Viewpoints

An operational definition of the biome for global change research

Summary

Biomes are constructs for organising knowledge on the structure and functioning of the world's ecosystems, and serve as useful units for monitoring how the biosphere responds to anthropogenic drivers, including climate change. The current practice of delimiting biomes relies on expert knowledge. Recent studies have questioned the value of such biome maps for comparative ecology and global-change research, partly due to their subjective origin. Here we propose a flexible method for developing biome maps objectively. The method uses range modelling of several thousands of plant species to reveal spatial attractors for different growth-form assemblages that define biomes. The workflow is illustrated using distribution data from 23 500 African plant species. In an example application, we create a biome map for Africa and use the fitted species models to project biome shifts. In a second example, we map gradients of growth-form suitability that can be used to identify sites for comparative ecology. This method provides a flexible framework that (1) allows a range of biome types to be defined according to user needs and (2) enables projections of biome changes that emerge purely from the individualistic responses of plant species to environmental changes.

Introduction

Our current use of the term biome traces back to Schimper (1903). Although Schimper did not use the word biome, his concept that plant growth forms are distributed across the globe in accordance with the availability of light, water, nutrients and heat underpins current use of the term. Schimper's premise that climate and soils shape vegetation at global scales has prevailed as a system for understanding and comparing ecosystem dynamics (Mucina, 2019) and evolutionary processes (Crisp *et al.*, 2009; Donoghue & Edwards, 2014) in different parts of the world.

There is, however, no universal truth when it comes to biomes. Biomes are simplifications of reality that can serve in the process of testing hypotheses and thereby revealing ecological truths (Higgins *et al.*, 2016). The implication is that the ecological and evolutionary questions that a study aims to answer will define what kind of biome concept is needed. For example, when biomes are defined as vegetation units with contrasting phenological response to climate parameters, they can serve as useful constructs for monitoring

changes in terrestrial ecosystems using satellite remote sensing (Higgins *et al.*, 2016). Other biome concepts may be appropriate for organising our knowledge on how ecosystems function. For example, global syntheses of ecosystem productivity, plant function or biodiversity patterns are organised by biomes that are defined by the dominant plant growth form (e.g. evergreen needle leaf forest, deciduous broadleaf forest) (Churkina & Running, 1998) or by a combination of vegetation physiognomy and climate descriptors (e.g. tropical rain forest, temperate grassland) (Whittaker, 1975; Reich *et al.*, 1997; Schultz, 2005; Echeverría-Londoño *et al.*, 2018). Alternatively, if a study aims at comparing or projecting ecosystem responses to external drivers, a functional definition of biomes may be more appropriate. For example, if effects of elevated atmospheric CO_2 concentrations are assessed, the photosynthetic pathways of the dominant plant growth forms may be considered as a criterion for defining biomes (Higgins & Scheiter, 2012). The various demands on biome maps necessitate a flexible, transparent and repeatable method for constructing fit-for-purpose biome maps.

Existing biome maps are constructed in different ways (for a recent review see Mucina, 2019). The frequently used WWF biome map (Olson *et al.*, 2001) for example uses expert knowledge and regional vegetation maps (in turn created by experts) to construct a global biome map. Such expert-based maps are influenced by the mappers' conception of what, for example, a Mediterranean shrubland, boreal forest or savanna is. As a consequence, expert-based maps are not repeatable because different experts may group and delimit vegetation units differently. Moreover, climatic thresholds are often used to delineate biome boundaries. In such cases, the analysis and prediction of biome distributions in response to climate are confounded (Moncrieff *et al.*, 2016). A solution may be to use satellite-based biome maps, which are more objective in this regard. The MODIS land cover type map (Friedl *et al.*, 2010), for example, classifies pixels with different remotely sensed reflectance patterns using a supervised classification algorithm. The algorithm is trained with pixels assigned to biomes by experts. The MODIS map, although trained by expert knowledge, is repeatable as long as one agrees with the a priori biomes assigned to pixels for training of the algorithm.

Deficiencies in the definition and mapping of biomes by experts reduce the usefulness of biome concepts for comparing ecosystems in different parts of the world (Higgins *et al.*, 2016; Moncrieff *et al.*, 2016). Symptoms of such deficiencies are that the same biome in different regions often occupies dissimilar positions in environmental space (Moncrieff *et al.*, 2015), different biomes can exist in the same environmental space (Moncrieff *et al.*, 2014) and different floristic instances of the same biome may behave in ecologically divergent ways (Lehmann *et al.*, 2014). This hampers our ability to draw general conclusions from comparisons of ecological and evolutionary processes in different parts of the world, because places

with convergent selection for plant growth forms by climate and soils may be incorrectly identified.

A solution may be to follow a deductive approach in which biomes emerge from the data rather than being defined a priori. A deductive approach is deployed in dynamic global vegetation models (DGVMs), which model the abundance of different plant functional types (PFTs) based on physiological principles (Prentice *et al.*, 2007). DGVMs represent the huge functional diversity of plants with a small number of PFTs (Lavorel *et al.*, 1997), assuming that all species of a PFT respond in exactly the same way to environmental conditions. This simplifying assumption makes DGVM-derived biomes reliant on the individual PFTs being adequately conceptualised and parameterised. Another example of a deductive biome map is the Functional Biome Map of Higgins *et al.* (2016), which uses remotely sensed phenological information to classify vegetation by their height and productivity and the extent to which a pixel's vegetation appears to be limited by cold or moisture. The Higgins *et al.* (2016) approach is repeatable because, although expert knowledge motivates the algorithm and selection of vegetation attributes to discern biomes, the biomes that emerge are not defined a priori. However, the Functional Biome Map is not explicitly true to Schimper's concept of convergent selection of plant growth forms by climate. This is because it is derived largely from ecosystem-level attributes detectable using Earth observation satellites. The Functional Biome Map misses important plant-level attributes; for example, the map does not discern between needle leaf vs broad leaf, evergreen vs deciduous, or C_3 vs C_4. It may thus be of limited applicability in studies seeking to compare ecological, evolutionary or biogeochemical dynamics of the same ecosystems in different parts of the world. In conclusion, we still lack a deductive biome map that represents the selection of functionally important growth forms by climate and soils, and at the same time overcomes the dependency of DGVMs on PFT parameterisation.

In this contribution, we propose and operationalise a deductive biome concept that overcomes the limitations of expert-based and existing deductive concepts. Specifically, we show how range modelling of many thousands of plant species can reveal spatial attractors for different growth-form combinations that represent biomes. Recent advances in data availability and analysis methods make this workflow realisable. A previous obstacle has been a lack of distribution data for sufficient species, but large global databases of plant occurrence records now exist. Together with advances in species distribution modelling (SDM) and increases in computational capacities, we can now characterise the niches of large numbers of plant species (Evans *et al.*, 2016). Moreover, global databases on plant traits have been compiled in recent years, allowing us to combine the modelled ranges of thousands of species with growth-form information to infer a site's suitability for different growth forms. The biomes that result from such a procedure are the result of convergent response of plant growth form to selection by climate and soils. An advantage over DGVMs is that the biomes are based on many thousands of species-specific range models. This greatly reduces effects of misparameterising the environmental preferences of plant types.

Below, we first outline a general protocol for constructing deductive and repeatable biome maps. We then demonstrate two applications of the protocol. In the first application we construct biome maps for Africa under ambient and future climatic conditions to illustrate the applicability of our approach in global change research. In the second application we show how the scheme can be used to select study locations for comparative research and meta-analyses.

Protocol

Fig. 1 provides an overview of the main steps involved in our protocol. First, we fit a niche model to species distribution data and project the potential ranges of each species in the study region (Fig. 1a–c). Second, we use the potential ranges of all species belonging to a growth form to calculate the preference of that growth form for each grid cell (geographical location) (Fig. 1d). Finally, we calculate a growth-form spectrum for each grid cell, thereby generating a grid cell by growth-form preference matrix (Fig. 1e), which in turn is used to classify the grid cells into biomes (Fig. 1f).

Niche projections for each species

A first step is to obtain distribution data for a representative sample of species in the study region (Fig. 1a). For instance, in the following example we used *c.* 23 500 African species for which we had growth-form data and sufficient distribution data to fit an SDM. It is important to obtain distribution data for a sufficient sample of species from each growth form used in a study (see next section) and to sample species from different floristic regions (e.g. deciduous trees from East and West African savannas). The latter allows the effects of biogeographical differences in how species of the same growth form respond to environmental drivers to be included in the analysis.

A variety of potential data sources exist for obtaining distribution data. Herbarium records provide information on spatial plant occurrences for large sets of species, and many herbaria have contributed their data to GBIF (www.gbif.org). Another source is vegetation plot data, which include presence–absence or abundance data for species from plot-based vegetation surveys. Several plot databases exist, such as sPLOT (Bruelheide *et al.*, 2019) or the European Vegetation Archive (EVA) (Chytrý *et al.*, 2016). The most comprehensive global database is BIEN (http://bien.nceas.ucsb.edu/bien/), which combines plant observations from herbarium, plot and trait records. Depending on the data source, cleaning the data is often necessary to deal with nomenclature changes, records from subspecies, ecotypes and varieties, and potential errors in the location information attached to the species records. Tools for facilitating such data cleaning exist (e.g. Zizka *et al.*, 2019).

