the Landsat 5 TM and Landsat 7 ETM+, with a pixel size of 30 m × 30 m (https://landsat.gsfc.nasa.gov/). Landsat imagery was processed to surface reflectance using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) atmospheric correction algorithm (*47*). This algorithm is designed to work with Landsat-5 to 8 products using the spectral response functions developed for the Terra MODIS instrument (*48*). MODIS data are geometrically and atmospherically corrected and include a reliability index of data quality based on the environmental conditions in which the data were recorded and ranging from 0 (good quality data) to 4 (raw or absent data) (*49*). At each point, we calculated the mean annual NDVI for each year between 2000 and 2015 and then averaged for the entire period; for that, we discarded the data with MODIS reliability index >1. NDVI estimates from MODIS and LANDSAT were very similar (r = 0.85). Hence, we report results from MODIS because data from this satellite matches more closely the spatial resolution of the aridity interpolations used.

Vegetation fractional cover- NDVI may saturate at very low vegetation fractional cover due to the existence of high solar reflectance in the soil (*42*). Thus, its signal is less reliable in zones with very low vegetation cover. For these cases, MODIS provides a product of vegetation fractional cover estimation (MOD44B, *50*) by making corrections using a machine-learning process with proven small error (*50*). We used MOD44B to extract vegetation fractional cover estimations (tree + non-tree vegetation cover), and further cross-validated it with other remote sensing product, the Global Bare Ground circa 10 (*51*, accessed from www.landcover.usgs.gov/glc). This database, which provides better spatial resolution using LANDSAT 7 ETM+ data, yields very similar results to those provided by MOD44B (r = 0.94). Again, and for the same reasons noted above, we kept MOD44B extractions as measurements of the fractional cover of vegetation. It is worth noting that issues related to the ability of MOD44B to detect trees with less than 5 m (which has raised some recent discussion; *52*, *53*) do not affect our results because we used the sum of tree + non-tree cover layers.

Albedo - Albedo refers to the sunlight reflection to the atmosphere after reaching Earth surface. It is related to soil typology, vegetation cover and topography (*54–56*). It correlates with ecosystem functioning in drylands (*57*) and has important feedbacks on climate (*30*). The actual land surface albedo may also be influenced by atmospheric conditions, the season of the year and the zenithal position of the sun (*58*). To avoid problems with the atmosphere and solar zenithal position, we used white sky-albedo (WSA). It occurs in the case of completely diffuse and isotropic illumination, i.e., when light from all directions is the same. We extracted WSA in the shortwave spectral domain (i.e., 0.3-5 µm) from MODIS MCD43B3 BRDF/Albedo Model Parameters Product (Collection 5) (*59*). This is calculated by using the RossThick-LiSparse-Reciprocal (Ross-Li) BRDF model, and is independent of the atmosphere and solar zenithal position (*54*). To avoid problems related to seasonal differences in albedo, we calculated WSA from May to September in the northern hemisphere, and from November to March for the southern hemisphere. We averaged values on a yearly basis for the period 2000-2015.

Land use type (shrublands)- We extracted vegetation types using the maps provided by FAO (*39*), which were also used to remove agricultural and urban landscapes in a previous step. We did so to classify each site regarding the dominant plant form (e.g., forest, grassland, shrubland, savanna). As explained above, FAO maps provide detailed information on vegetation types from 2000 to present. When changes in the dominant vegetation had occurred in time, we kept the most representative (i.e., the land type recorded during most years for the period 2000-2015) for each site. In particular, we were interested on the occurrence of shrublands, which have been associated with processes of encroachment with increasing aridity (*24*). Thus, we created a binary variable with values 1 (when FAO maps classified a given site as open or dense shrubland) or 0 (when a site was covered with any other vegetation type). Desert states, which are devoid of perennial vegetation, were excluded for the analysis of this variable.

Vegetation sensitivity index (VSI)- This index is extracted from Seddon et al. (*25*), and measures the sensitivity of vegetation concerning climatic fluctuations. For doing so, VSI quantifies the ratio of change of enhanced vegetation index (EVI, a vegetation index very similar to NDVI, *60*) divided by precipitation and temperature changes within a year (then averaged for the period 2000-2013). This index is an important component of vegetation resilience (ability to remain undisturbed with climatic fluctuations), and thus summarizes an important dynamical component of the ecosystem by assessing the relative response rate of vegetation to environmental variability. Values of VSI had a spatial resolution of 5 km (around 2.5 arc minutes).

Inter-annual precipitation variability- As rainfall is extremely variable in drylands, and it is expected to become even more variable with climate change (*41*, *61*, *62*), we extracted precipitation variability as an important metric of the unpredictable climatic conditions faced by these ecosystems (*63*). We calculated annual precipitation for each year and then derived the coefficient of variation (standard deviation/mean) among years (period 2000-2015) using data derived from the Climatologies at High resolution for the Earth's Land Surface Areas (*64*) at 30 arc-secs spatial resolution. The CV of rainfall is widely used to obtain estimates of precipitation inter-annual variability and has certain values (~ 30%) considered critical for the equilibrium of grazing areas (*65*, *66*), which are key for the livelihoods of dryland populations in developing countries (*67*, *68*). It has been suggested that areas with values of this CV higher than 30% are non-equilibrium ecosystems, in which the carrying capacity of the ecosystem to grazing fluctuates in a way such that herbivores need to accommodate their population each year due to these fluctuations (*65*, *66*).

Standardized field surveys

We used data from a standardized field survey conducted at 236 dryland ecosystems worldwide (21 countries from six continents, *14*, *69*). These sites encompass all dryland subtypes except hyper-arid, and span a wide range of environmental conditions and vegetation and soil types (*14*, *69*). In brief, vegetation was sampled *in situ* at each site using four 30-m long transects interspaced 8 m from each other. Along these transects, adjacent quadrats of 1.5 by 1.5 m were located, and the relative cover of each species on each quadrat was measured (80 quadrats per site were surveyed). Also, five composite soil samples (0-7.5 cm depth) beneath the dominant plants and other five in bare ground areas devoid of vascular vegetation were collected. These soil samples were dried and shipped to the laboratory of Rey Juan Carlos University (Móstoles, Spain) for analyses. See refs. *69* and *14* for additional details on the field survey and the handling of soil samples. We completed the data from this global survey with data from independent

sources obtained using comparable survey methodologies. For instance, microbial abundance and fungal functional types were expanded adding sites from Australia (see unpublished data in ref: *70*). Soil carbon data were expanded by adding plots from Australia (*71*) and North-America (*72*). In all cases, we used data collected with standardized field and laboratory protocols to ensure their comparability (*73*). Overall, the variables from field surveys included in this study are the following:

Soil organic carbon (N = 627)- Measured by colorimetry after oxidation with a mixture of potassium dichromate and sulfuric acid (*74*) using samples coming from bare ground areas devoid of vascular vegetation.

Vegetation effect on soils (N = 204)- Vegetation can modify its environment at local scales via its effects on soils (*24*, *75–77*). Indeed, soils under plant canopies in drylands typically show increased nutrient contents and biological activity compared to adjacent bare ground areas (the so-called "fertility island" effect, *76*, *78*). We assessed this effect by calculating the differences in soil organic carbon under plant canopies and adjacent bare ground areas in each of the sites surveyed. Thus, this metric measures the net soil organic carbon gain in microsites covered with vegetation regarding bare ground microsites.

Plant-plant interactions (N = 185)- The use of spatial patterns as a proxy for biotic interactions has been widely used in plant ecology, particularly in drylands where biotic interactions produce a strong spatial signal (*79–81*). We approximated the type (positive-negative) and dominance of plant-plant interactions using spatial network analyses of the vegetation surveyed as described in Saiz et al. (*82*). In short, we assessed the mean link weight distribution in the community using co-occurrence matrices of the species for each site. This index represents the dominant type of spatial pattern in the community, with values > 0 and < 0 indicating spatial aggregation and segregation, respectively. To account for the effect of species abundance distribution, for each site the index was calculated as the difference between the observed value and the expected value calculated from 2000 networks generated by randomizing the spatial distribution of abundances at the site. The occurrence of species and their relative abundances were extracted from the 1.5 x 1.5 m quadrat samples surveyed in the field.

Stability of soil aggregates (N = 220)- Soil aggregation is a key structural parameter with important implications for soil fertility and erosion resistance (*83*). We measured the size distribution of soil aggregates, expressed as their mean weight diameter, as a proxy of this feature. We did so by using a modified protocol by Kemper and Rosenau (*84*). Briefly, the pre-sieved samples (< 2 mm) were passed through a stack of four sieves (1, 0.5, 0.212 and 0.053 mm) hence separating the soil into five fractions of decreasing particle size. To avoid abrasion, we vertically moved the sieve stack four times to allow the soil aggregates to separate through the different mesh sizes. Weights of the different fractions were calculated as the mean diameter of each fractions weighted by the proportion of total soil mass of the size fraction. By doing so the mean weight diameter (in mm) increases with increasing amounts of large aggregates in a sample.

