1.3 Navigating continent-wide geospatial data

1.3.1 Using imperfect global data for local decisions

The quality of biodiversity information varies substantially between different sources. Decisions that affect large geographical areas – like the whole of Africa – often rely on information from global sources that provide a consistent overview across the whole area. However, those decisions may be difficult to implement locally when they are inconsistent with specific local contexts. How we handle the potential mismatch between global datasets and local realities is critical for making effective evidence-based decisions.

Science-based policy assumes that management decisions are based on the best available scientific evidence. It makes sense then to assume that policy choices that affect all of Africa ought to be guided by scientific information also covering the whole continent. Unfortunately, there are significant trade-offs between the broad geographical coverage of data, and how accurately they represent any specific locality. National stakeholders will simply refuse to use global datasets that fail to represent reality based on their own experiences.

These feature maps supposedly all present the terrestrial ecosystems of Mozambique. Even without presenting the map legend (it is not feasible to show legends for all 162 vegetation types in the most detailed map, for instance), it is obvious that the three maps differ in material ways. The first map shows the extents of ecoregions[1], which derive from a global dataset based on broad landforms and vegetation types. Ecoregions are often used to assess ecological diversity at large spatial scales (see Topic 2.2.2) because they use a consistent classification scheme to cover the whole planet. By contrast, the second map is based on continent-wide mapping of ecosystems based on machine-learning classification of topographical and remotely sensed land cover information[2]. This is more detailed than the global ecoregion map, but lacks the detail of the third map, which shows 162 Mozambican ecosystem types based on field surveys, historical vegetation and soil maps, and expert assessments[3].

When it comes to policy decisions related to Mozambique, it is obvious that the detailed map is most useful. However, policy discussions considering larger areas – say, the whole of Southern Africa – cannot rely on these detailed data, which may be unavailable or inconsistent in neighbouring countries. This leaves scientists with the tough decision on whether to use a patchwork of inconsistent national datasets, or less accurate but consistent global data. The latter approach is often preferred for practical reasons (i.e. ease of reliable access, or the technical limitations of harmonising disparate datasets), but not without limitations.

A major limitation of global data is that they are not validated for every local context. For instance, global forest maps based on earth observation technologies are remarkably accurate at identifying trees. However, to a satellite sensor, trees in natural forest appear similar to trees in artificial timber plantations. This is not a failure of the dataset, which identified the trees correctly. But it can contribute to major mistakes when policy decisions are based on the incorrect assumption that all tree-covered landscapes are beneficial for biodiversity. Blindly calculating indicators based on these forest data (e.g. "*Forest area as a proportion of total land area*", a complimentary indicator of Goal A of the Global Biodiversity Framework) would misrepresent the situation on the ground.

Similarly, global datasets intended to identify ecological degradation are at risk of misattributing the effects of natural climate variation to degrading pressures. Many arid ecosystems throughout Africa transition between barren and vegetated landscapes at the onset of seasonal rainfall[4]. To satellite sensors, the reflectance of bare ground in pristine arid landscapes is indistinguishable from heavily degraded rangelands elsewhere. Again, in the absence of local context with long seasonal time-series, global datasets can grossly misrepresent the true state of biodiversity.

Natural climate variation can change ecosystems in profound ways that are not immediately obvious from remote sensing.

These repeat photos from the Nama Karoo, South Africa, show the dramatic ecological differences at the same site during wet (2011, top) and dry (2016, bottom) years. These grassy and bare states have different reflectance according to satellite-based sensors, so snapshot images risk misattributing natural seasonal cycles to ecological degradation when sufficiently long time-series data are unavailable.

Source: von Maltitz, G.P. et al. (2024) *Coupled earth system and human processes: an introduction to SPACES and the book*. In von Maltitz, G.P et al. (eds): Sustainability of Southern African ecosystems under Global Change. Springer. Cham, Switzerland. Image: Graham von Maltitz et al. under Creative Commons License CC BY 4.0 DEED.

While important, the accuracy of global dataset is not the only thing that matters. Globally derived information is only useful when it aligns with local contexts and needs. Thus, the characteristics of end-users are as important as the data themselves. During the International Congress for Conservation Biology in Kigali, Rwanda in July 2023, a group of researchers from more than 20 international organisations identified seven preconditions of useful biodiversity information[5]. These preconditions are equally important and indivisible, so failing to consider all of them limits the uptake of information for evidence-based biodiversity reporting, policy and action. These preconditions are useful for diagnosing why global datasets do not lead to local actions and identifying interventions to improve the uptake of evidence.

	Preconditions	Supporting interventions
PRECONDITIONS – User-focused	**Awareness:** Users should know the data exist	**Outreach:** Improve marketing, advertising, publicising, and communication
	Capacity: Users should have the technical ability to use the data	**Training:** Produce tutorials, user manuals, demos, or present training workshops
	Infrastructure: Users should have sufficient internet access and computing power	**Data optimisation and investment:** Package data more efficiently (for example, for mobile access) Enhance or share information and communications technology infrastructure"
	Usability: Information should be in an easy-to-use and accessible format and language	**Design and translation:** Design data platforms to be more user-friendly and intuitive Produce jargon-free explanations in multiple languages"
PRECONDITIONS – Data-focused	**Legitimacy:** Data should be considered authoritative	**Advocacy and diplomacy:** Advocate and negotiate such that data are accepted by national and international authorities
	Saliency: Data should be fit for purpose	**Engagement and exchange:** Interact with end-users to better understand their needs
	Credibility: Data should be scientifically robust and representative	**Validation:** Ensure scientific validity of data through independent verification and peer review.