The distribution data are then used to fit an SDM (Fig. 1b) that projects the potential range of each species (Fig. 1c). To fit SDMs, researchers can choose from a diversity of modelling approaches that range from process-based to correlative (Dormann *et al.*, 2012). In this context, it is notable that process-based niche modelling of many species is no longer limited by the availability of physiological or demographic parameters, as these can now be estimated for large numbers of species using hierarchical and
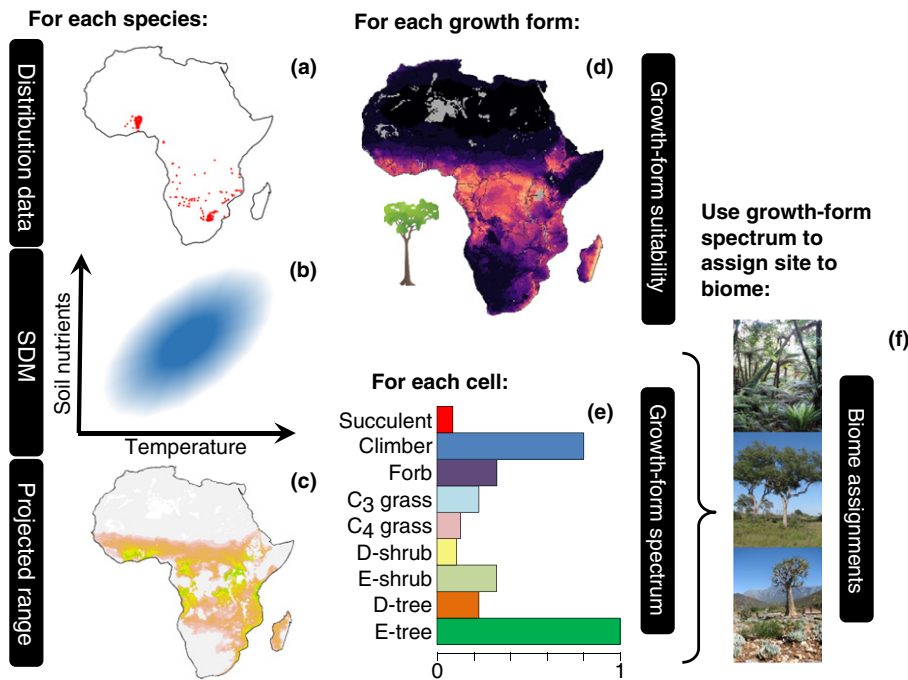
For each species:

(a)

Distribution data

(b)

SDM

Soil nutrients

Temperature

Projected range

(c)

For each growth form:

(d)

Growth-form suitability

Use growth-form spectrum to assign site to biome:

For each cell:

(e)

Succulent
Climber
Forb
C_3 grass
C_4 grass
D-shrub
E-shrub
D-tree
E-tree

0   1

Growth-form spectrum

(f)

Biome assignments

Fig. 1 Overview of the proposed workflow to construct biomes as described in the 'Protocol' section. Species distribution data (a) are used to fit a species distribution model (SDM; b) that enables projecting the potential range of a species in geographical space (c). Stacking the potential ranges of many species of a growth form reveals the suitability of grid cells for this growth form (d). Each cell's suitability for various growth forms is calculated (e), and cells are classified into biomes based on their suitability for growth forms (f). In (e) the abbreviations D- and E- indicate deciduous and evergreen, respectively.

inverse modelling techniques (Hartig *et al.*, 2011; Evans *et al.*, 2016). SDMs differ in how they use environmental information to project the ranges of species. This obviously has consequences for the biome maps being produced. Contrasting biome maps generated with models based on different algorithms or assumptions allows one to explore hypotheses about the controls over plant species distributions and the consequences for biome assembly. It also enables sensitivity analysis with respect to model selection and to produce model-ensemble biome maps. Irrespective of which SDM is used, the goal is to infer where a species could grow, that is its potential range, not its realised range. This is because in the next step, we stack the potential ranges of all species of a growth form to create suitability surfaces for each growth form. These surfaces represent spatial attractors for each growth form. Because the goal is to delimit potential rather than realised ranges, overfitting of SDMs is not desired because this might underestimate potential ranges and biases inferred attractors towards the spatial density of distribution records. It is important that methodical choices and underlying assumptions are reported in order to make the resulting biome maps reproducible.

Growth forms and their preferences

The next step in the protocol is to create suitability surfaces for each growth form. This is done by merging the species range projections with growth-form information. Because biomes can be defined as large-scale vegetation units characterised by their growth-form spectra, this merge provides the link between the ranges of species and biomes. Information on growth forms is included in floras, although the growth-form definitions do vary with flora. Databases such as BIEN and TRY (Kattge *et al.*, 2011) contain information on growth forms. There are often several entries per species and some of these may yield inconsistent information (e.g. 'shrub' vs 'tree'). A

pragmatic approach may be required, such as using the most frequent entry, or creating intermediate categories (e.g. 'shrub or tree'), depending on the objectives of the study. Broad growth-form categories (e.g. tree, grass) can be combined with other traits such as whole plant vegetative phenology or photosynthetic pathway to obtain finer and ecologically more relevant growth-form categories (e.g. deciduous vs evergreen trees, C_3 vs C_4 grasses). Phenological information is available for many species in BIEN and TRY, and databases on specific functional traits exist, such as the global database of C_4 photosynthesis in grasses (Osborne *et al.*, 2014). Missing functional trait data for individual species can be imputed using a variety of methods, such as taking the most frequent value in that genus or family, or using more sophisticated imputation methods (for a comparison of methods see Penone *et al.*, 2014).

Researchers have to decide which and how many growth forms they want to use to characterise biomes. This is an important decision, because it will affect the number of biomes and the extent to which they can be interpreted as vegetation units with similar functional responses to environmental forcing. The strength of our approach is that the resulting biome maps will reflect the researcher's hypotheses regarding the physiognomic attributes that discern vegetation types, and the plant growth forms that are functionally important.

Once all species are assigned a growth form, a suitability score for each growth form is computed for each grid cell. This suitability score is calculated as the proportion of species of each growth form that can grow in that grid cell. For example, if 200 out of 1000 evergreen trees and 400 out of 800 deciduous trees can grow in a grid cell, the grid cell has suitability scores of 0.2 and 0.5 for evergreen trees and deciduous trees, respectively. We can thus create suitability surfaces for each growth form (Fig. 1d) and characterise each grid cell by a growth form spectrum (Fig. 1e).

Classifying growth-form spectra into biomes

In the final step, the suitability scores for each growth form are used to create a grid cell by growth-form preference matrix. This matrix is then classified into groups, which are interpreted as biomes (Fig. 1f). The resulting biomes can be conceptualised as attractors for different growth-form combinations. Researchers can choose between a variety of clustering algorithms and distance metrics to identify the different biomes. In addition, methods exist to determine the optimal number of clusters/biomes (Borcard *et al.*, 2011). However, the final decision on the number of clusters will also be influenced by the level of abstraction required to answer the research question of a study.

As in any biome concept, the number of clusters/biomes is an arbitrary decision. There is no correct number of biomes. As we have alluded to above, biomes are simplifications of reality that help us to organise our knowledge about the Earth's vegetation and can serve for hypothesis testing. It is the level of detail required by users of a biome map that really determines what the optimal number of biomes is. For example, a computer algorithm may say that 2 is the optimal number of clusters, and discriminate between vegetated land and deserts. Few would find such a biome map useful. A strength of our method is that the growth-form preference matrix can be used to flexibly determine the desired number of clusters. Therefore, we advocate that adopters of our approach should make these matrices available. If others agree that the environmental preferences of growth forms revealed by these matrices are useful, they can use the matrices to make their own biome maps that are tailored to their research.

Sensitivity analyses

The suitability surfaces are at the core of the proposed workflow and their robustness needs to be evaluated. They are the aggregate output of hundreds to thousands of species-specific SDMs. Aggregating the output of many SDMs means that only models with high predictive accuracy should be used in the aggregation to avoid the propagation of prediction errors (Thuiller *et al.*, 2019). Metrics for evaluating the predictive accuracy of SDMs exist (e.g. Allouche *et al.*, 2006) and should be incorporated in the workflow. In addition, SDMs differ in how they use environmental information to project the ranges of species. As a consequence, there is often disagreement between model predictions (Cheaib *et al.*, 2012), which means that model-based uncertainties in aggregate outputs such as the suitability surfaces can be substantial and should be assessed (Thuiller *et al.*, 2019).

Adopters of the approach also need to ensure that the amount, coverage and quality of their distribution and trait data is sufficient. For example, researchers should check that different biogeographical zones are adequately represented in their sample of distribution records. In addition, how well the growth-form imputation works should be assessed. Moreover, the number of species per growth form should be sufficient to adequately estimate the growth-form suitability surfaces. This can be tested for each growth form by taking increasingly larger subsamples of the range models, computing suitability scores based on each subsample, and

calculating the correlation with the suitability scores based on the full set of species. If the correlation is already around 1 when smaller subsets are used, this would indicate that the suitability surfaces would not change if more species than in the full set were included. If the correlation approaches 1 only when a very large subset is used, this would indicate that even the full set may not be enough to accurately estimate suitability surfaces for that growth form.