Soil microbial communities (N = 100)- Soil DNA was extracted from 0.25 g of defrosted soil samples using the Powersoil® DNA Isolation Kit (Mo Bio Laboratories, Carlsbad, CA, USA). The total abundance of bacteria and fungi was quantified using qPCR and the Eub 338-Eub 518 and ITS 1-5.8S primer sets, as described in ref (*85*). To characterize the relative abundance of fungal functional groups, amplicon sequencing was conducted using Illumina MiSeq platform (*86*) and the FITS7/ITS4 primer set (*87*, *88*). Bioinformatic analyses were conducted using UPARSE (*89*) and MOTHUR (*90*). Operational Taxonomic Units (OTU) were picked at 97%

sequence similarity. Taxonomy was assigned to OTUs in MOTHUR using the UNITE version 6 dataset (*91*). The resulting OTU abundance tables were rarefied. We identified these fungal functional groups (e.g., fungal pathogens, saprobes, mycorrhizal fungi) using FUNGUILD (*92*). We calculated the relative abundance of each functional group as the sum of the relative abundance of taxa within each fungal group, as described in ref. *93*.

Plant species richness (N = 236)- Plant species richness is of paramount importance to ecosystem functioning (*69*) and stability (*94*) in drylands. We estimated this variable for each site as the total number of plant species found in the 80 1.5 x 1.5 m quadrat samples surveyed.

Plant functional traits at the community level (N = 119) - We assessed the specific leaf area (SLA) of perennial vascular plants, calculated as the ratio between the weight of a leaf and its area, by extracting these data from global datasets and local floras (see more details in ref. *29*). This is a trait summarizing the resource economic spectrum of plant leaves, i.e., opposing leaves with high metabolic activity, weak stress-tolerance and low lifespan to the opposite syndrome (*95*, *96*). In particular, SLA tends to decrease with increasing aridity in drylands as smaller and tougher leaves endure better water scarcity conditions (*20*, *97*, *98*). Under extreme aridity conditions, plants may also increase SLA by producing lighter leaves, which are not good to endure water scarcity periods but that are easy and less costly to produce, thus can be fast growing during the periods of favorable photosynthetic activity (after sporadic rains, *21*, *29*, *99*). These leaves might be discarded afterwards, thus configuring a strategy of stress avoidance. Using the recorded abundance of each species, we estimated the community-weighted mean of SLA that reflects the most dominant plant functional strategy within communities (*100*).

We completed our information on plant functional traits using photosynthetic rates from the database created by Maire et al. (*20*) and Wang et al. (*31*). Both datasets provide a comprehensive (more than 10,000 individual records of 809 plant species) and standardized measurement of the photosynthetic activity and leaf N content of vascular plants from all biomes, but it is biased towards dryland ecosystems. We discarded from both databases plants that did not come from dryland areas (sites with aridity index > 0.65). For the rest (N = 1,903), we focused on two key functional traits: light-saturated photosynthetic rate and leaf N content. Details on the methods to assess both traits can be found in ref. *20*. In brief, light-saturated photosynthetic rate was measured on leaves enclosed into a chamber that measures incoming and outgoing CO_2 and H_2O fluxes under optimal light conditions (while other environmental conditions are kept to ambient conditions). Leaf nitrogen content is the percentage of nitrogen mass in respect to dry leaf weight. As nitrogen is a major compound in the composition of photosynthetic enzymes (*20*), this functional trait is directly related to potential photosynthetic activity. In addition, nitrogen content in green leaves is a good proxy of nitrogen remaining in leaf litter after senescence period. As such, it can also be considered as a key parameter affecting leaf decomposition through the regulation of microbial activity (*101*, *102*).

Data analyses

Before data analysis and for map attributes, we first matched the spatial scales of assessed metrics. To do so, when metrics were of broader resolution than aridity (i.e., > 30 arc-sec), we matched the spatial scales by taking the median of aridity indices within the spatial extent of the response variable (e.g., 5 km in vegetation sensitivity index). When ecosystem attributes had a better resolution than the aridity index (e.g., NDVI), we used again the aggregated median of the attribute to match the scale of the aridity index (30 arc-secs). Other aggregation methods (such as the mean or the maximum value) or taking the centroid value of the coarser pixel yielded very

similar results (correlation between aridity values estimated with different interpolations was always > 0.98). Furthermore, the same best model and threshold values were chosen regardless of the interpolation method used ($F_{3,396}$ = 1.368/1.271; P value = 0.25/0.28 for Vegetation sensitivity index/NDVI variables respectively).

Evaluation of linear and non-linear responses to aridity

We fitted linear and non-linear (quadratic and general additive models [GAM] *103*) regressions to the relationships between all our ecosystem variables and aridity, and used the Akaike information criterion (AIC) to decide the model that provided the best fit in each case. This criterion penalizes model fit when more parameters (as used in non-linear regressions) are used, so that the most likely model has the lowest AIC value (*104*). In general, differences in AIC higher than 2 indicate that the models are different (*104*). The linear model is the null hypothesis and assumes a gradual response of a given ecosystem attribute in response to increases in aridity. Quadratic and GAM models evidence a nonlinear but continuous trend throughout the aridity gradient. We chose quadratic to synthetize the simplest case of nonlinear trend, and GAM to summarize more complex trends (through smoothing parameters, *103*).

Thresholds detection

Only when non-linear regressions were a better fit to the data, thresholds may be present. Therefore, we explored the presence of thresholds only when non-linear models were a better fit to the data. We did so because threshold models (e.g., segmented, step and stegmented regressions) force the existence of at least one threshold, and therefore applying these methods to relationships that best fit linear regressions will lead to over-fitting and the detection of spurious thresholds. We typified the responses of a non-linear trend by actively searching the two types of thresholds according to the definition of Groffman et al. (*17*): continuous and discontinuous. Following this definition, we consider a threshold as the point in aridity in which a given variable either changes abruptly its value (discontinuous threshold, or breaking point) or its relationship with aridity (continuous threshold). Continuous thresholds may be well fitted to segmented regressions (i.e. a linear regression that modifies its slope at a certain value of the predictor, or threshold, *105*). Also, when fitting segmented regressions to models that are better fitted to smooth nonlinear continuous trends (such as models that best fit GAM regressions), segmented regressions evidence the point of maximum curvature of the fit. This point can be considered a threshold in the sense that it shows a peak of change in the response of the variable to aridity, even if the fit of segmented regressions is poorer than that of GAM or other nonlinear models. Discontinuous thresholds attain an overall change in the intercept, apart from the slope, and may be fitted to either step (linear regression that changes only intercept at a given point or threshold) or a combination of step + segmented regressions (stegmented; exhibits changes both in intercept and slope at a given point or threshold, *106*, *107*).
Thus, on those variables that were better fitted by either quadratic or GAM models than by linear models, we fitted segmented, step and stegmented regressions. Each of these models renders a parameter describing the point in the predictor (aridity) that evidences the shift in the relationship (in slope, intercept or slope + intercept for segmented, step and stegmented regressions, respectively). We consider this point as the threshold in a given non-linear relationship evaluated. To select among the three thresholds yielded by segmented, step and stegmented regressions, we used AIC criteria to choose the model that best fitted the data. If GAM

regressions where the best model when compared to threshold models, we reported so (Table S2) and selected the threshold yielded by segmented regressions to evidence the point of maximum curvature of the regression. In the particular case that there are clearly two curvatures in a nonlinear regression (e.g., those that would better fit cubic regressions such as is occurring in photosynthetic efficiency at the species level, Fig. S4) we also fitted a segmented regression with two thresholds. We did so because these trends evidence the existence of two thresholds in the same dataset and, therefore, the best fit cannot be performed using approaches best suited for the presence of a single threshold. To obtain the variance associated to each threshold, we performed 200 bootstrap samplings on each database, thus allowing us to find a set of 200 plausible thresholds for each variable. For the probability of occurrence of shrublands variable (that is categorical), we conducted the same methodology described above but using generalized linear models with a binomial distribution link-function.

We used the same analytical approach for variables obtained from interpolations, albeit in these cases the bootstrap procedure was performed by selecting subsets of data. We did so because the *chnpt* package (used to fit stegmented and step regressions) is not efficient with very large datasets due to the immense amount of permutations necessary. Additionally, we enforced homogeneous distribution of aridity levels when running this bootstrap procedure with remote sensing variables. This is important because the sampling design for remote sensing variables is spatially homogeneous. Thus, points located at certain aridity levels may exert a major influence in the trend just because of the different geographical coverage of different aridity levels (e.g., because there are more areas in the globe with an arid than with a dry-subhumid climate (*2*) more points have been sampled in arid areas, and this may influence the overall trend of the relationships). To correct for this issue, we divided the dataset according to aridity in windows taken each 0.005 aridity units (168 windows), and we subsampled 100 points in each window. By doing so, the bootstrap subsampling in which regressions are fitted is built with homogeneous distribution of aridity levels.