Seven preconditions of useful biodiversity information.

The usefulness of biodiversity information depends on the interplay between demand-driven preconditions determined by users (shades of red) and supply-side preconditions determined by the data (shades of blue). A variety of interventions can improve the uptake of biodiversity information in national policies and programmes.

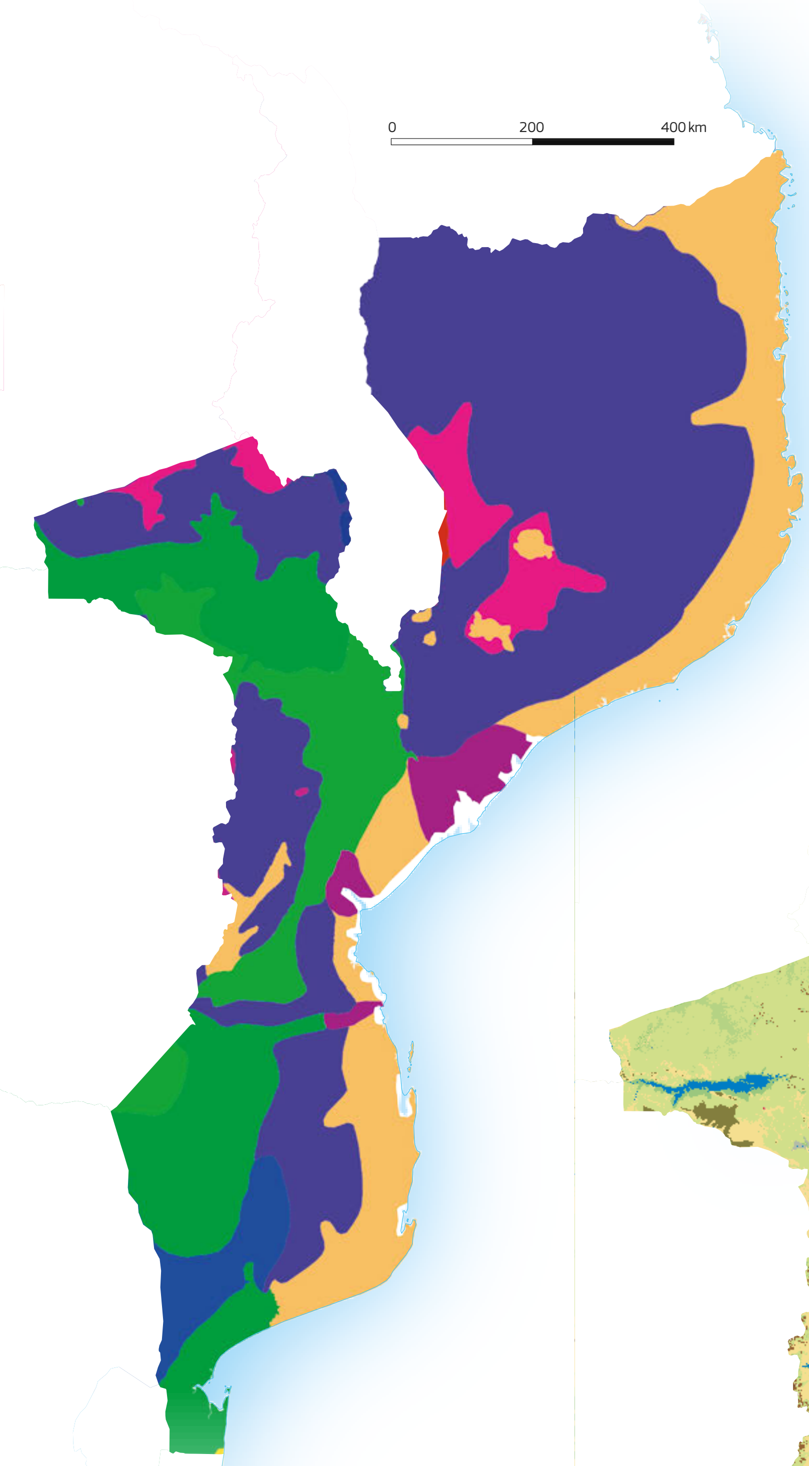
Source: Buschke, F.T. et al. (2023) Make global biodiversity information useful to national decision-makers. *Nature Ecology and Evolution*, 7, 1953-1956.

Ultimately, global datasets are essential for continent-wide scientific assessments, such as those presented throughout this Atlas. While trade-offs are inevitable, the known limitations of global information must be recognised and considered whenever translating information into policy and action. The next four topics in this chapter describe how spatial scale, geographical and taxonomic biases, as well as methodological changes affect our ability to map continent-scale biodiversity information accurately.

Timber plantations mapped as forest in global datasets.