Making a biome map

Here we illustrate the framework by constructing a biome map for Africa. The biome map, the underlying grid cell by growth-form preference matrix and R code to produce the map from the matrix are available from https://doi.org/10.5061/dryad.4j0zpc87t.

Distribution data and growth-form information

First, we construct a biome map for ambient conditions following the protocol outlined above. We extracted distribution data of all African plant species available in the BIEN database v.4.1, using the BIEN R package (Maitner *et al.*, 2018). Distribution records flagged as non-native or cultivated were also used. Only species with records in seven or more different 1-km grid cells in Africa were retained. This resulted in an initial sample of *c.* 25 700 plant species.

We grouped the species into nine growth forms: evergreen and deciduous trees, evergreen and deciduous shrubs, C_3 and C_4 grasses, forbs, succulents, and climbers. We propose that these nine growth forms are adequate to characterise biomes at the continental scale and expected them to have contrasting functional responses to the environment. We classified all species belonging to the families Poaceae, Cyperaceae, Juncaceae, Flagellariaceae and Restionaceae as grasses, and used the database of Osborne *et al.* (2014) to identify grass species with C_4 photosynthetic pathway. For all other species we extracted information on whole plant growth form and whole plant vegetative phenology from BIEN. If multiple entries with contrasting information were available for a species (e.g. a species had entries as a shrub and a tree), we used the most frequent entry. If species-level trait information was not available, we used the most frequent entry for the genus, and if that was also not available, we took the most frequent entry for the family. Species classified in BIEN as herb, geophyte or fern were assigned to the forb category, and species classified as epiphyte, liana or vine were assigned to the climber category. For 1000 species we could not assign a growth form, leaving us with *c.* 24 700 species. References for BIEN occurrence records and growth-form data can be found in Supporting Information Table S1.

Range projections

For each of the 24 700 species we modelled its potential range with two SDMs, Maxent (Phillips *et al.*, 2006) and the TTR-SDM (Higgins *et al.*, 2012). Maxent is a frequently used correlative SDM (Guillera-Arroita *et al.*, 2015). The TTR-SDM is a physiologically based SDM for plants that is closer to the process-end of the correlative–process-based continuum described by Dormann *et al.*

(2012). The TTR-SDM model is based on Thornley's transport resistance model (hence 'TTR', Thornley, 1998). The TTR model is a series of dynamic equations that represent how resource assimilation (carbon and nitrogen in this case), plant growth and allocation between resource sinks and sources influence the biomass accumulation of an individual plant. The TTR-SDM as described in Higgins *et al.* (2012) includes a series of functions that simulate how environmental factors (light, temperature, soil water, soil nitrogen) influence key processes in the TTR model. In this contribution, we used a new variant of the TTR-SDM that includes a Farquhar-type (Farquhar *et al.*, 1980) photosynthesis model. This allows the model to formally link ecophysiological knowledge of how photosynthesis is co-limited by light, temperature and atmospheric CO_2 concentration and how this co-limitation differs for C_3 and C_4 plants. The TTR-SDM thus explicitly considers how C_3 and C_4 plants may respond to temperature and atmospheric CO_2 concentrations. Therefore, the TTR-SDM will project shifts in ranges with elevated CO_2, whereas other SDMs will not. As in DGVMs, photosynthesis is parameterised by leaf-level observations. That is, we do not infer Farquhar photosynthesis parameters from distribution data. A more detailed description of the TTR-SDM is provided in Methods S1.

The environmental data (at 1-km resolution) used for fitting both SDMs were defined by the requirements of the TTR-SDM: minimum, mean and maximum monthly temperatures (Hijmans *et al.*, 2005), solar radiation (Trabucco & Zomer, 2010), soil nitrogen content (Shangguan *et al.*, 2014) and monthly soil moisture contents (Trabucco & Zomer, 2010). The soil moisture data from Trabucco & Zomer (2010) represent soil moisture available for evapotranspiration and are not influenced by the biome state of a grid cell. Both SDMs used the same presence and pseudo-absence points. The TTR-SDM further assumed an atmospheric CO_2 concentration of 338 ppm (Meinshausen *et al.*, 2011). Methods S1 provides more information on model fitting. After discarding models that did not converge or that had low predictive accuracy (True Skill Statistic (TSS) < 0.7), we were left with *c.* 23 500 (TTR-SDM) and 23 200 species (Maxent).

Growth-form preferences and classification to biomes

We used the modelled potential ranges of the species to calculate suitability scores for each growth form in each grid cell. As described above, the suitability score is simply the proportion of species of each growth form that can grow in a grid cell according to the SDM used. Fig. 2 shows the suitability surfaces based on the TTR-SDM. The Maxent SDMs showed a tendency to overfit, as revealed by lower false positive rates (Fig. S1), and the resulting suitability surfaces (Fig. S2) appear to be influenced by the density of distribution records (Fig. S3). This problem could be reduced by cross-validation procedures.

We conducted a number of sensitivity tests that confirmed the robustness of the TTR-SDM suitability surfaces to the TSS threshold used, to the distributions of growth forms and biogeographical regions in the occurrence records, and to the imputation of growth-forms (see Methods S2 and Figs S3–S7).

The suitability scores for each growth form (each ranging from 0 to 1; see definition earlier) were then used to create a grid-cell by growth-form preference matrix and this matrix was then classified in groups (i.e. biomes) using the *clara* algorithm (Kaufman & Rousseeuw, 2009). *Clara* is a clustering algorithm that partitions data into *k* groups. Unlike *k*-means algorithms, where the clusters are defined by the mean value of all data points in a cluster, the *clara* clusters are defined by the medoid, which is the existing data point within the cluster with the lowest dissimilarity to other points in the cluster. Using medoids instead of means makes the method less sensitive to outliers.

We used average silhouette widths to determine the optimal number of clusters (biomes). The four best clustering solutions for the TTR-SDM suitability scores were two, five, four and three biomes, respectively, and are shown in Fig. S8. Here we present the fifth best clustering solution with 14 biomes because we consider this an appropriate level of abstraction for a continental-scale map and for discussing how our biome constructs are to be interpreted. The same number of clusters was used for the Maxent biome map. We refined cell assignment to the clusters using discriminant analysis of principal components (DAPC; Jombart *et al.*, 2010). This method optimises the separation of cells into predefined groups (the 14 *clara* clusters) based on discriminant functions of principal components analysis (PCA) components. The PCA components were computed from the growth-form preference matrix. Our DAPC used nine PCA components and nine discriminant axes.

Interpretation of the biome map

Fig. 3(a) shows the emergent biomes based on the TTR-SDM suitability surfaces. Table 1 shows the differences between the biomes in terms of the suitability for different growth forms. A biome map based on the Maxent suitability surfaces is shown in Fig. S9.

It is useful to look at Fig. 2 and Table 1 when interpreting the modelled biomes (Fig. 3a). The dark red biome (biome 1) is characterised by extremely low suitability for all growth forms and corresponds to Africa's desert areas. Light and dark grey are arid biomes (13 and 14) adjacent to Mediterranean-type climate areas (e.g. the Succulent and Nama Karoo in South Africa and steppe in northern Africa) and are characterised by a higher suitability for succulents (especially in southern Africa), a few shrubs and grasses. These arid biomes are discerned from arid biomes closer to the equator (e.g. in the Sahel and Horn of Africa) coloured in lighter red shades (2 and 3), which tend to attract shrubs, grasses and deciduous trees more than succulents. In green and yellow (4–8) are biomes with high suitability for grasses (especially C_4) and trees. These biomes correspond to different types of savanna and seasonal forest. The fact that all these biomes have relatively high suitability for trees and grasses means that there is potential for alternative states to occur, albeit disturbance-maintained alternative states (Bond *et al.*, 2005) cannot be modelled with our approach. The dark green biome (8) already has a higher suitability for evergreen trees and occurs close to the wettest biomes (light and dark blue; 9 and 10), which are most suitable for evergreen trees and climbers,
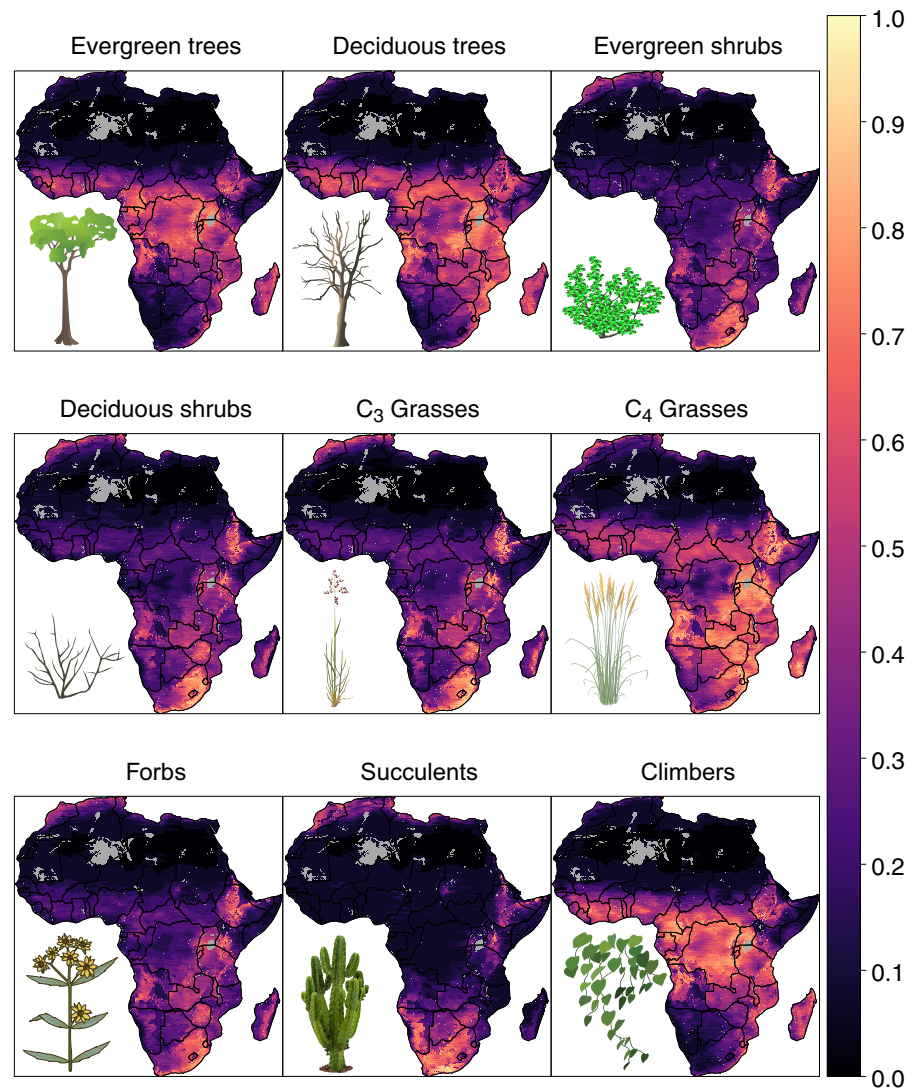
Fig. 2 Environmental suitability for plant growth forms. Suitability is expressed as the proportion of species of each growth form that can grow in a cell according to the TTR-SDM. No predictions were made for the grey areas because of incomplete environmental data.