Finally, although threshold regressions work well for identifying breaks in continuous trends, there is one case in which they fail to evaluate them: when the variable of interest follow a bimodal distribution. In this case, and because the linear regressions that underlie them rely on changes of the mean, threshold regressions fail on detecting changes in the mode. These trends, in which modes of a variable are interchanged through a gradient (here aridity) and usually overlap for a certain gradient range, are the common expected output of variables following catastrophic shifts (*108–110*). Hence, they are meaningful to infer potential abrupt changes. It is important to note, however, that these regressions do not necessarily evidence critical shifts, for which time-series analysis are necessary. In these cases (stability of soil aggregates, soil organic carbon; both from field and remote sensing; soil organic nitrogen, albedo, plant cover, NDVI and sand content; see Fig S8), the analysis needs to focus on the central tendency of the variable, rather than on the mean, to spot abrupt changes in the most common values within the data. Thus, in the cases where a bimodal distribution of the variable was present (as evaluated by gmdistribution.fit function in MATLAB), we changed the threshold regressions (segmented, step and stegmented) to use quantile regressions instead of regular linear ones when finding the threshold value. Quantile regressions correct the maximum likelihood estimation of linear models, which rely on ordinary least squares of the residuals, to track a specific quantile of distances with residuals of the estimated trend (*111*). If quantile regressions are tuned to match the median, they effectively track the central tendency of the variable along with the predictor and this allows estimating thresholds in variables that overlap for a given range of X-axis values

(see Supplementary text for details). Additionally, quantile regression has important advantages. First, it down-weights outliers, which have a large influence on the detection of thresholds (*112*) and are not easily identified in vast datasets (*113*). Second, quantile regressions have been proven successful to relate variables with an unequal variation of one variable for different ranges of the predictor (*111*), something common in some of our variables. It is worth noting that when ecosystem attributes follow more normal-like distributions and meet assumptions of homoscedasticity along the aridity gradient, fitting quantile regressions will be analogous to fitting common linear regressions.

We used the *chngpt* (*106*) and *gam* (*103*) packages in *R* (*114*) to fit segmented/step/stegmented and GAM regressions, respectively and the *segmented* package (*105*) to fit segmented regressions with two thresholds (in the case of photosynthetic efficiency).

Using other criteria than AIC to select the best models such as the Bayesian Information criterion (BIC) did not affect our results except for fungal animal pathogens (Table S2). In this variable the best model suggested by BIC was step regression, whereas that suggested by AIC was stegmented. However, the thresholds found in both models were indistinguishable (0.796 c.f., 0.798; ANOVA $F_{1,398} = 0.1194$, P value = 0.729).

An analysis of the semivariograms for the residuals of all the variables analyzed showed very low spatial autocorrelation in our results (Fig. S9). We must note, however, that some variables such as soil organic carbon, soil nitrogen, sand content, precipitation interannual variability, albedo, NDVI or fractional cover, show patterns that may indicate some minimal autocorrelation, with a characteristic scale of 30km. To ensure that this was not influencing our results, we recalculated the thresholds for these variables by masking out points that where closer than 30km from others. Results showed that the threshold values were almost identical to the ones previously calculated (Table S3), thus this issue do not affect conclusions of the paper.

Validating the significance of the thresholds identified

To further test whether the thresholds identified significantly affected the slope and/or intercept of the fitted regressions, we bootstrapped linear regressions at both sides of each threshold for each variable. We then extracted the slope and the predicted value of the variable evaluated before and after the threshold and compared them using a Mann-Whitney U test. In all cases, we found significant differences in at least one of these two parameters (Figs S2 and S3).

Building homogeneous groups of thresholds

Once all threshold values were extracted and validated, we performed cluster analysis on the bootstrapped thresholds to evaluate whether or not we could organize them in groups or *phases*. We used the Elbow method (*115*) to find the number of optimum homogeneous groups of thresholds. This methodology analyzes the within clusters sum of squares for several clusters (from 1 to 14). The optimum number of clusters is identified with the broken stick method (i.e., the number of clusters from which variance absorbed tends to an asymptote). We identified three clusters (Fig. S10), with centroids on aridity values = 0.54, 0.69 and 0.84 corresponding to each of the phases reported in the main text (i.e., vegetation decline, soil disruption and ecosystem collapse). We assigned variables to the three phases mentioned in the text based on the most common cluster assignment of their thresholds. We used the cluster centroids as the aridity threshold identifying the phase (to be used in the predictions reported in Fig. 3). These three phases and their associated aridity values were very similar when analyzing differences in the

thresholds found for each pair of variables using the Mann-Whitney U test (with each of the 20 ecosystem structural and functional features included as a factor and the bootstrapped threshold values as response variable; data not shown).

Maps for future predictions

We used the aridity maps provided by Huang et al. (*3*) to locate those areas that are likely to cross the main aridity thresholds identified due to ongoing increases in aridity driven by climate change. As thresholds to be evaluated, we selected the three groups found in cluster analysis explained above. Extrapolations extracted from these maps provide a widely recognized prospect for aridity trends until 2100 using two different scenarios (Representative Concentration Pathways [RCP] 8.5 and 4.5, inferring exponential and saturated increases of CO_2 emissions respectively, see ref: *116*).

It is worth noting that the results of this exercise are temporal extrapolations of results obtained using spatial gradients, and therefore constitute what is known as a space-for-time substitution approach (*117*). The use of this approach is common in ecology (*118–120*), as it allows to infer hypotheses related to temporal changes when temporal series are not available or do not suffice to cover processes that operate at very slow temporal rates (*117*, *119*). While space-for-time substitutions have been proven successful in some situations (*121*), they have also been criticized as spatial gradients may include drivers different than those driving temporal changes in ecosystem variables and do not include adaptation of ecosystems to the new environment (*122*, *123*). The interpretation of these results, thus, must consider this limitation. For instance, prospects in areas that are not drylands today, which delimit the spatial extent of the data used in our analyses, are extrapolations. These extrapolations may be unreliable due to, for example, the importance of environmental variables other than aridity (e.g., temperature in the case of boreal forests) as drivers of changes in the ecosystem attributes investigated. To acknowledge this, areas crossing the aridity thresholds identified in our study in the future but that are not drylands today are made transparent in Figs. 3 and S7. Due to the uncertainty surrounding these results, they should be interpreted with caution. We present in the manuscript results from RCP 8.5 scenario, but those from RCP 4.5 can be found in Fig. S7.

Data accessibility

All analyses were performed using R (*114*). The codes used as well as the data extracted are available from ref. *124*.

Supplementary text: On the relevance of tracking the central tendency to identify thresholds in bimodal variables

Introduction to the problem and its relevance

There is a particular case in which threshold models (segmented, stegmented and step models) fail to identify the threshold of a variable of interest (here ecosystem features) through an x-axis variable (here aridity). This occurs when the variable of interest exhibits two trends that overlap in a certain range of x-axis conditions. Situations like this are important, as they exemplify circumstances in which the variable of interest changes its mode through the x-axis but not the mean of overall values. These cases match what is expected when the variable of interest follows two regimes that co-exist for a certain range of conditions (summarized by the x variable). For instance, the existence of different stable states has been hypothesized to follow these trends (109). In these cases, the dynamics of the system exhibit two attractors whose stable states coexist for a certain range of x conditions as is the case of systems exhibiting critical transitions. Other situations may also yield coexisting regimes without meaning that these states are interchangeable through time (e.g., the existence of savannas or different types of forests depending on temperature or soils). These overlapping states, are anyway different regimes and the fact that they change abruptly through increases in the x-axis is relevant to study, even if it is not possible to infer temporal changes or catastrophic shifts. Therefore, the finding of the threshold in which modes of variables of interest change is important to track the possibility of their existence through the x-axis. In particular, in these situations, we are interested in the threshold from which a shift between regimes is more likely to occur, and those are better exemplify by the changes of the mode (or the central tendency) than by changes in the mean values of the variable of interest.

Fitting bimodal data

The chngpt R package (106) fits two glm regressions to the data using different thresholds and finds the threshold that minimizes OLS residuals. This would show greater uncertainty when there is an overlap in values of Y for a given value of X (aridity in our case), which is what would be expected in a hysteresis-type response (109). Stegmented regressions calculated using OLS, as implemented in the chngpt R package, rely on the mean of least squares, and this mean is biased by the increasing number of outliers found when two levels of Y overlap under a given range of X. As a result, applying mean least squares residuals in variables that follow two regimes that overlap for a certain range of the x-axis often results on detection of the threshold at the point in which the overlap zone starts or finishes (Fig. S11A). To minimize this issue is why, we used quantile regressions. Quantile regressions weight the residuals to fit the central tendency, thus correcting for the bias and finding the optimum fit in the proper aridity value (Fig. S11B). To do so, we developed a custom function that works similarly to the chngpt original function, but uses quantile regression instead of linear models.

Example with soil organic carbon data

We illustrate this point with our soil organic carbon data. When using regressions based on linear OLS, we found a (soft) threshold around 0.6 (where a smooth change in slope is detected in the

smoothed values shown in Fig S12A). However, when applying quantile regressions, we found that there is a clear threshold at aridity = 0.7 instead of 0.6 (see both fittings at Fig. S13).

The aridity value of 0.7 matches the threshold found using the Livina and Lenton´s approach (109, Fig. S14), which was designed precisely to study bi-stable states and uses potentials (based on data points distribution) to track exactly the central tendency of the data. The threshold at 0.7 also matches the shift in soil C values seen when smoothing the data using the median instead of the mean (Fig. S12 B), minimizing the increasing uncertainty caused by overlap in Y values under this type of behavior.