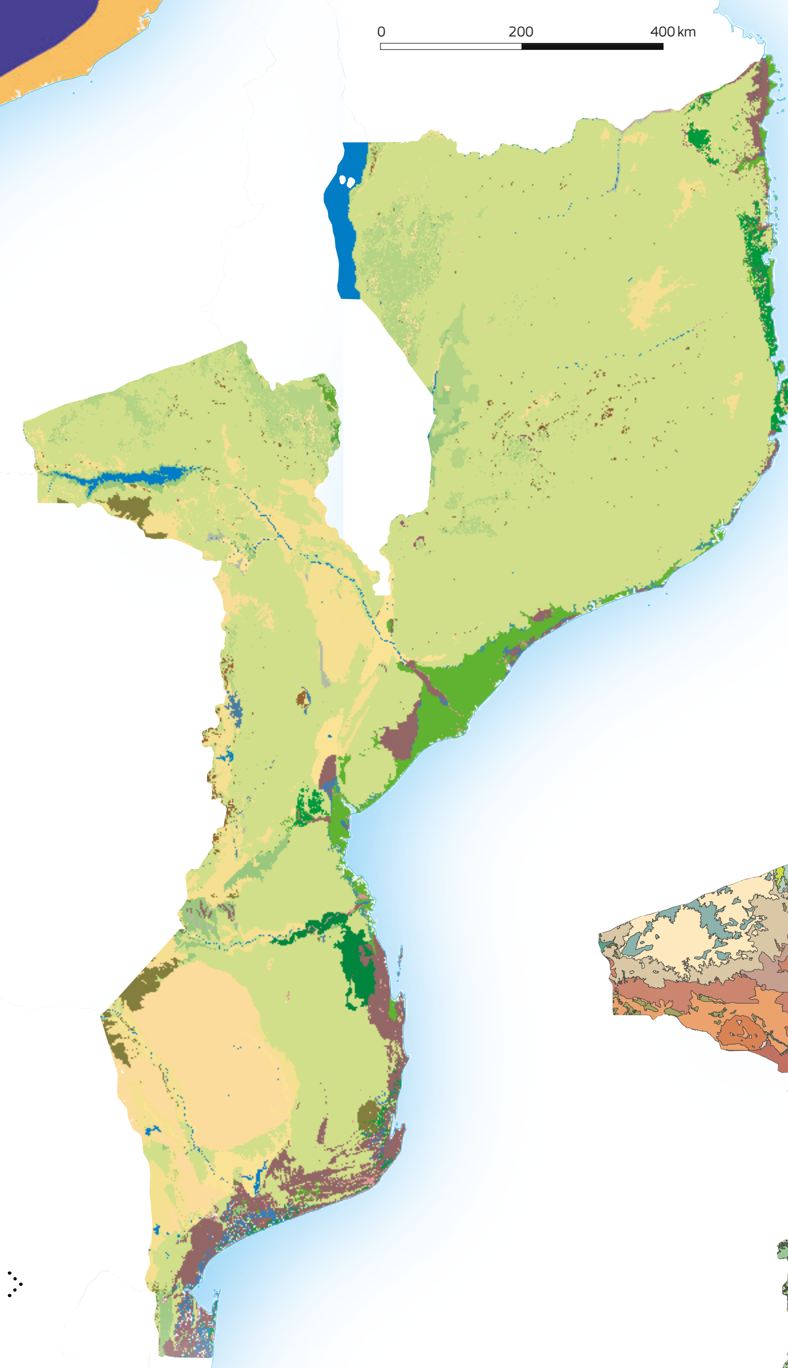
Global datasets based on remote sensing identify habitat features without necessarily considering the specific conservation context. In South Africa, for example, global forest maps identify tree-covered landscapes with great accuracy without distinguishing that these trees are within artificial timber plantations with little value for conservation.

Source: Bourgoin, C., et al. (2023): Global map of forest cover 2020 - version 1. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/10d1b337-b7d1-4938-a048-686c8185b290



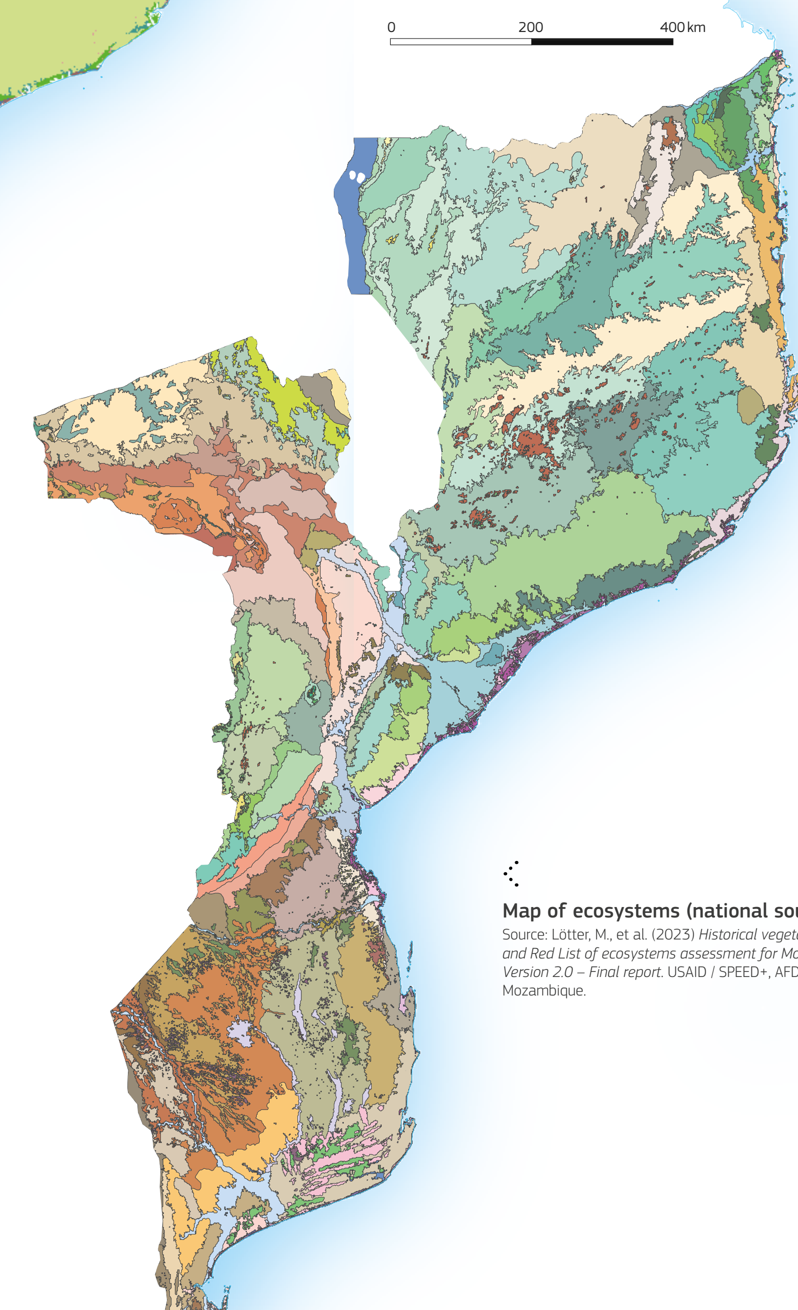
Map of ecoregions (global source).