but have a low suitability for other growth forms. In particular, the dark blue biome (10) is only suitable for a lower number of species and these are mostly evergreen trees and climbers. This biome occurs under the wettest and hottest conditions. Note that the rather low overall suitability of this biome does not necessarily imply that it is species-poor. This biome just has a low suitability for most species that are not evergreen trees or climbers.

The spatial arrangement of the modelled biomes, although differing in many ways from White (1983) or Olson *et al.* (2001) (Fig. 3b,c), does capture many widely recognised vegetation patterns in Africa. For example, the sequence from the Sahara to the Congo basin is often shown in vegetation maps of Africa, as is the sequence in Madagascar from wet tropical forest on the east coast to grass-dominated savannas in the central highlands to woody savannas and seasonal forest on the west coast (e.g. Schultz, 2005; Pfadenhauer & Klötzli, 2014). Similarly, the savanna intrusions into the wet forests in southern Togo and neighbouring countries and into the Ethiopian highlands are also emulated.

Although it is tempting to compare our biomes with existing biome maps of Africa, one needs to bear in mind the different biome concepts that underlie these different maps. Most importantly, our biome constructs represent environmental attractors for growth-form combinations. That is, they show where different growth-form combinations are likely to assemble based on environmental conditions. As such, they do not always show the locally dominant growth form. For instance, disturbances such as fire, herbivory or human land use may push realised growth-form combinations away from the modelled attractors.

For example, although Mediterranean shrub-dominated ecosystems with winter rainfall (i.e. South Africa's Fynbos and the Maghreb coast) do appear on our biome map (biome 12), our classification does not distinguish them from upland grasslands with summer or year-round rainfall (e.g. South Africa's Highveld). The Highveld grasslands, for instance, are indeed climatically suitable for woody plants (Wakeling *et al.*, 2012), as the TTR-SDM correctly predicts. However, the model does not consider that the relatively short growing season confines woody plants to fire-
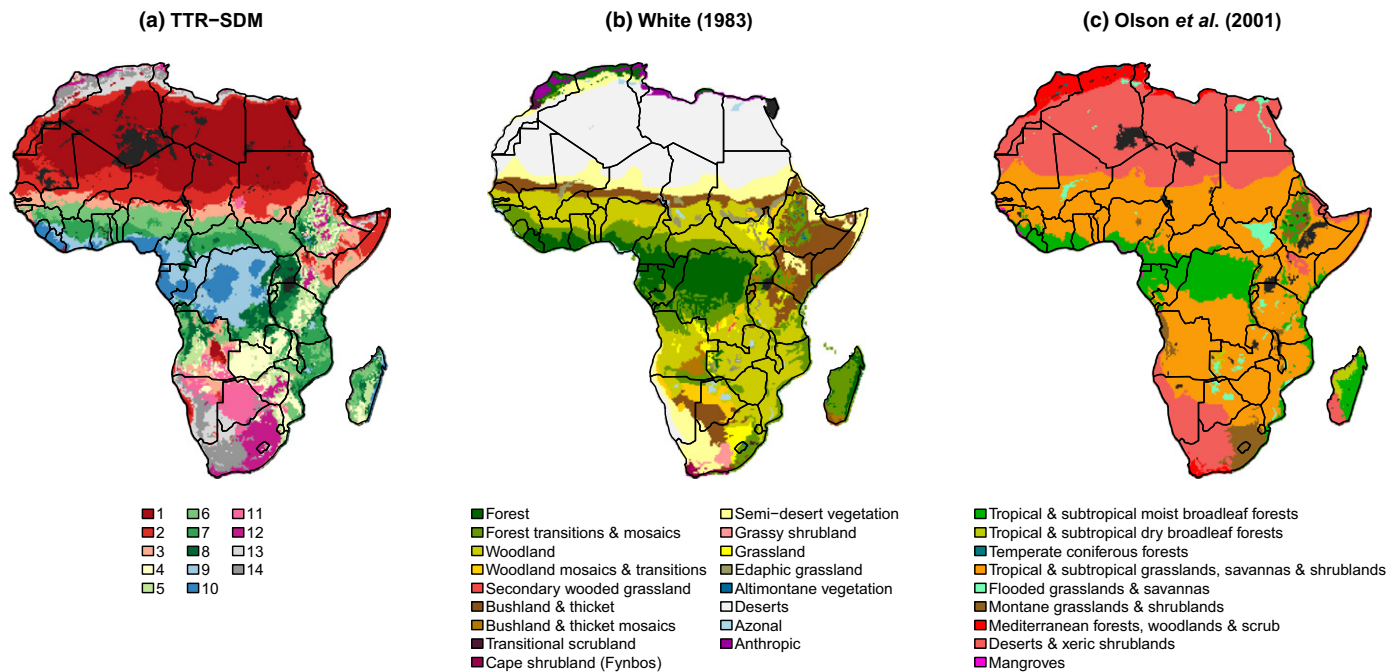
Fig. 3 Biome concepts for Africa according to (a) environmental attractors for growth form combinations estimated using the TTR-SDM applied to *c.* 23 500 plant species (this study), (b) White (1983) and (c) Olson *et al.* (2001). Black cells have no data.

Table 1 Mean suitability scores for plant growth forms in African biomes.

Biome	Tree-E	Tree-D	Shr-E	Shr-D	C_3-G	C_4-G	Forbs	Succ	Climb
1	0.00	0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.00
2	0.02	0.05	0.04	0.06	0.04	0.14	0.05	0.05	0.03
3	0.11	0.22	0.09	0.13	0.12	0.30	0.10	0.05	0.13
4	0.28	0.41	0.31	0.33	0.33	0.55	0.33	0.19	0.32
5	0.37	0.39	0.43	0.45	0.51	0.56	0.46	0.23	0.40
6	0.22	0.38	0.13	0.18	0.18	0.40	0.15	0.04	0.27
7	0.37	0.49	0.18	0.21	0.22	0.44	0.20	0.04	0.44
8	0.52	0.55	0.28	0.29	0.31	0.53	0.32	0.08	0.57
9	0.46	0.41	0.16	0.14	0.18	0.29	0.18	0.01	0.53
10	0.32	0.24	0.10	0.07	0.12	0.13	0.11	0.01	0.37
11	0.10	0.22	0.22	0.23	0.20	0.41	0.22	0.25	0.12
12	0.20	0.24	0.52	0.56	0.57	0.47	0.49	0.45	0.21
13	0.03	0.07	0.15	0.15	0.15	0.19	0.15	0.27	0.04
14	0.07	0.12	0.32	0.29	0.31	0.30	0.29	0.47	0.08

The biome numbers are the same as in Fig. 3(a). Standard errors of the means were always < 0.01. Tree-E, evergreen trees; Tree-D, deciduous trees; Shr-E, evergreen shrubs; Shr-D, deciduous shrubs; C_3-G, C_3 grasses; C_4-G, C_4 grasses; Succ, succulents; Climb, climbers.

protected topographic positions such as river valleys and gorges (Wakeling *et al.*, 2012).