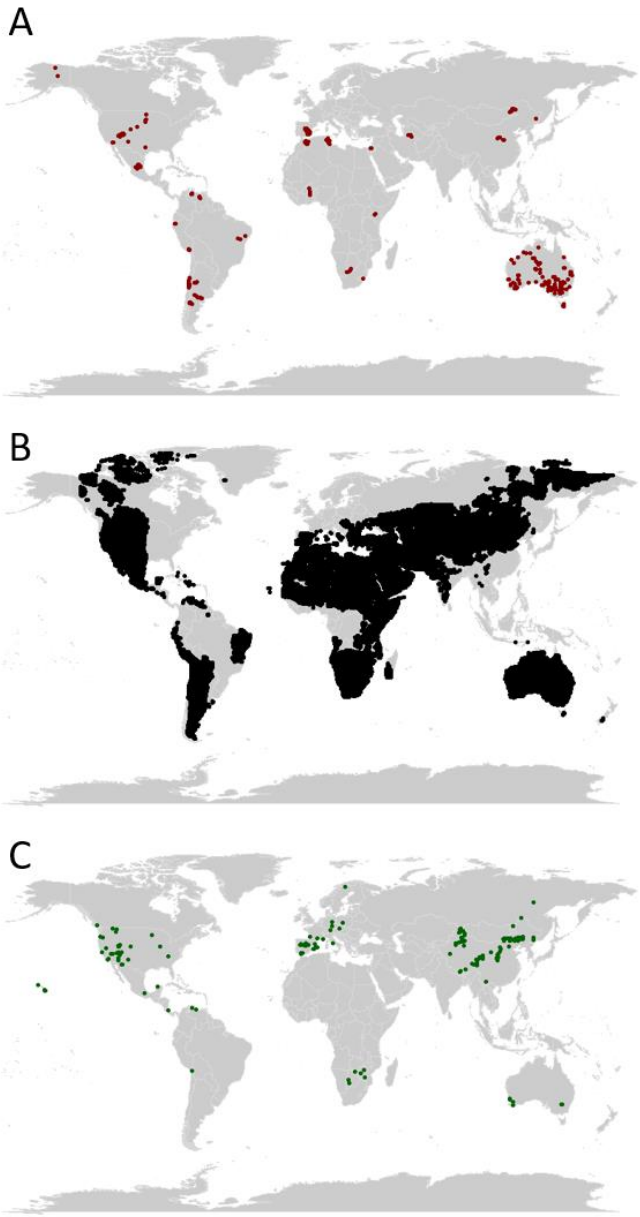
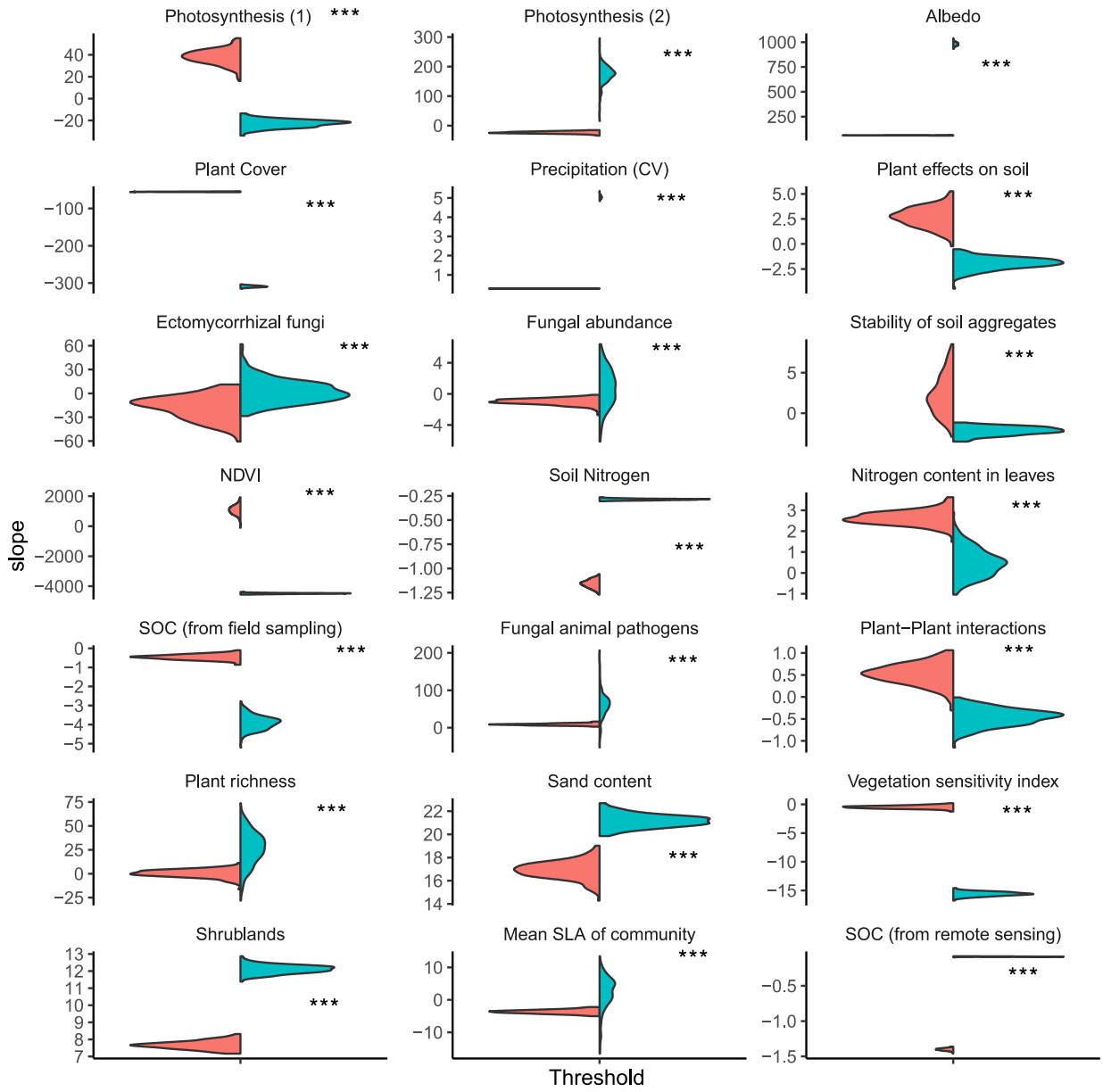


Fig. S1. Location of the data points used in the study.

A: variables obtained from a global field survey (species richness, stability of soil aggregates, ratio of positive vs negative plant-plant interactions, plant effects on soil organic carbon; specific leaf area; soil organic carbon from field studies, abundance of microbial OTUs and relative abundance of fungal groups); B: variables obtained from global maps and interpolations (albedo, vegetation fractional cover, coefficient of variation of interannual precipitation; NDVI; soil organic carbon and nitrogen from remote sensing, sand content, shrubland occurrence and vegetation sensitivity index): C: variables from functional traits databases (photosynthetic efficiency and nitrogen content in leaves).



485

Fig. S2. Differences in the predicted value of the variable at the aridity threshold.

Violin diagrams showing bootstrapped predicted values of each variable at each side of every aridity threshold found in the study (red: regression before the threshold; blue, after the threshold). Asterisks indicate significant differences when conducting a Mann-Whitney U test between before and after the threshold where: ; *= P value >0.05; **= P value >0.01; ***= P value <0.01. n.s.= not significant. SOC = soil organic carbon; NDVI: Normalized difference vegetation index; Photosynthesis (1 or 2) indicate the two thresholds found for photosynsthesis under controlled conditions.