Source: Olson, D.M., et al. (2001). Terrestrial ecoregions of the world: a new map of life on earth. *BioScience*, 51, 933–938.



Map of ecosystems (continental source).

Source: Sayre, R. (2023) Africa Terrestrial Ecosystems: U.S. Geological Survey data release, https://doi.org/10.5066/ P9BHBKA2



Map of ecosystems (national source).

Source: Lötter, M., et al. (2023) *Historical vegetation map and Red List of ecosystems assessment for Mozambique – Version 2.0 – Final report*. USAID / SPEED+, AFD/FFEM. Maputo, Mozambique.

The ecosystems of Mozambique.

Different data sources intended to depict the same biological features can differ significantly. The global map of ecoregions shown here was derived from landforms and vegetation cover. The continental map of ecosystems relied on a machine learning algorithm, GIS, and remote sensing. The national ecosystem map combined field data, historical vegetation and soil maps, and expert assessments. National data is clearly more detailed, but too often decisions are based on global datasets that tend to be more easily accessible. (Note: the maps do not include legends due to space constraints).

Sources: In individual captions.

References

[1] Olson, D.M., et al. (2001). Terrestrial ecoregions of the world: a new map of life on earth. *BioScience*, 51, 933–938.

[2] Sayre, R. (2023) *Africa Terrestrial Ecosystems*: U.S. Geological Survey data release, https://doi.org/10.5066/P9BHBKA2.

[3] Lötter, M., et al. (2023) *Historical vegetation map and Red List of ecosystems assessment for Mozambique – Version 2.0 – Final report*. USAID / SPEED+, AFD/FFEM. Maputo, Mozambique.

[4] von Maltitz, G.P. et al. (2024) *Coupled earth system and human processes: an introduction to SPACES and the book*. In von Maltitz, G.P et al. (eds): Sustainability of Southern African ecosystems under Global Change. Springer. Cham, Switzerland.

[5] Buschke, F.T. et al. (2023) Make global biodiversity information useful to national decision-makers. *Nature Ecology and Evolution*, 7, 1953-1956.

1.3.2 Source of uncertainty: the spatial scale of biodiversity assessment

Most protected areas have lists of common species. Yet anyone who has visited one knows the frustration of not sighting rare species from these lists. This is not because the lists are necessarily wrong, but because monitoring biodiversity is fundamentally scale-dependent. Just because a species occurs inside a protected area, does not mean that it also occurs in the area immediately surrounding the game-viewing vehicle. Scale-dependency also affects how species are monitored for the whole of Africa. Something as foundational as mapping a single species' geographical distribution is far from straightforward.

When researchers record a species during field surveys, they mark coordinates of their sightings and upload these to international databases, such as the Global Biodiversity Information Facility (GBIF)[1]. Alongside field records from researchers, GBIF also includes historical records from museum specimens and modern sightings recorded by the public as part of citizen science programmes. While occurrence records are the most reliable information about where species do occur (i.e. their true presence), they are very likely to underestimate occurrences in areas that are rarely visited and poorly studied (i.e. their false absence).

A well-studied species like the Chacma baboon (*Papio ursinus*) has many occurrence records throughout southern Africa. These unmistakable and charismatic primates occur in many different habitats around cities, farmlands, mountains, and deserts. But other less obvious species might not be recorded with as much regularity or accuracy.

The simplest way to extrapolate a species geographic range into understudied regions is to draw a shape around the outermost occurrence records. These types of ranges are known as Extents of Occurrence, and are commonly used in field guides to show where species are likely to occur. In the process of developing the IUCN Red List, experts delineated and refined the Extents of Occurrence for practically all vertebrate species worldwide[2–4]. In contrast to point records, Extents of Occurrence are susceptible to overestimating species ranges because they include large areas where species might not actually occur (i.e. false presences). For well-studied baboons, experts knew to cut out a portion of their range in the Kalahari Desert, where the species was known to be absent. For lesser known species, however, such exclusions may be overlooked.

Several different modelling techniques aim to fill the gap between fine-scale occurrence records and coarse Extents of Occurrence maps. These techniques rely on statistical models to find relationships between known occurrence records and a suite of predictors, like climate, elevation, or land cover. A widely used method is Maxent[5], a machine-learning approach that uses entropy maximisation to predict species distributions. Methods like these are very powerful, but they have a steep learning curve to understand all the possible analytical choices and validate model outputs. The species distribution model for the baboon is sophisticated enough to predict that this adaptable species can occur in habitats as vastly different as dry Namibian deserts or cold mountains in Lesotho.