Projecting biome shifts

In this section, we illustrate how our biome protocol can be used to project future biome changes in Africa. Warming and precipitation change in Africa are predicted to be stronger than the global mean, and novel climates without present-day analogues will develop in

parts of the continent (Williams *et al.*, 2007). In addition, the elevated atmospheric CO_2 concentrations are likely to shift Africa's vast savannas into more woody states because woody plants with C_3 photosynthetic pathway benefit more from elevated CO_2 concentrations than C_4 grasses (Bond & Midgley, 2012; Higgins & Scheiter, 2012). As stated in the Introduction, our method is suitable for projecting biome changes because the modelled biomes are the outcome of the individualistic response of thousands of plant species to climate and soils. Moreover, the biome boundaries are not defined by climatic thresholds, and the method does not assume that biomes are units that will shift. Rather, the biome shifts that our protocol projects are the outcome of the potential range shifts of thousands of species. This makes the method robuster than a PFT-based DGVM that projects biome shifts on the basis of a type-specimen for a PFT.

Model projection

We used the fitted range models to project where the species will be able to grow under future climatic conditions. The future projections were made separately based on the TTR-SDM and Maxent models. We used three climate scenarios for the year 2070 predicted by the MPI-ESM Global Circulation Model (GCM) of the Max Planck Institute for Meteorology (Giorgetta *et al.*, 2013) under the 2.6, 4.5 and 8.5 Representative Concentration Pathways (RCPs) (van Vuuren *et al.*, 2011). The climate data for these scenarios were downloaded from Hijmans *et al.* (2005). Total soil nitrogen and solar radiation were assumed to be the same in 2070 as under ambient conditions. Future monthly soil moisture contents were modelled using the equations in Trabucco & Zomer (2010) and 2070 climate data from the MPI-ESM GCM from Hijmans
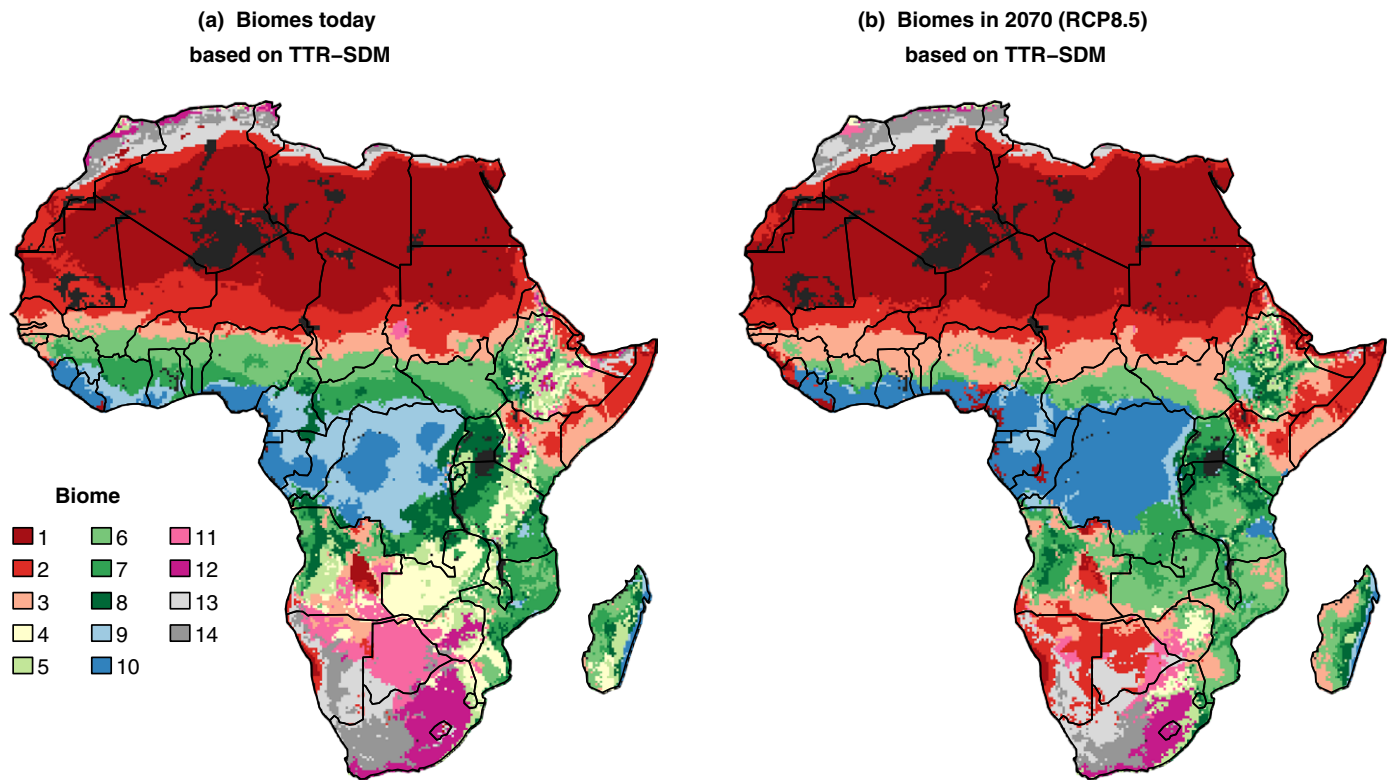
Fig. 4 Current and future biome distributions based on basins of attraction for plant growth forms estimated using the TTR-SDM applied to *c.* 23 500 plant species. (a) Biomes under ambient environmental conditions. (b) Projected biome distribution in 2070 under the RCP8.5 emission scenario. Black cells have no data.

et al. (2005). Again, the soil moisture values represent soil moisture available for evapotranspiration and are not influenced by land cover or biome type. For the modelling horizon 2070 the TTR-SDM projections assumed atmospheric CO_2 concentrations of 438, 524 and 677 ppm for RCPs 2.6, 4.5 and 8.5, respectively (Meinshausen *et al.*, 2011).

We then calculated growth-form suitability scores for each grid cell based on the future ranges of all 23 500 species. These future suitability scores were then used to predict the biome membership of each grid cell based on the discriminant functions derived from the DAPC above. This allowed us to assess the biome state of each cell in 2070 (Fig. 4) and to assess uncertainty in biome-shift predictions originating from RCPs and SDM choice (Fig. 5). We used only one GCM for the purpose of illustration, because GCM is a minor source of uncertainty in comparison to RCP and SDM (Thuiller *et al.*, 2019). However, empirical applications of the approach should also evaluate uncertainty originating from GCM choice when making future projections.

Interpretation of projected biome shifts

The future projection based on the TTR-SDM for RCP8.5 (Fig. 4b) suggests a profound transformation of African biomes that is in agreement with theoretical expectations, observations and previous modelling studies. For instance, our model predicts that savannas with high suitability for C_4 grasses (yellow and light green; 4–5) will transition to a more woody state (6–7). This is consistent with the prediction that woody C_3 plants will benefit relatively

more from elevated atmospheric CO_2 concentrations than C_4 grasses (Bond & Midgley, 2012; Higgins & Scheiter, 2012; Conradi, 2018). Woody encroachment is already ongoing in African savannas and cannot be easily explained by a lack of fire or herbivory, suggesting that CO_2 is a chronic driver of encroachment (Buitenwerf *et al.*, 2012). Another marked change is the expansion of highly unsuitable biomes (1–3) north of the Congo basin that are projected to replace the savannas of that region (6–7) despite increasing rainfall in this region (IPCC, 2014). One reason for this may be that gains in rainfall are going to be offset by increasing temperatures. The red biome (1) will also appear locally in the Congo basin and West African rain forests. This illustrates that our biomes should not be interpreted in terms of climate (e.g. that these will become arid locations), but in terms of their suitability for species from different growth forms. Here, it means that only a low proportion of evergreen tree and climber species will be able to grow in these locations in the future. A further major change is theexpansion of the dark blue 'rain forest' biome (10; a biome with low suitability for all growth forms except evergreen trees and climbers) in the Congo basin and its disappearance in Madagascar. This is in line with the projected increasing rainfall in the Congo basin and decreasing rainfall in Madagascar (IPCC, 2014).