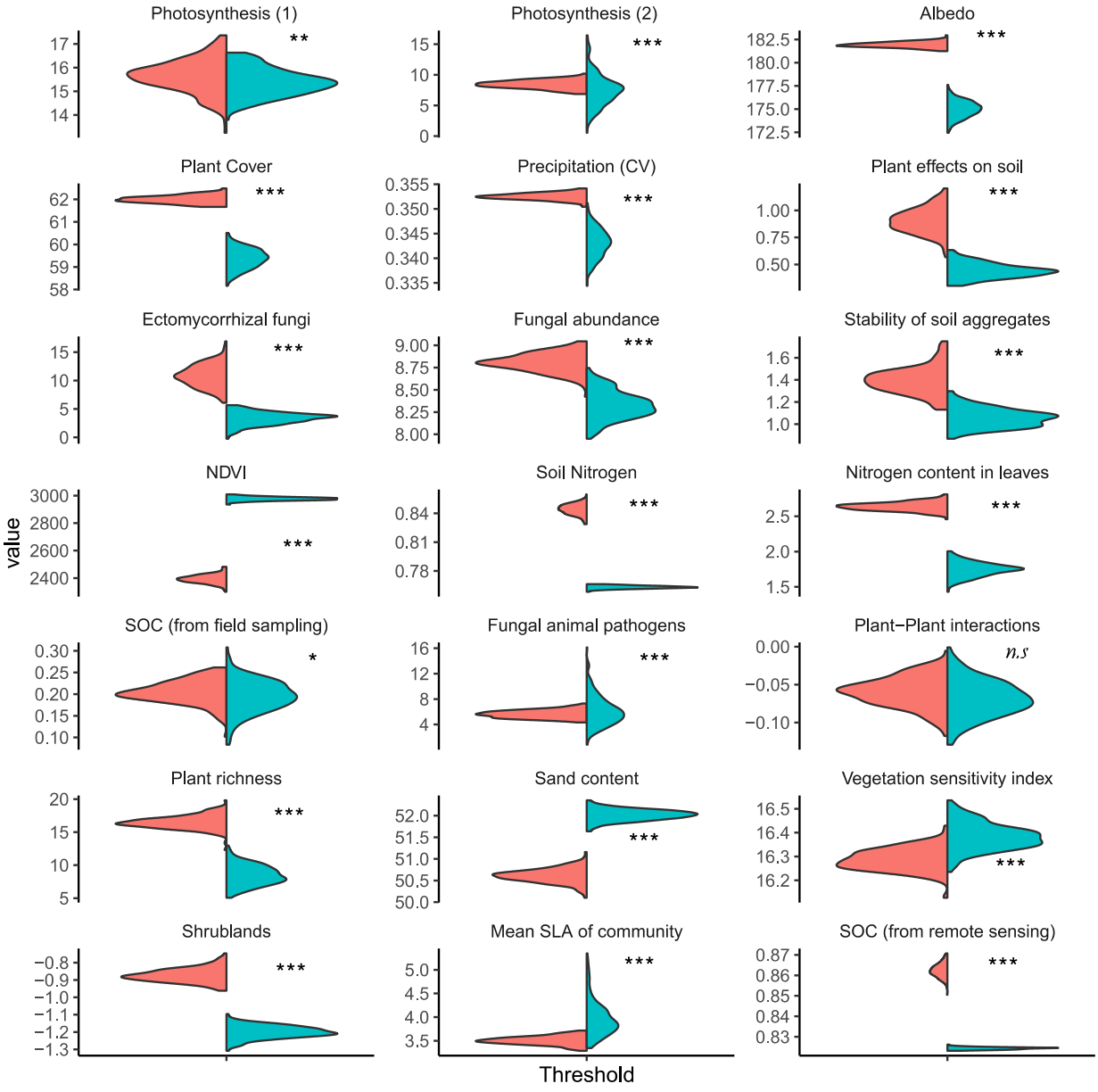


Fig. S3. Differences in the slope at both sides of the aridity threshold.

Violin diagrams showing bootstrapped values of the slope of the two regressions existing each side of every aridity threshold found in the study (red: slope before the threshold; blue, after the threshold). Rest of legend as in Fig. S1.

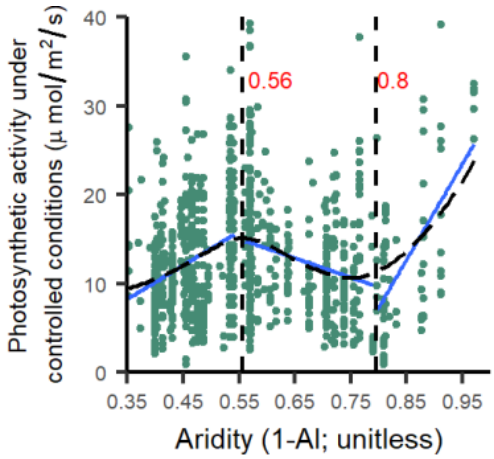


Fig. S4. Additional example of ecosystem attribute (photosynthesis measured under controlled conditions) showing non-linear responses in the vegetation phase.

Rest of legend in Fig. 2.

505

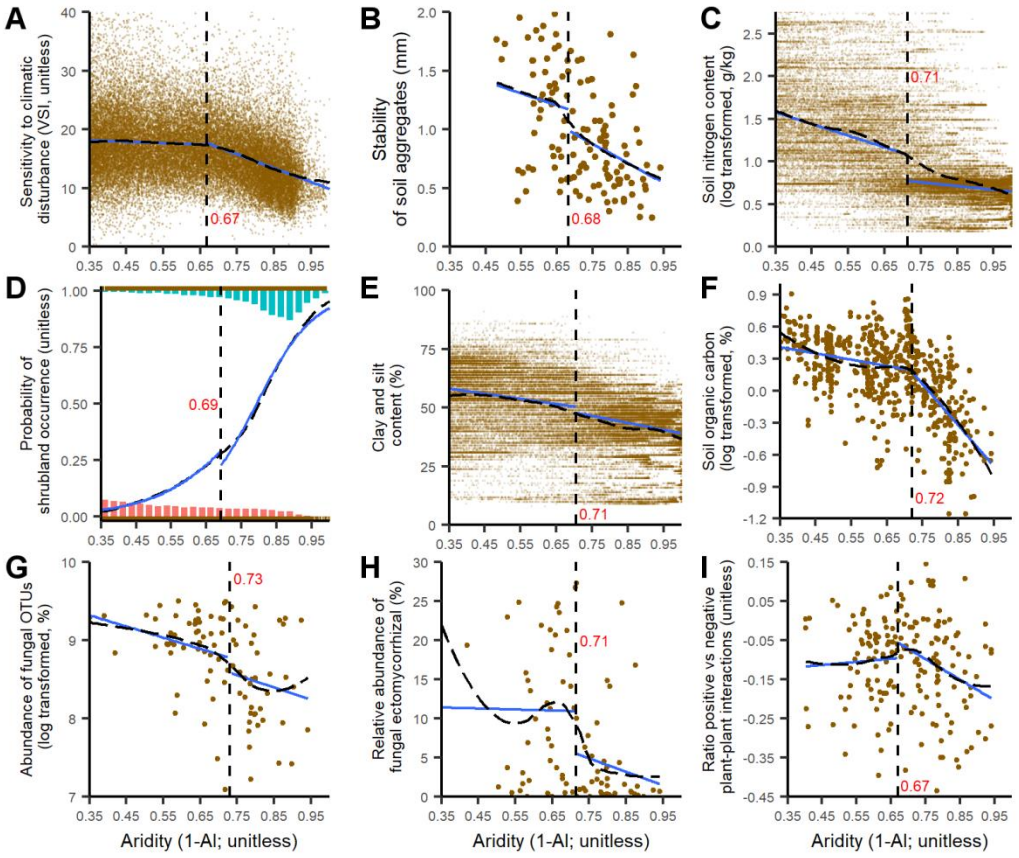


Fig. S5. Additional examples of ecosystem attributes showing abrupt responses in the soil disruption phase.

A: vegetation sensitivity to climatic disturbance (Vegetation sensitivity index); B: soil aggregate stability; C: soil nitrogen content interpolated from global maps; D: probability of occurrence of shrubland vegetation; E: clay and silt content in soils; F: soil organic carbon obtained from standardized field surveys; G: microbial abundance of soil fungi; H: positive/negative plant-plant interactions ratio estimated from co-occurrence networks; I: relative abundance of ectomycorrhizal fungi. Rest of legend as in Fig. 2.

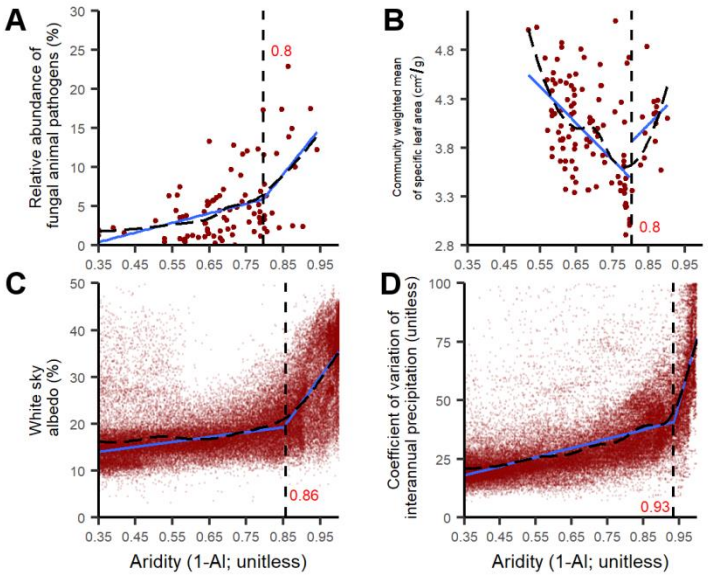
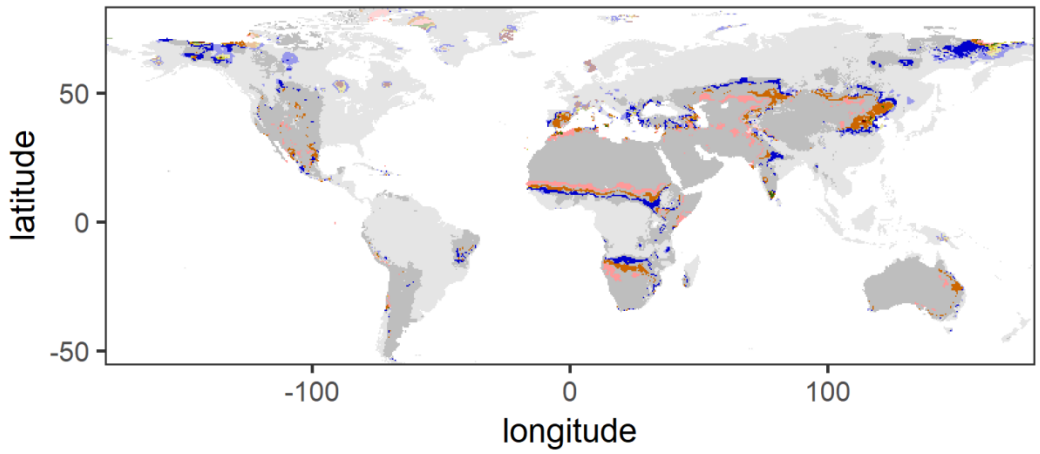


Fig. S6. Additional examples of ecosystem attributes showing non-linear behaviour during the systemic collapse phase.