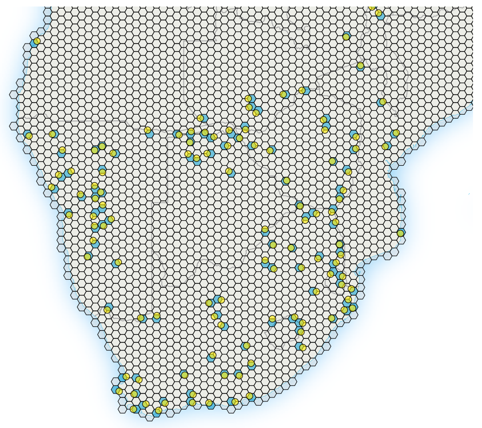
Mapping the range of a single species is tricky enough, but when researchers need to aggregate the distributions of multiple species at the same time, they rely on sampling grids. Range maps of different species are stacked so the number of species occurring together within the sampling grid cell give an indication of the biodiversity of that area.

The choice of grid size can have a huge effect on biodiversity estimates. For example, counting the number of hexagonal grid cells that contain occurrence records for the baboon will differ a lot depending on the size of the grid. When grid cells are only 50 km across, most cells will contain zero occurrences while a few cells might have one or two records. However, when grid cells are 250 km across, a single cell can easily encompass multiple records. Summing the area of the grid cells containing occurrence records quantifies the geographic range as the 'Area of Occupancy'.

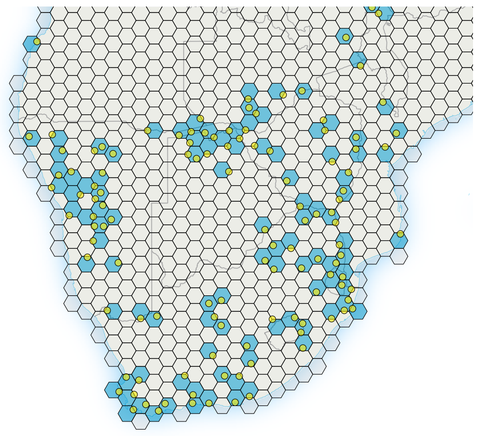
For the baboon, we can see that the estimates of its geographic range size as the Area of Occupancy varies considerably for each combination of range map and scale of assessment. Range size estimates based on point records in 50 km grid cells are 20 times smaller than ranges estimated from the Extent of Occurrence in a 250 km grid. Yet both approaches can be justified under the right circumstances.

Ultimately, the onus in on users to understand and interpret whether the source of distribution data and the scale of assessment are appropriate for their specific purposes. This holds true for single protected areas, countries, regions, and the whole continent of Africa.

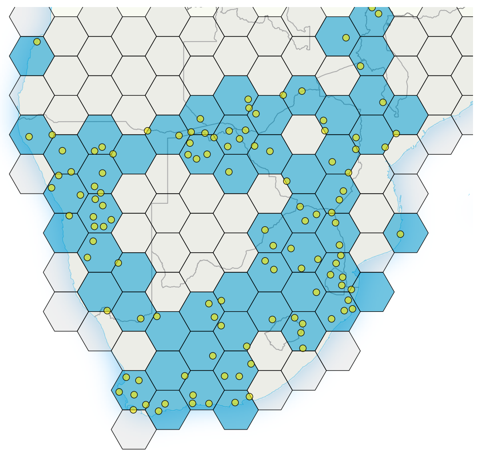
50 km hexagonal grid:



100 km hexagonal grid:



250 km hexagonal grid:



The Area of Occupancy at three different spatial scales.

The geographic range for the baboon (*Papio ursinus*) expressed as the Area of Occupancy, by counting grid cells (50 km, 100 km, 250 km across) that contain occurrence records.

Source: GBIF.org (2023), Global Biodiversity Information Facility (GBIF) Homepage. Available from: https://www.gbif.org [17 November 2023].

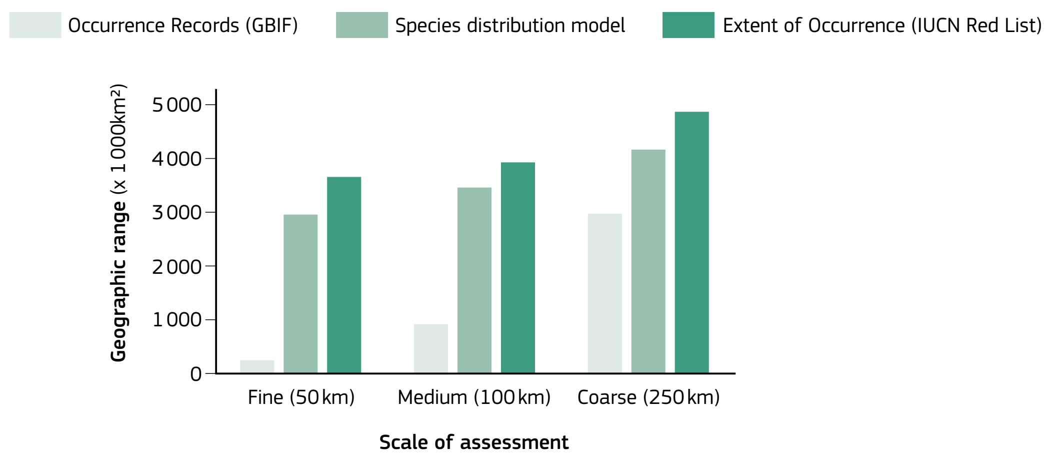
Mapping species' distribution ranges for the whole of Africa is not a simple task. Results vary significantly depending on how, and at which spatial scale, species' ranges are defined. Biodiversity maps are simplified depictions of complex realities, shaped by scientific views. To best interpret them, it is key for the users to understand the trade-offs between different mapping technics depending on e.g. symbolism. It is as important for the scientists to report mapping choices and the limits of their data in order to best support decision making.