Selecting sites for comparative research

Predictive ecology is hampered by the fact that ecological dynamics are influenced by a mixture of biophysics and evolutionary and ecological history. These historical factors conspire to make the
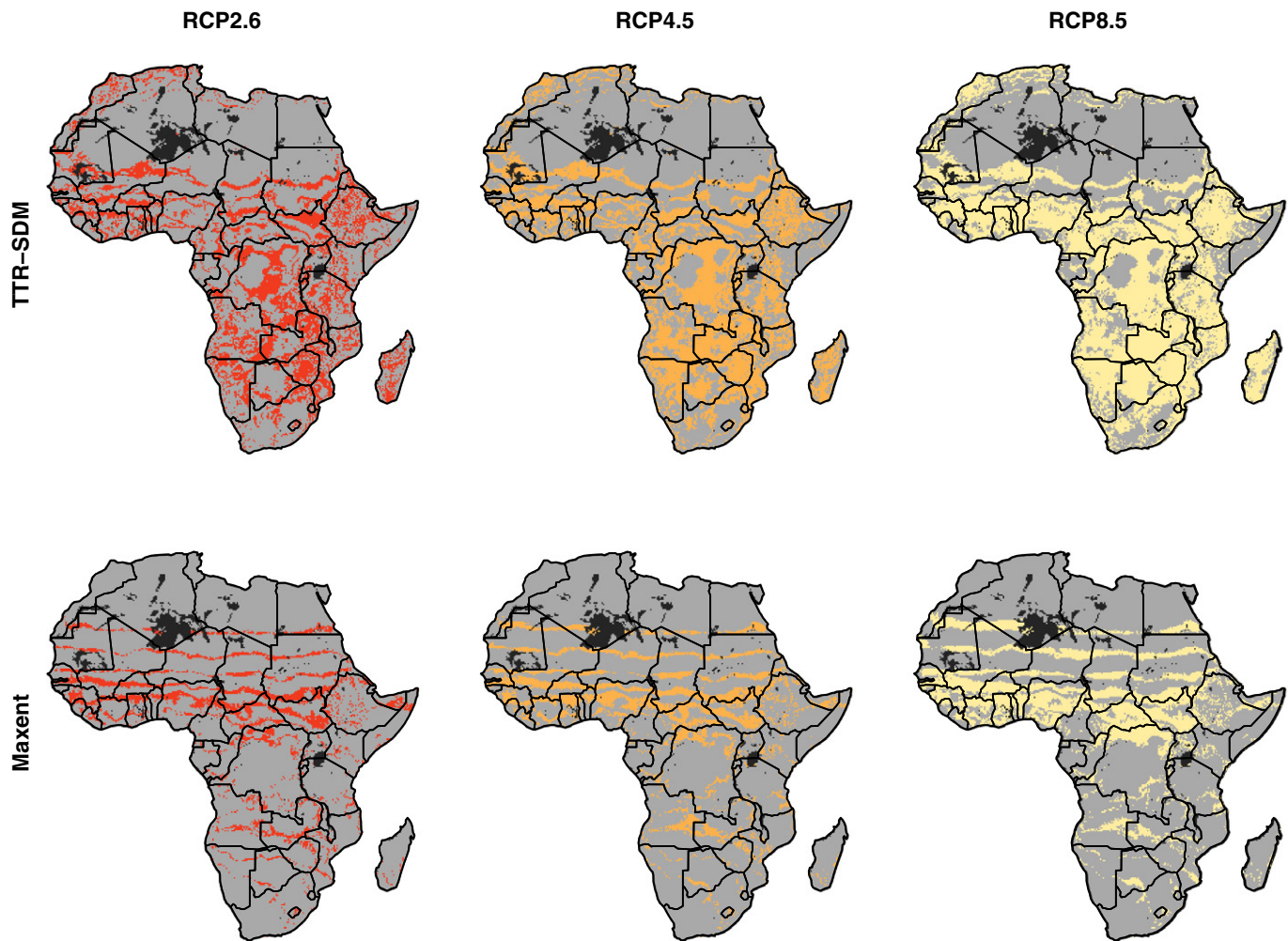
Fig. 5 Projected biome shifts in Africa in 2070 based on different range models (TTR-SDM and Maxent) and different Representative Concentration Pathways (RCPs). Pixels in colour indicate biomes that will shift in state, warmer colours indicate higher certainty of change (i.e. they will change under a greater proportion of RCP scenarios). The TTR-SDM considers how changes in temperature, soil moisture and atmospheric CO_2 concentration influence species ranges, and thus biome distributions, whereas Maxent is not sensitive to CO_2.

dynamics of each ecological system unique. To make progress in interpreting the unique dynamics of ecosystems, ecologists compare estimates of ecological rates measured in different ecosystems, and in different instances of the same ecosystem at different geographical locations, for example using meta-analyses or globally distributed experiments. However, this endeavour relies on an appropriate a priori method for assigning an ecosystem to types, such as biome types.

Here we demonstrate how our biome map protocol can be used to identify sites for comparison. For example, should a researcher wish to compare their measurements made near Skukuza in the Kruger National Park, South Africa, to other similar sites in Africa, they could use the cell by growth-form matrix to find grid cells in Africa that have similar growth-form spectra to Skukuza. Fig. 6 shows that we would choose different grid cells for this comparison if our selection were based on growth-form similarity rather than environmental similarity. We propose that for comparative research it is more appropriate to select study locations based on

growth-form spectra, because sites with similar growth-form spectra represent similar environmental attractors.

Alternatively, the researcher could be more explicit and ask for all sites with a high suitability for deciduous trees and shrubs, high suitability for C_4 grasses and low suitability for succulents. It is up to the researcher to define what they regard as meaningful units for comparison, and our protocol provides a flexible means to delimit such units.

Outlook

This study proposes and demonstrates a protocol for creating biome maps. The protocol provides a data-driven pipeline for defining the environmental preferences of the growth forms that define biomes, and for delimiting land surface units that have internally similar basins of attraction for different growth form combinations. This contrasts with previous methods that are based on the idea of a type-specimen or benchmark-site for a biome (e.g.
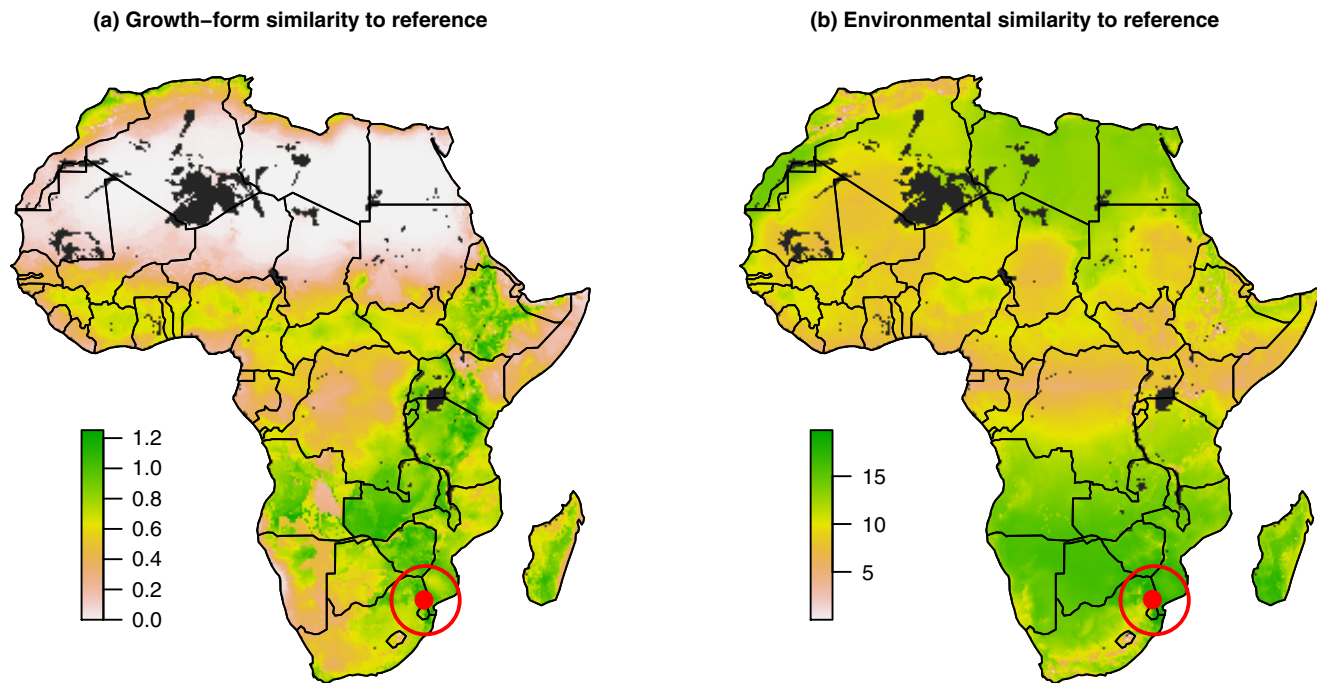
(a) Growth–form similarity to reference   **(b) Environmental similarity to reference**

Fig. 6 Similarity of African grid cells to Skukuza in the Kruger National Park (a) in terms of their growth-form spectra and (b) environmental conditions. The Skukuza grid cell is shown with the red dot. Similarity is expressed in Euclidean distances, calculated from (a) the growth-form spectra and (b) the (scaled) monthly input variables of the TTR-SDM. In (b), a slight discontinuity occurs in equatorial cells. This is because for the calculation of similarity, we changed the order of the monthly environmental variables of northern-hemisphere cells to match the seasons of the southern hemisphere. This is necessary to calculate environmental similarity based on the conditions the plants in both hemispheres perceive during their respective growing seasons, but introduces a small artefact at the Equator.

Olson *et al.*, 2001; Friedl *et al.*, 2010). Such approaches are reliant on the type specimens or benchmarks being appropriately defined. In our protocol the biomes emerge as a mechanical classification of the growth-form spectra of sites. That is, the user makes decisions on which growth forms are to be used in the classification, but the resulting biomes and their distributions emerge from the analysis.