A: Relative abundance of fungal animal pathogens; B: community weighted mean of specific leaf area; C: white sky albedo; D: precipitation inter-annual variability. Rest of legend in Fig. 2.



Legend	Threshold Crossed	% emerged lands affected	% drylands affected
	0.54	4.6 %	5.9 %
	0.54 & 0.69	0.6 %	0.4 %
	0.54, 0.69 & 0.83	0.4 %	0.1 %
	0.69	2.7 %	5.5 %
	0.69 & 0.83	0.2 %	0.3 %
	0.83	2.1 %	4.3 %

Fig. S7. Areas of the globe that will cross each of the thresholds described (or several of them) under the RCP4.5 scenario, which assumes an asymptotic increase of CO_2 emissions. Rest of legend as in Fig. 3.

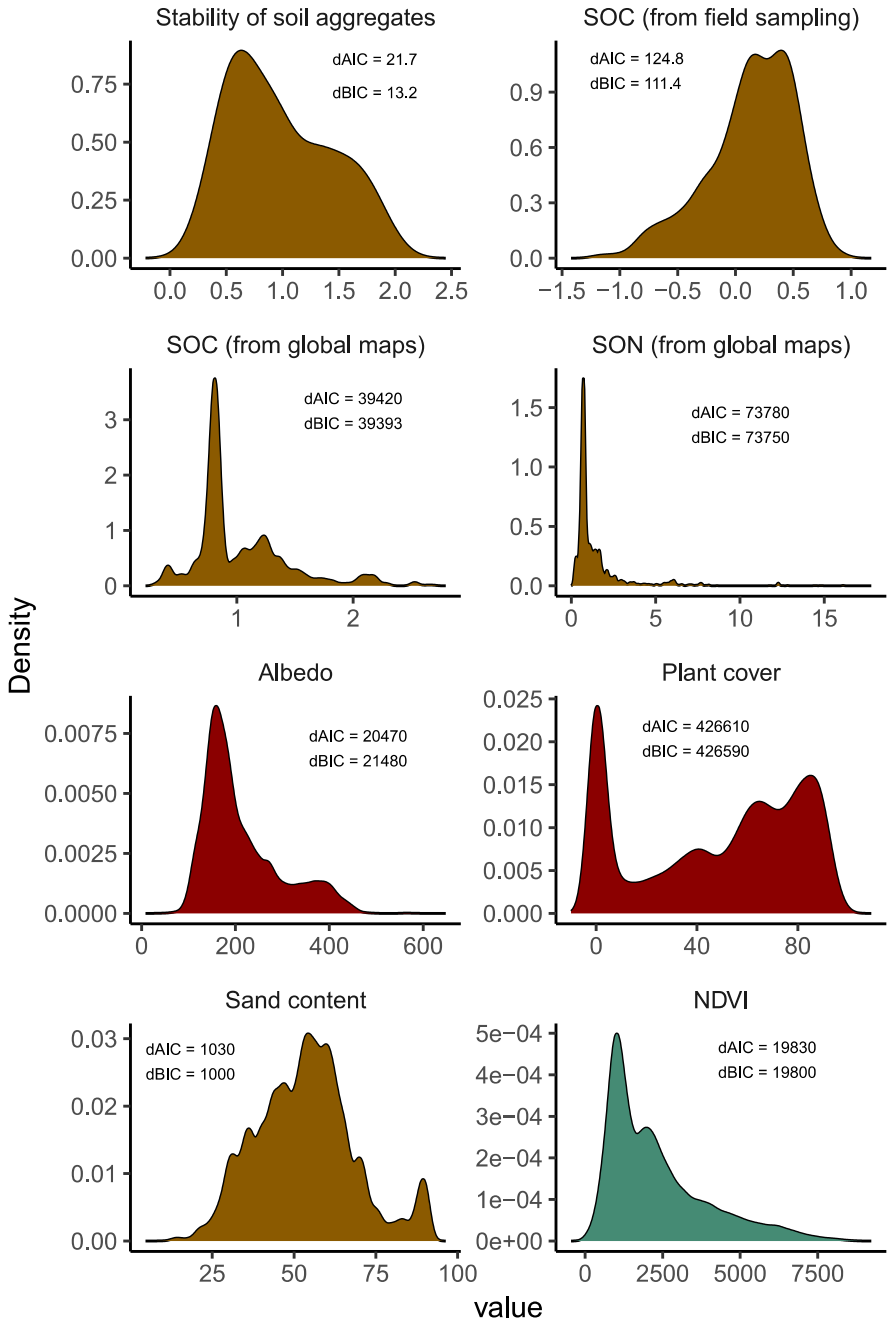
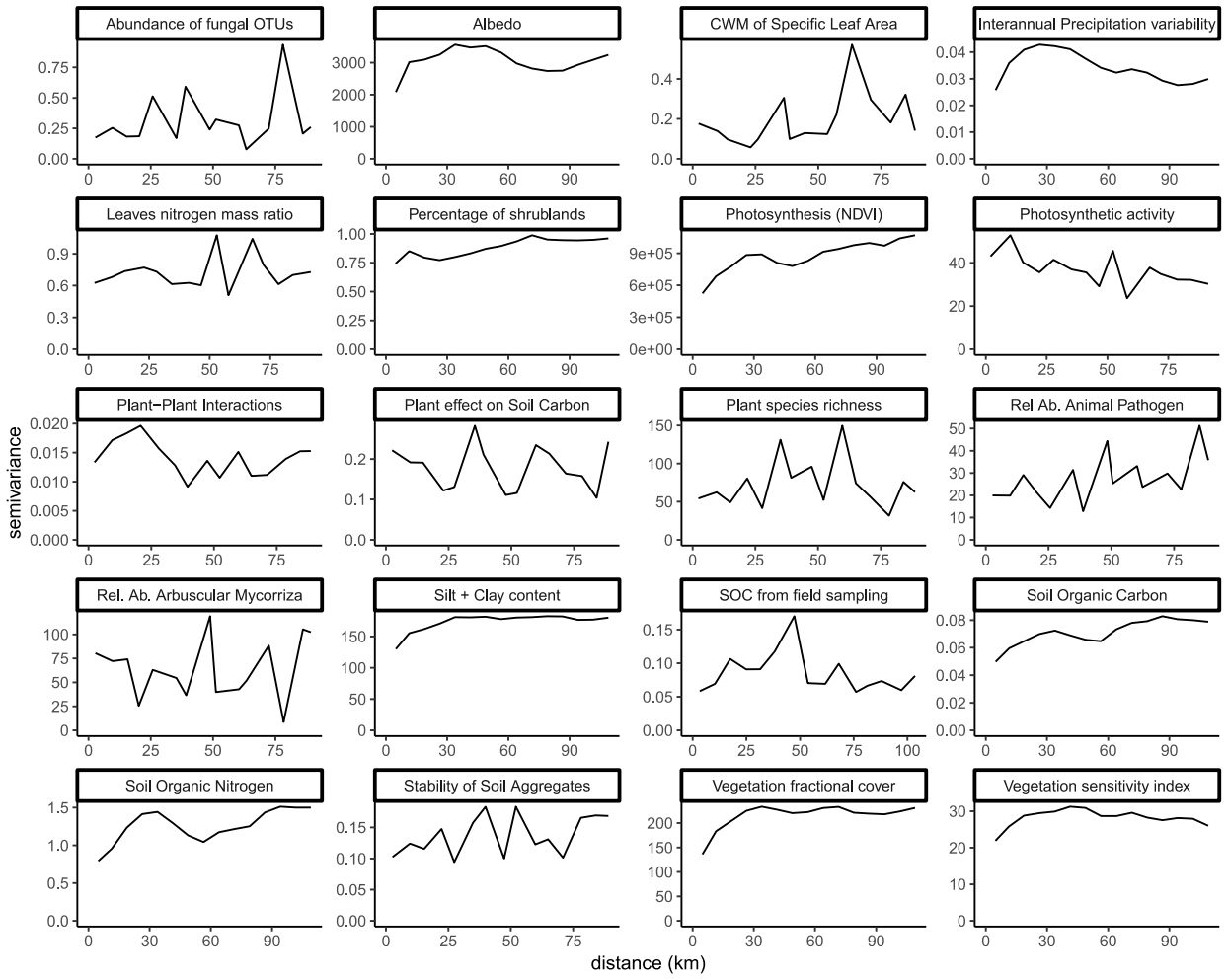


Fig. S8. Variables that present bimodality.

Kernel density function of the variables that presented bimodality. In these variables, quantile regression is preferred over simple linear regression to depict changes in the central tendency. dAIC/BIC= difference in AIC/BIC values from fitting two modes vs fitting one mode (if positive the variable is bimodal).



535

Fig. S9. Spatial Autocorrelation.

Semivariograms of the residuals of the variables analyzed after fitting their response to aridity.