Chacma baboon.

The Chacma baboon is very adaptable to life in many different habitats, from the deserts of Namibia, to the mountains of Lesotho. This makes it challenging to map its distribution range accurately.

Source: Bernard DUPONT on Wikimedia Commons CC BY-SA 2.0.

Types of distribution

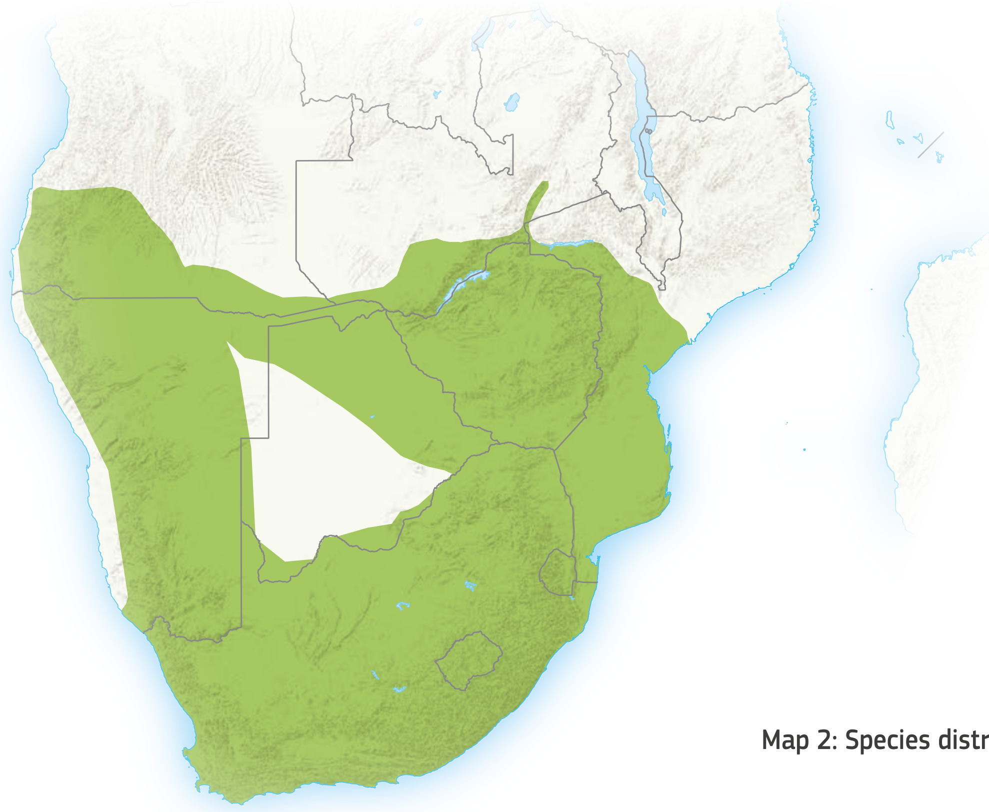


Different, yet equally valid, estimates of the baboon's Area of Occupancy.

The geographic range for the Chacma baboon (*Papio ursinus*) expressed as the Area of Occupancy, for each combination of distribution type (Extent of Occurrence, species distribution model, and occurrence records) and scale of assessment (a grid with cells measuring 50 km, 100 km, 250 km across).

Source: Own calculations.

Map 1: Extent of Occurrence (Expert delineations for the IUCN Red List)



Three technically valid depictions of the baboon's geographical distribution.

The geographic range for the baboon (*Papio ursinus*) displayed as its Extent of Occurrence (from the IUCN Red List), a species distribution model (climate-based suitability modelled using Maxent), and occurrence records (based on points from the Global Biodiversity Information Facility, GBIF).