The protocol uses range models for thousands of species to characterise the environmental preferences of each growth form. We hypothesise that this model averaging makes projections of biome shifts more robust and avoids a problem PFT-based DGVMs have with biome shift projections. DGVMs are important tools for making predictions about ecosystem change under future climates because they use physiological sub-models to scale from plant-level rates to ecosystem rates and to biome distributions (Prentice *et al.*, 2007). However, instead of modelling individual species, PFT-based DGVMs represent the taxonomic and functional diversity of plants using a small number of PFTs (Lavorel *et al.*, 1997). The spatial distribution and abundance of the PFTs is then modelled, and vegetation types are delimited based on the modelled relative dominance of different PFTs (Bonan *et al.*, 2002). That is, in such DGVMs the projections are reliant on the PFTs characterising a biome being correctly conceptualised and parameterised. For example, DGVM projections of the likelihood of Amazon forest die-back (Cox *et al.*, 2004; Rammig *et al.*, 2010) are reliant on the broad-leaf evergreen PFT being correctly conceptualised and parameterised. Even though current-generation DGVMs include more PFTs (Sitch *et al.*, 2008) or consider

variation in the traits that defined PFTs (Scheiter *et al.*, 2013), this problem that functional diversity is often under-represented in DGVMs remains. Our approach circumvents this problem by parameterising thousands of species separately and letting them assemble to biomes. Of course, DGVMs include a higher level of biophysical realism than SDMs. Both methods have strengths and weaknesses and using both methods is advocated to better evaluate uncertainty in biome shift projections.

Our approach mimics a Gleasonian view of an ecological community. The Gleasonian view is that communities are manifestations of the environmental preferences of individual species and not of the biotic linkages between species. Similarly, our biomes are manifestations of the environmental preferences of individual growth forms and contain no ecological hypothesis of why these growth forms co-occur and how they interact with one another. That is, the effects of biotic interactions and disturbance (fire, herbivory, land use) are either in the error terms of the statistical model or included (misspecified) in the parameters which describe the physiological preferences of species. This problem is well known (Wisz *et al.*, 2013), but its significance is difficult to assess.

Some features of our protocol are dependent on the SDM used. The TTR-SDM, for example, allows physiological processes to control the projected distribution of species. That is, the model has a higher level of process realism than purely correlative models and this, in principle, should reduce the potential for over-fitting (Higgins *et al.*, 2012). In this study we further reduced the potential for over-fitting by using a Farquhar-style (Farquhar *et al.*, 1980)

model to characterise how carbon gain is co-limited by temperature, light and atmospheric CO_2 concentration. That is, the model uses data derived from studies of leaf-level gas exchange to constrain the species distribution projections. We emphasise however that the general principle of the protocol can be realised using other species distribution models.

As we argued in the Introduction, different biome maps may be appropriate for different purposes. Indeed in the section 'Selecting sites for comparative research' we suggested that the distance between the growth-form spectra of grid cells can be used to flexibly group cells into groups for meta-analyses. The authors of such meta-analyses of course do not wish to perform analyses as computer-intensive as our protocol, but producers of such biome map products could and should make the cell by growth-form preference matrices available, which would provide a portable method for meta-analysis authors to create their own biome categories.

One could ask what Schimper would have done if the wealth of today's biodiversity data was available to him. Would he have sought to describe continuous variation in vegetation rather than simplifying the Earth's vegetation into 'like' subsets? Some authors find the idea of classification outdated in modern predictive science. We would argue that there is a place for both approaches. In fact, our suitability surfaces (Fig. 2) emphasise that vegetation varies continuously in space in response to environmental variation, and proponents of the 'gradient view' of vegetation may find these more useful products of our protocol than a biome map. In our opinion, there remains a place for classification into abstract constructs such as biomes because such constructs help us to organise the diversity of the Earth's vegetation. As an example, it is difficult to consider all the variation depicted in the nine panels of Fig. 2 simultaneously. And this is only one continent and nine growth forms. This variation is much easier for us to visualise when we represent it in units with internally similar structure, as in the biome map (Fig. 3a). Such a simplification allows us to apply knowledge gained in one instance of that biome to other instances, and forms a language for conversations with others working in similar systems. Biomes are also useful in ecosystem management. For example, if one applies a conservation treatment in localities with similar growth-form spectra, one would (often) rightly assume that the outcome of the treatment will be similar in all instances. The reason is that the boundaries between biomes are not arbitrarily drawn by the clustering algorithm (or an expert), but reflect nonlinear changes in the suitability surfaces (or vegetation physiognomy) (Fig. 2).

In summary, the protocol outlined here provides a data-driven method for mapping biomes and projecting biome shifts (for both fore- and hind-casting). This paper serves to articulate the concept and illustrate its feasibility. The next steps would be to develop global biome maps and make global biome-shift projections. Global analyses will require more careful consideration of the role of dispersal limitation for empirical species occurrence data as well as its role for projections of species ranges. Global analyses will also require different growth forms to those considered here. Perhaps more fundamentally, the concept of what we call growth forms in this paper needs to be expanded to include other sources of information such as phylogenetic information (e.g. conifer or nonconifer trees), life history information (e.g. annual, biennial or perennial forbs) and ecological information (e.g. rain green vs summer green deciduous, phenology of growth, frost tolerance, fire tolerance). Ecologists do not lack creativity when it comes to expanding classification schemes, but rather data will limit the feasibility of implementing such expanded schemes.

Author contributions

TC and SIH planned and designed the research. TC ran the models and analysed the data, with contributions from AHS and HN. TC wrote the manuscript with substantial contributions from JAS, GFM, HN, AHS and SIH.

ORCID

Timo Conradi (iD) https://orcid.org/0000-0003-2360-9284
Steven I. Higgins (iD) https://orcid.org/0000-0001-5695-9665
Guy F. Midgley (iD) https://orcid.org/0000-0001-8264-0869
Henning Nottebrock (iD) https://orcid.org/0000-0003-2238-2700
Andreas H. Schweiger (iD) https://orcid.org/0000-0003-2656-5918
Jasper A. Slingsby (iD) https://orcid.org/0000-0003-1246-1181

Timo Conradi[1]* (iD), **Jasper A. Slingsby**[2,3] (iD), **Guy F. Midgley**[4] (iD), **Henning Nottebrock**[1] (iD), **Andreas H. Schweiger**[1] (iD) and **Steven I. Higgins**[1] (iD)

[1]Plant Ecology, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Universitätsstraße 30, 95447 Bayreuth, Germany;
[2]Fynbos Node, South African Environmental Observation Network (SAEON), Centre for Biodiversity Conservation, 7735 Cape Town, South Africa;
[3]Department of Biological Sciences, Centre for Statistics in Ecology, Environment and Conservation (SEEC), University of Cape Town, 7701 Rondebosch, South Africa;
[4]Global Change Biology Group, Department of Botany and Zoology, Stellenbosch University, 7602 Stellenbosch, South Africa
(*Author for correspondence: tel +49 921552805; email timo.conradi@uni-bayreuth.de)

References

Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**: 1223–1232.

Bonan GB, Oleson KW, Vertenstein M, Levis S, Zeng X, Dai Y, Dickinson RE, Yang Z-L. 2002. The land surface climatology of the community land model coupled to the NCAR community climate model. *Journal of Climate* **15**: 3123–3149.

Bond WJ, Midgley GF. 2012. Carbon dioxide and the uneasy interactions of trees and savannah grasses. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367**: 601–612.

Bond WJ, Woodward FI, Midgley GF. 2005. The global distribution of ecosystems in a world without fire. *New Phytologist* **165**: 525–538.

Borcard D, Gillet F, Legendre P. 2011. *Numerical ecology with R.* New York, NY, USA: Springer.

Bruelheide H, Dengler J, Jiménez-Alfaro B, Purschke O, Hennekens SM, Chytrý M, Pillar VD, Jansen F, Kattge J, Sandel B et al. 2019. splot – a new tool for global vegetation analyses. *Journal of Vegetation Science* **30**: 161–186.

Buitenwerf R, Bond WJ, Stevens N, Trollope WSW. 2012. Increased tree densities in South African savannas: >50 years of data suggests CO_2 as a driver. *Global Change Biology* **18**: 675–684.

Cheaib A, Badeau V, Boe J, Chuine I, Delire C, Dufrêne E, François C, Gritti ES, Legay M, Pagé C et al. 2012. Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty. *Ecology Letters* **15**: 533–544.

Churkina G, Running SW. 1998. Contrasting climatic controls on the estimated productivity of global terrestrial biomes. *Ecosystems* **1**: 206–215.

Chytrý M, Hennekens SM, Jiménez-Alfaro B, Knollová I, Dengler J, Jansen F, Landucci F, Schaminée JH, Acic S, Agrillo E et al. 2016. European vegetation archive (EVA): an integrated database of European vegetation plots. *Applied Vegetation Science* **19**: 173–180.

Conradi T. 2018. Woody encroachment in African savannas: towards attribution to multiple drivers and a mechanistic model. *Journal of Biogeography* **45**: 1231–1233.

Cox PM, Betts RA, Collins M, Harris PP, Huntingford C, Jones CD. 2004. Amazonian forest dieback under climate-carbon cycle projections for the 21st century. *Theoretical and Applied Climatology* **78**: 137–156.

Crisp M, Arroyo M, Cook L, Gandolfo M, Jordan G, McGlone M, Weston P, Westoby M, Wilf P, Linder H. 2009. Phylogenetic biome conservatism on a global scale. *Nature* **458**: 754–756.

Donoghue M, Edwards E. 2014. Biome shifts and niche evolution in plants. *Annual Review of Ecology, Evolution, and Systematics* **45**: 547–572.