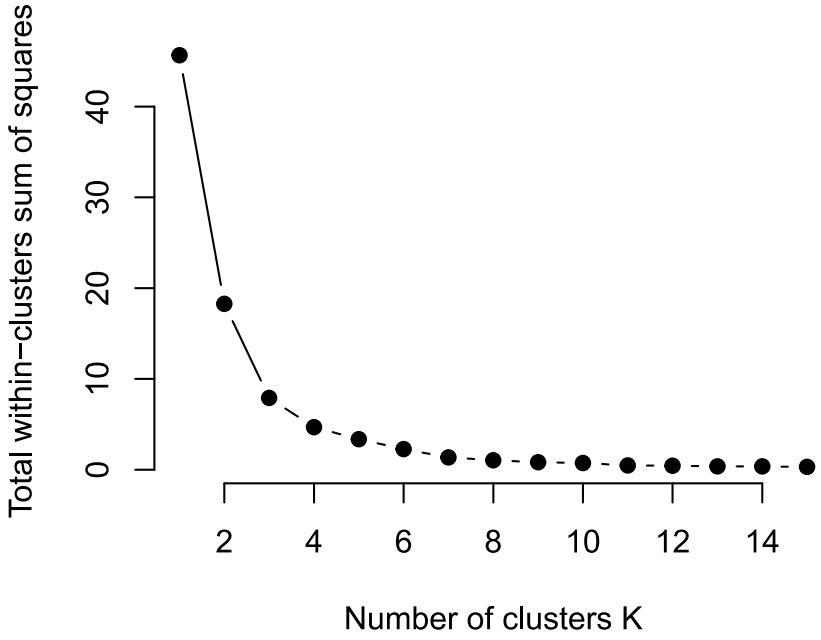


Fig. S10. Cluster analysis on threshold phases.

Total within-cluster sum of squares for several number of clusters. The optimal number of clusters is that from which adding more clusters results in similar variance explained; here three.

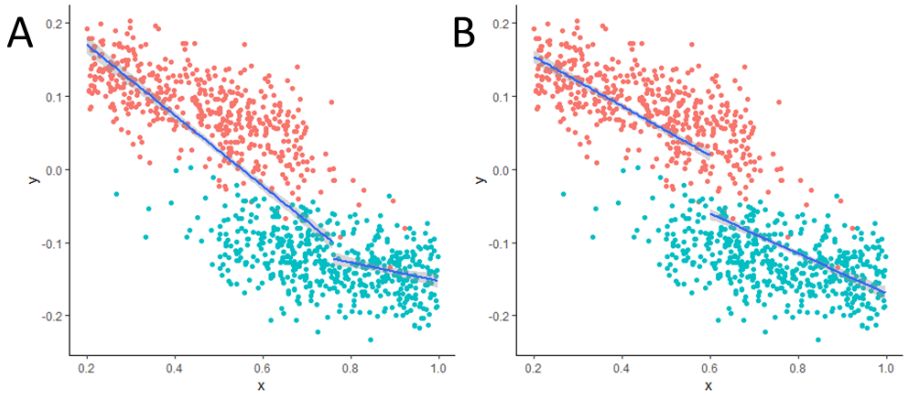


Figure S11. Example with simulated values of Y for X using two models (points in red c.f. blue) that overlap for the X range of 0.5-0.7, with increasing probability of being fitted by the second model around X = 0.6. Fitted stegmented regression lines are applied using generalized linear models (A) and quantile regressions (B) with contrasting values of the threshold found.

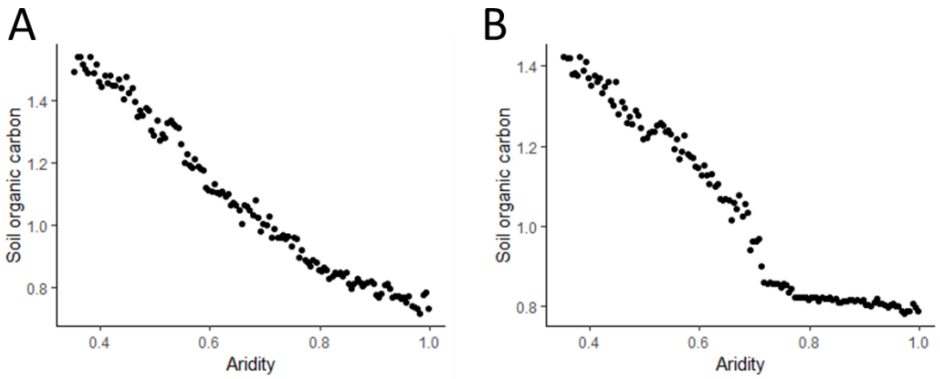
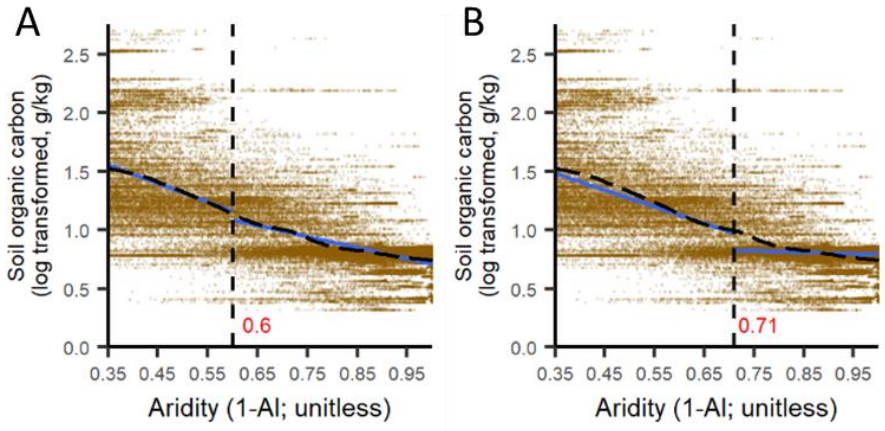


Figure S12. Smoothed trend of soil organic carbon throughout aridity using the mean (A) and the median (B).



555 **Figure S13**. Results of the fitted values of a stegmented model using glms (A) and quantile regressions (B) to effectively track the central tendency of the data.

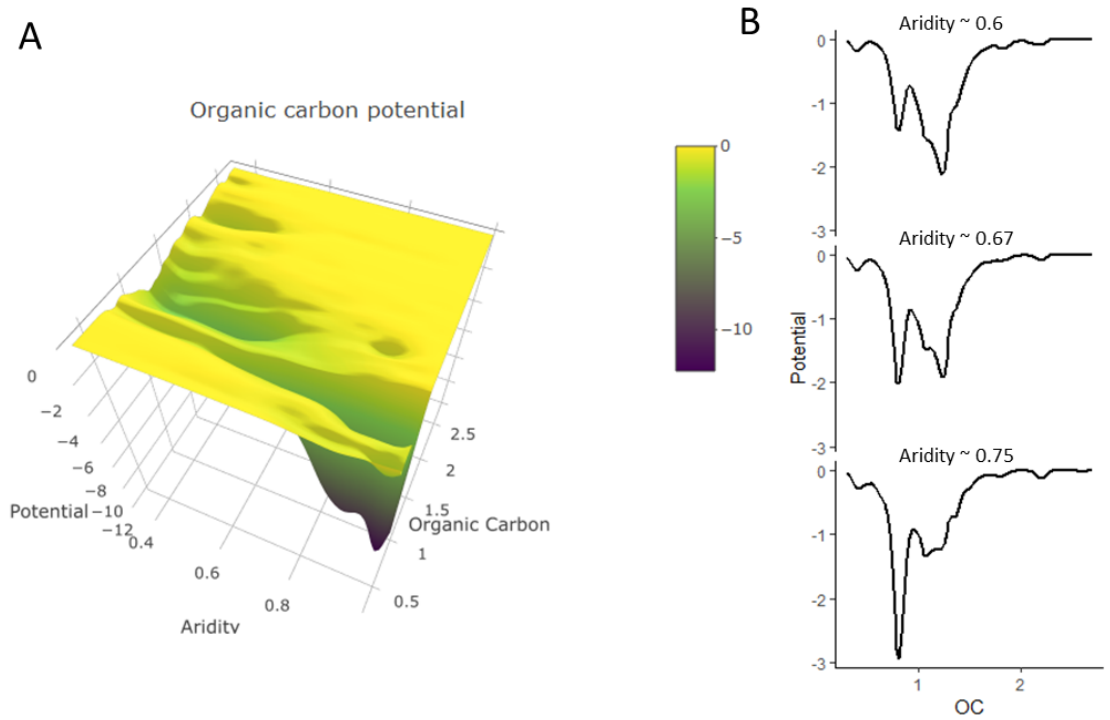


Figure S14. Pattern of the potentials (density of points) of soil organic carbon (SOC) values through increasing aridity. The data show a clear pattern resembling those expected for alternative states (109), with a clear threshold around aridity values of 0.7 where there is a clear shift from high to low soil carbon values. In B three slices of A showing the potential (equivalent to the inverse of a density plot of SOC values) at different aridity values.

Table S1.

Variables used in the study. Description and origin of variables used in this study.