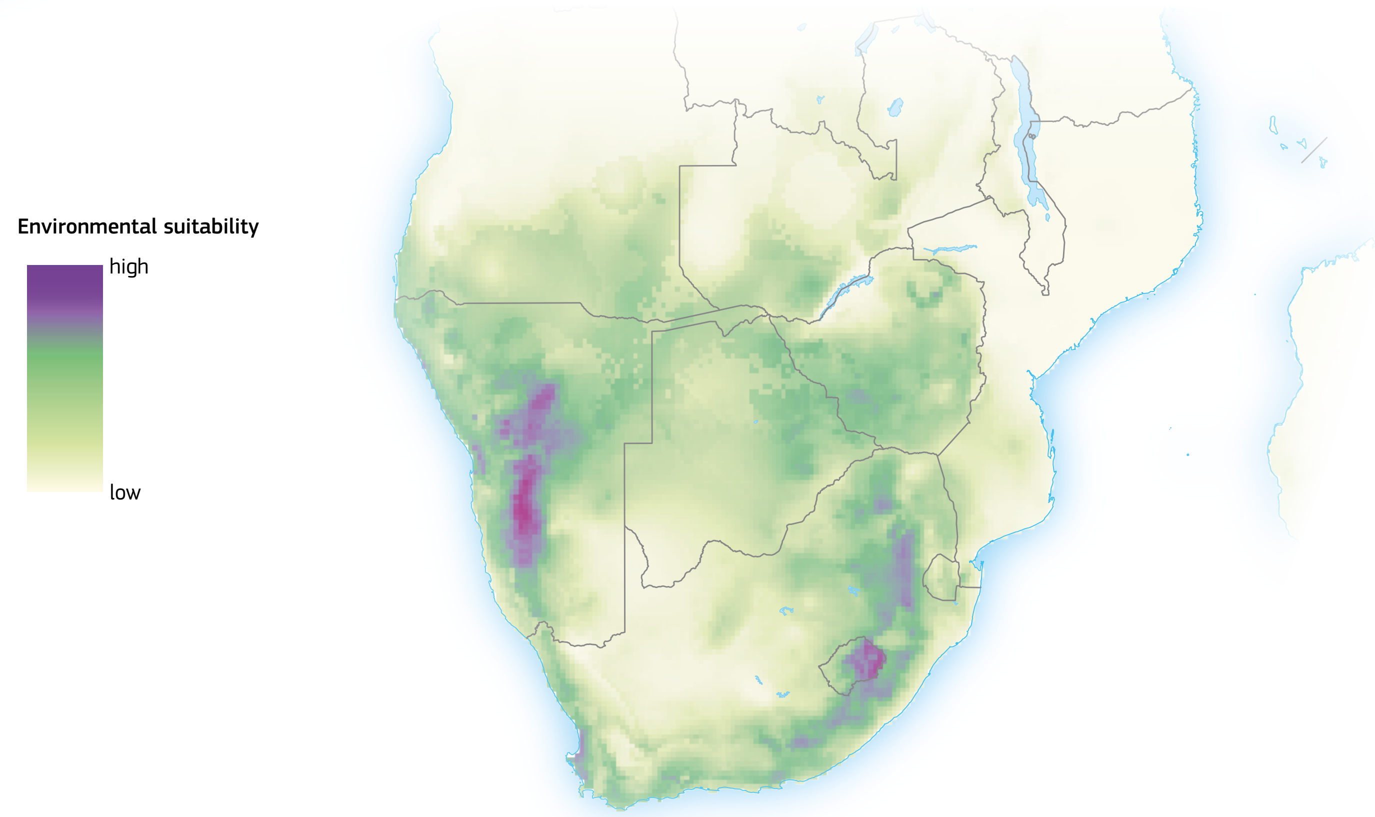
Sources:

Map 1: Sithaldeen, R. (2019) Papio ursinus (errata version published in 2020). The IUCN Red List of Threatened Species 2019: e.T16022A168568698. https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T16022A168568698.en

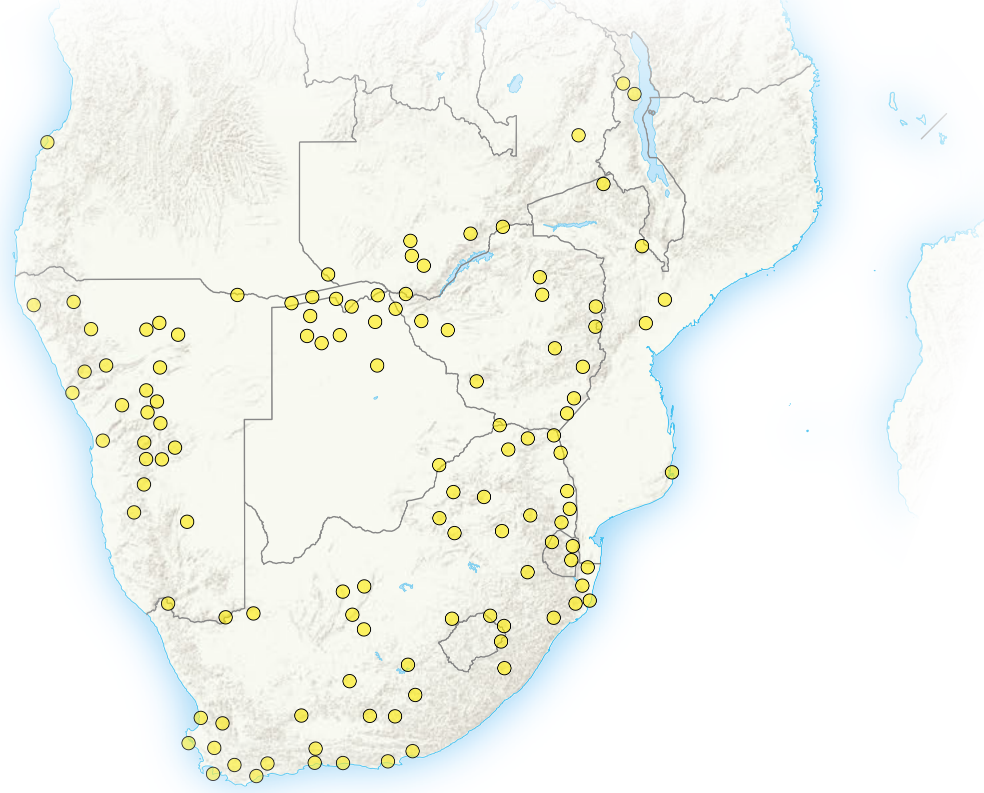
Map 2: Own modelling.

Map 3: GBIF.org (2023) Global Biodiversity Information Facility (GBIF) Homepage. Available from: https://www.gbif.org [17 November 2023].

Map 2: Species distribution model (Climatic suitability based on Maxent modelling)



Map 3: Occurrence Records (GBIF)



References

[1] Edwards, J. L. (2004) Research and societal benefits of the Global Biodiversity Information Facility. BioScience 54, 486.

[2] Orme, C. D. L. et al. (2005) Global hotspots of species richness are not congruent with endemism or threat. Nature 436, 1016–1019.