Dormann CF, Schymanski SJ, Cabral J, Chuine I, Graham C, Hartig F, Kearney M, Morin X, Römermann C, Schröder B et al. 2012. Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography* **39**: 2119–2131.

Echeverría-Londoño S, Enquist BJ, Neves DM, Violle C, Boyle B, Kraft NJB, Maitner BS, McGill B, Peet RK, Sandel B et al. 2018. Plant functional diversity and the biogeography of biomes in North and South America. *Frontiers in Ecology and Evolution* **6**: 219.

Evans MEK, Merow C, Record S, McMahon SM, Enquist BJ. 2016. Towards process-based range modeling of many species. *Trends in Ecology & Evolution* **31**: 860–871.

Farquhar GD, von Caemmerer S, Berry JA. 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta* **149**: 78–90.

Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X. 2010. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment* **114**: 168–182.

Giorgetta MA, Jungclaus J, Reick CH, Legutke S, Bader J, Böttinger M, Brovkin V, Crueger T, Esch M, Fieg K et al. 2013. Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the coupled model intercomparison project phase 5. *Journal of Advances in Modeling Earth Systems* **5**: 572–597.

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* **24**: 276–292.

Hartig F, Calabrese JM, Reineking B, Wiegand T, Huth A. 2011. Statistical inference for stochastic simulation models – theory and application. *Ecology Letters* **14**: 816–827.

Higgins SI, Buitenwerf R, Moncrieff GR. 2016. Defining functional biomes and monitoring their change globally. *Global Change Biology* **22**: 3583–3593.

Higgins SI, O'Hara RB, Bykova O, Cramer MD, Chuine I, Gerstner E-M, Hickler T, Morin X, Kearney MR, Midgley GF et al. 2012. A physiological analogy of the niche for projecting the potential distribution of plants. *Journal of Biogeography* **39**: 2132–2145.

Higgins SI, Scheiter S. 2012. Atmospheric CO_2 forces abrupt vegetation shifts locally, but not globally. *Nature* **488**: 209–212.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978.

IPCC. 2014. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**: 94.

Kattge J, Díaz S, Lavorel S, Prentice IC, Leadley P, Bönisch G, Garnier E, Westoby M, Reich PB, Wright IJ et al. 2011. Try – a global database of plant traits. *Global Change Biology* **17**: 2905–2935.

Kaufman L, Rousseeuw PJ. 2009. *Finding groups in data: an introduction to cluster analysis.* Chichester, UK: John Wiley & Sons.

Lavorel S, McIntyre S, Landsberg J, Forbes T. 1997. Plant functional classifications: from general groups to specific groups based on response to disturbance. *Trends in Ecology & Evolution* **12**: 474–478.

Lehmann CER, Anderson TM, Sankaran M, Higgins SI, Archibald S, Hoffmann WA, Hanan NP, Williams RJ, Fensham RJ, Felfili J et al. 2014. Savanna vegetation–fire–climate relationships differ among continents. *Science* **343**: 548–552.

Maitner BS, Boyle B, Casler N, Condit R, Donoghue J II, Durán SM, Guaderrama D, Hinchliff CE, Jørgensen PM, Kraft NJ et al. 2018. The Bien R package: a tool to access the botanical information and ecology network (BIEN) database. *Methods in Ecology and Evolution* **9**: 373–379.

Meinshausen M, Smith SJ, Calvin K, Daniel JS, Kainuma MLT, Lamarque J-F, Matsumoto K, Montzka SA, Raper SCB, Riahi K et al. 2011. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change* **109**: 213–241.

Moncrieff GR, Bond WJ, Higgins SI. 2016. Revising the biome concept for understanding and predicting global change impacts. *Journal of Biogeography* **43**: 863–873.

Moncrieff GR, Hickler T, Higgins SI. 2015. Intercontinental divergence in the climate envelope of major plant biomes. *Global Ecology and Biogeography* **24**: 324–334.

Moncrieff GR, Scheiter S, Bond WJ, Higgins SI. 2014. Increasing atmospheric CO_2 overrides the historical legacy of multiple stable biome states in Africa. *New Phytologist* **201**: 908–915.

Mucina L. 2019. Biome: evolution of a crucial ecological and biogeographical concept. *New Phytologist* **222**: 97–114.

Olson D, Dinerstein E, Wikramanayake E, Burgess N, Powell G, Underwood E, D'Amico J, Itoua I, Strand H, Morrison J et al. 2001. Terrestrial ecoregions of the world: a new map of life on Earth. *BioScience* **51**: 933–938.

Osborne CP, Salomaa A, Kluyver TA, Visser V, Kellogg EA, Morrone O, Vorontsova MS, Clayton WD, Simpson DA. 2014. A global database of C_4 photosynthesis in grasses. *New Phytologist* **204**: 441–446.

Penone C, Davidson AD, Shoemaker KT, DiMarco M, Rondinini C, Brooks TM, Young BE, Graham CH, Costa GC. 2014. Imputation of missing data in life-history trait datasets: which approach performs the best? *Methods in Ecology and Evolution* **5**: 961–970.

Pfadenhauer J, Klötzli F. 2014. *Vegetation der Erde.* Berlin, Germany: Springer.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231–259.

Prentice IC, Bondeau A, Cramer W, Harrison SP, Hickler T, Lucht W, Sitch S, Smith B, Sykes MT. 2007. Dynamic global vegetation modeling: quantifying

terrestrial ecosystem responses to large-scale environmental change. In: Canadell JG, Pataki DE, Pitelka LF, eds. *Terrestrial ecosystems in a changing world*. Berlin, Germany: Springer, 175–192.

Rammig A, Jupp T, Thonicke K, Tietjen B, Heinke J, Ostberg S, Lucht W, Cramer W, Cox P. 2010. Estimating the risk of amazonian forest dieback. *New Phytologist* **187**: 694–706.

Reich PB, Walters MB, Ellsworth DS. 1997. From tropics to tundra: global convergence in plant functioning. *Proceedings of the National Academy of Sciences, USA* **94**: 13730–13734.

Scheiter S, Langan L, Higgins SI. 2013. Next-generation dynamic global vegetation models: learning from community ecology. *New Phytologist* **198**: 957–969.

Schimper AFW. 1903. *Plant-geography upon a physiological basis*. Oxford, UK: Oxford University Press.

Schultz J. 2005. *The ecozones of the world*. Berlin, Germany: Springer.

Shangguan W, Dai Y, Duan Q, Liu B, Yuan H. 2014. A global soil data set for Earth system modeling. *Journal of Advances in Modeling Earth Systems* **6**: 249–263.

Sitch S, Huntingford C, Gedney N, Levy PE, Lomas M, Piao SL, Betts R, Ciais P, Cox P, Friedlingstein P *et al.* 2008. Evaluation of the terrestrial carbon cycle, future plant geography and climate- carbon cycle feedbacks using five dynamic global vegetation models (DGVMs). *Global Change Biology* **14**: 2015–2039.

Thornley JH. 1998. Modelling shoot : root relations: the only way forward? *Annals of Botany* **81**: 165–171.

Thuiller W, Guéguen M, Renaud J, Karger DN, Zimmermann NE. 2019. Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications* **10**: 1446.

Trabucco A, Zomer R. 2010. *Global soil water 611 balance geospatial database*. CGIAR Consortium for Spatial Information. [WWW document] URL https://cgiarcsi.community [accessed 18 January 2011].

van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F *et al.* 2011. The representative concentration pathways: an overview. *Climatic Change* **109**: 5.

Wakeling JL, Cramer MD, Bond WJ. 2012. The savanna–grassland 'treeline': why don't savanna trees occur in upland grasslands? *Journal of Ecology* **100**: 381–391.

White F. 1983. *The vegetation of Africa: a descriptive memoir to accompany the Unesco/AETFAT/UNSO vegetation map of Africa*. Paris, France: United Nations Educational, Scientific and Cultural Organization.

Whittaker RH. 1975. *Communities and ecosystems*. New York, NY, USA: Macmillan Publishing.

Williams JW, Jackson ST, Kutzbach JE. 2007. Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences, USA* **104**: 5738–5742.

Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, Damgaard CF, Dormann CF, Forchhammer MC, Grytnes J-A, Guisan A *et al.* 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* **88**: 15–30.

Zizka A, Silvestro D, Andermann T, Azevedo J, Duarte Ritter C, Edler D, Farooq H, Herdean A, Ariza M, Scharn R *et al.* 2019. Coordinatecleaner: standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* **10**: 744–751.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 False positive rates of the TTR-SDMs and Maxent models.

Fig. S2 Suitability surfaces of plant growth forms based on Maxent models.

Fig. S3 Density of species distribution records used for range modelling.

Fig. S4 Comparison of suitability scores based on TSS > 0.7 and > 0.6.

Fig. S5 Comparison of suitability scores based on TSS > 0.7 and > 0.8.

Fig. S6 Results of the growth form subsampling test.

Fig. S7 Effect of removing species with family-level imputation of growth form.

Fig. S8 Alternative TTR-SDM biome maps based on different numbers of clusters.

Fig. S9 Biomes of Africa based on Maxent range models.

Methods S1 Description of the TTR-SDM and model fitting.

Methods S2 Sensitivity analyses of the suitability surfaces.

Table S1 Reference list for BIEN occurrence records.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.