Database type	Variable typology	Variables taken	Description	Provider	Database (product)	Reference	Resolution
Aridity	Climate	Aridity	Aridity index calculated as annual precipitation/annual evapotranspiration	CIGAR	CIGAR-CSI. Global Aridity and PET database	38	30 arc-secs
Map variables	Soil	Soil organic carbon	Soil organic carbon content, interpolated from global soil typology maps	ISRIC	ISRIC- WISE Soil Propery Database.	40	30 arc-secs
	Soil	Soil nitrogen content	Soil nitrogen content, interpolated from global soil typology maps				30 arc-secs
	Soil	Silt + Clay content	Soil Silt and Clay content, interpolated from global soil typology maps				30 arc-secs
	Climate	Albedo	White sky albedo (in case of totally diffuse illumination)	LP DAAC	MODIS (MCD43D product)	59	250m
	Climate	Inter-annual precipitation variability	Coefficient of variation of interannual precipitation	University of Delaware NCAR UCAR	CHELSA	64	30 arc-secs
	Vegetation	Vegetation fractional cover	Fractional cover of trees and non trees vegetation cover as interpolated from MODIS products	LP DAAC	MODIS (MOD44B product)	50	250m
	Vegetation	Plant productivity (NDVI)	NDVI averaged indices from locations evenly distributed around the world taken each	LP DAAC	MODIS (MOD13Q1 product)	46	250m
	Vegetation	Percentage area of grassland vs shrublands	Relative cover of places classified as shrbulands vs those classified as grasslands	FAO	MODIS (MCD12Q1)	39	30 arc-secs
	Vegetation-Climate	Vegetation sentivity index	Vegetation sensitivity to climate fluctuations	Alistair Seddon	Vegetation Sensitivity Index (database at LEFT project: https://www.left.ox.ac.uk/)	25	5km
Standardized field samplings	Soil	Stability of soil aggregates	Stability of soil aggregates meassured under controlled conditions in the laboratory with soil sampled in a standardized protocol coming from global drylands	F. T. Maestre	BIOCOM	unpublished	local

Database type	Variable typology	Variables taken	Description	Provider	Database (product)	Reference	Resolution
Standardized field samplings	Soil	Soil organic carbon	Soil organic carbon from standardize sampling and lab protocols.		BIOCOM +	70, 14, 72, 73	local
	Plant-Plant	Positive/Negative plant-plant interactions	Ration of positive versus negative links in species using co-occurrence matrices of species sampled under standardized field survey of global drylands	F. T. Maestre	BIOCOM	82	local
	Vegetation	Community weighted mena of Specific Leaf Area	Weighted mean of specific leaf area in the communities sampled in BIOCOM.	F. T. Maestre	BIOCOM	29	local
	Vegetation	Plant speries richness	Number of species of perennial plants recorded in BIOCOM	F. T. Maestre	BIOCOM	69	local
	Plant-Soil	Vegetation effect on soils	Difference between organic carbon of soils under dominant vegetation and soils in adjacent barren areas	F. T. Maestre	BIOCOM	69	local
	Microbial	Abundance of fungal OTUs	qPCR estimation of number of OTUs representing Fungi of soils coming from global standardized survey and meassuredi n the same laboratory under standardized conditions	F. T. Maestre			local
	Microbial	Rel. Ab. Animal pathogen	Relative abundances of fungi groups estimated from soil coming from global drylands which were analyzed under standardized laboratory conditions using miseq (see methods)	F. T. Maestre			local
	Microbial	Rel. Ab. Arbuscular Mycorrhiza		F. T. Maestre	BIOCOM +	69, 70	local
	Microbial	Rel. Ab. Ectomycorriza		F. T. Maestre			local
	Microbial	Rel. Ab. Lichenized		F. T. Maestre			local
	Microbial	Rel. Ab. Plant pathogens		F. T. Maestre			local
	Microbial	Rel. Ab. Saprophytes		F. T. Maestre			local
Standardized laboratory messurements	Vegetation	Leaves nitrogen mass ratio	Nitrogen mass/ leave mass meassured for individuals takn at contrasting aridity levels worldwide	Vincent Maire		20, 31	Individual
	Vegetation	Potential photosynthetic activity	Photosynthetic rate of plants taken at contrasting aridity levels worldwide but meassured under controlled standardize conditions.	Vincent Maire	Wang & Maire databases	20, 31	Individual

Table S2.

Best models obtained for each variable. Variables used with their corresponding Akaike (AIC) and Bayesian (BIC) values after fitting linear, nonlinear and best threshold models. Lower AIC/BIC values indicate a better fit of the model. Coefficients of determination (R2) from the lineal and best model fitted are also shown. * indicates variables that had a bimodal distribution; quantile regression was used instead of glm regression within threshold models to find the threshold in these cases (see Fig. S8). Rel. Ab. = Relative abundance; GAM = General Additive Models.

Database type	Variable typology	Variable name	AIC linear	AIC nonlinear	Best AIC	BIC linear	BIC nonlinear	Best BIC	R^2 lineal	R^2 threshold
Map variables	Soil	Soil organic carbon*	21868.9	20865.1	Stegmented = 20866.6	21895.4	20861.7	Stegmented = 20802.7	0.44	0.45
	Soil	Soil nitrogen content*	179982	178327.2	Stegmented = 178311.8	180008.5	178353.8	Stegmented = 178350.1	0.25	0.27
	Soil	Silt + Clay content*	409536.4	409256.4	GAM = 409256.4	409562.9	409282.9	GAM = 409282.9	0.19	0.19
	Climate	Albedo*	572735.3	558798.2	GAM = 558798.2	572761.8	558824.7	GAM = 558824.7	0.36	0.51
	Climate	Precipitation inter annual variability	-40816.6	-52988.7	Piecewise = -56923.8	-40790.1	-52962.2	Piecewise = -56888.5	0.52	0.65
	Vegetation	Vegetation fractional cover*	447914.6	432668	GAM = 432668	447941.1	432694.5	GAM = 432694.5	0.63	0.73
	Vegetation	Plant productivity (NDVI)*	856364.2	854695.3	GAM = 854695.3	856390.6	854721.7	GAM = 854721.7	0.38	0.4
	Vegetation	Percentage area of shrublands	36107.3	35881.7	Piecewise = 35879.8	36124.5	35898.9	GAM = 35898.9	0.28	0.29
	Vegetation-Climate	Vegetation sensitivity index	218018.3	217206.4	Piecewise = 217201.1	218043.7	217231.7	Piecewise = 217233.8	0.11	0.13
Standardized field samplings	Soil	Stability of soil aggregates*	138.9	130.6	Stegmented = 126.1	144.6	139.1	Stegmented = 140.3	0.21	0.35
	Soil	Soil organic carbon*	314.3	225.9	Stegmented = 206.3	327.6	239.2	Stegmented = 228.5	0.3	0.4
	Plant-Plant	Positive/Negative plant-plant interactions	-260.5	-264.8	Piecewise = -266	-250.9	-255.1	Piecewise = -255.1	0	0.05
	Vegetation	Community weighted mean of Specific Leaf Area	171.3	151.1	Stegmented = 146.4	179.6	159.5	Stegmented = 160.3	0.11	0.3
	Vegetation	Plant species richness	1618.8	1616.3	Step = 1606.6	1628.5	1626	Step = 1616.8	0.06	0.14
	Plant-Soil	Plant effect on soil organic carbon	292.1	264.5	Stegmented = 260.2	301.9	274.2	Stegmented = 268.8	0.03	0.24
	Microbial	Abundance of fungal OTUs	117.4	113.6	Step = 102.4	124.7	121	Step = 109.7	0.1	0.15
	Microbial	Rel. Ab. Animal pathogen	535.3	528.6	Stegmented = 524.2	542.9	536.2	Step = 534	0.21	0.34
	Microbial	Rel. Ab. Arbuscular Mycorrhiza	343.9	342.4	LINEAR	384.8	383.7	LINEAR	0	
	Microbial	Rel. Ab. Ectomycorrhiza	691.8	684.6	Step = 685.8	699.5	692.2	Step = 693.5	0.15	0.22
	Microbial	Rel. Ab. Lichenized	400.2	398.6	LINEAR	456.9	458.4	LINEAR	0.01	
	Microbial	Rel. Ab. Plant pathogens	520.7	520.8	LINEAR	589.3	591.3	LINEAR	0.02	
	Microbial	Rel. Ab. Saprophytes	624.2	623	LINEAR	695	697.1	LINEAR	0	
Standardized experimental messurements	Vegetation	Leaves nitrogen mass ratio	3154.4	3063.3	Stegmented = 3053.9	3169.9	3078.6	Stegmented = 3079.4	0.01	0.1
	Vegetation	Photosynthetic activity under controlled conditions	6149.4	6047.2	Double Piecewise = 6034.6	6163.9	6061.7	Double Piecewise = 6068.4	0.01	0.19

Table S3.

Thresholds of spatially autocorrelated variables. Thresholds (mean ± standard deviation) for variables exhibiting certain autocorrelation according to Fig. S9 with the values originally reported and those resulting from masking out proximal points (less than 30km away from others).

Variable	Threshold found in the main text	Threshold after removing SAC
Albedo	0.857 ± 0.008	0.859 ± 0.009
Vegetation fractional cover	0.786 ± 0.013	0.783 ± 0.015
Precipitation inter-annual variability	0.929 ± 0.008	0.928 ± 0.012
NDVI	0.536 ± 0.064	0.516 ± 0.065
Soil nitrogen	0.672 ± 0.135	0.677 ± 0.091
Soil organic carbon	0.707 ± 0.034	0.682 ± 0.024
Silt+Clay content	0.669 ± 0.128	0.677 ± 0.102