[3] Grenyer, R. et al. (2006) Global distribution and conservation of rare and threatened vertebrates. Nature 444, 93–96.

[4] Cox, N. et al. (2022) A global reptile assessment highlights shared conservation needs of tetrapods. Nature 605, 285–290.

[5] Phillips, S. J., et al. (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190, 231–259.

1.3.3 Geographic biases in biodiversity data

Although scientists currently know more about the distributions of plants and animals than at any other point in human history, large parts of Africa are essentially still unexplored. Reasons for this vary. Some countries have too few ecologists to cover large territories. In other countries, untraversable terrain, conflict zones, or poor road access prevent researchers from conducting biodiversity surveys. The outcome of these geographical biases is that it is sometimes hard to know whether an area has few species living there, or whether it is simply understudied.

This feature map shows all the occurrence records in the Global Biodiversity Information Facility's (GBIF) database[1]. Only a few countries – notably Benin, South Africa, and Kenya – are relative well covered by survey records. Most of the Sahel, Democratic Republic of the Congo, and Angola are data poor, despite these regions having some of the most unique biodiversity on the planet. Africa's biodiversity hotspots are amongst the most poorly studied worldwide, with only a tiny fraction of their total biodiversity represented in global databases[2].

Benin is unique in that its whole territory is covered by systematic surveys, such as the *Census of National Forest of Benin*[3]. But within most other countries, data are not spread evenly. In countries like Namibia, Botswana, and Ethiopia, occurrence records are clearly concentrated around urban centres, protected areas, and road networks. In such instances, the availability of data says more about the accessibility to researchers than the distributions of species.

How the Kunming-Montreal Global Biodiversity Framework considers data biases.

Target 21 of the Global Biodiversity Framework aims to:

"Ensure that the best available data, information and knowledge are accessible to decision-makers, practitioners and the public to guide effective and equitable governance, integrated and participatory management of biodiversity, and to strengthen communication, awareness-raising, education, monitoring, research and knowledge management..."

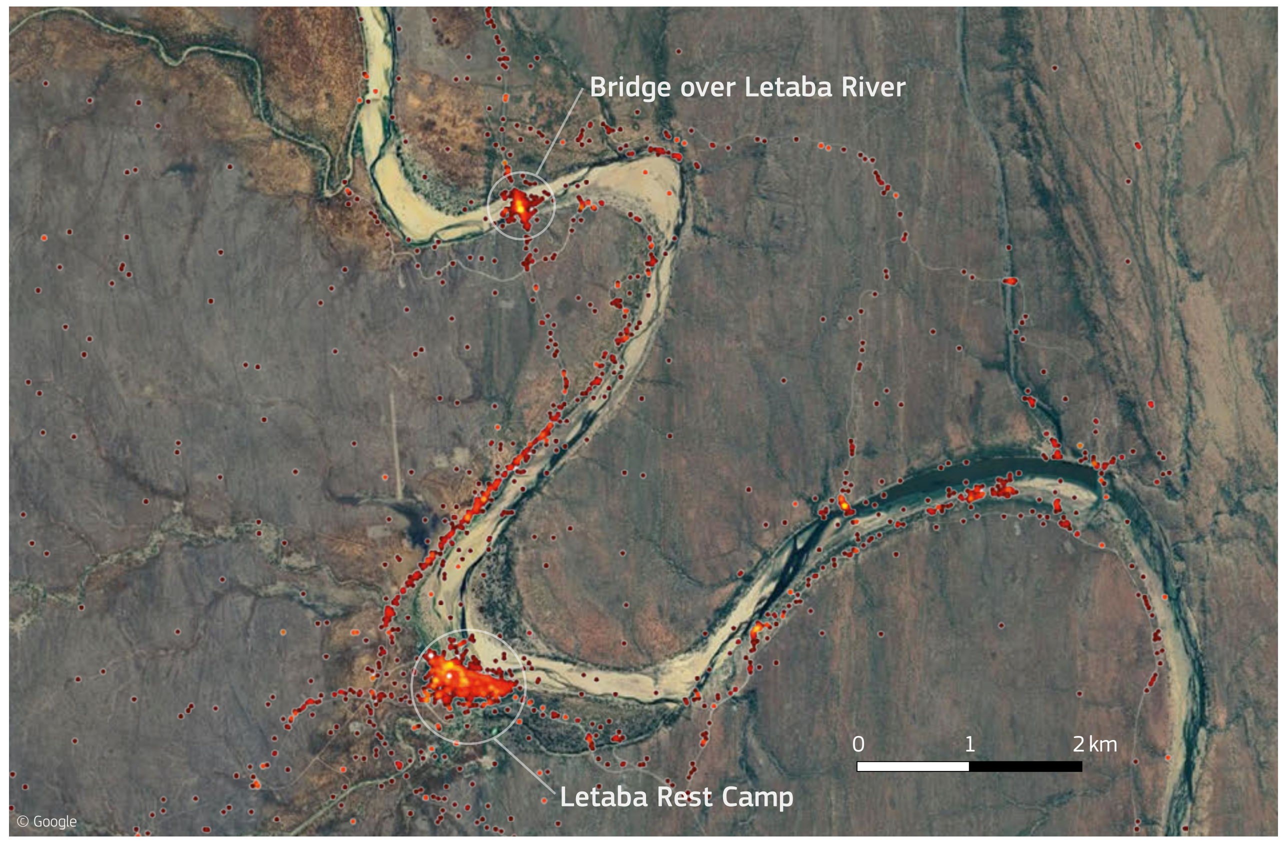
The indicator to monitor progress towards this targets is still under development, but it is proposed to be a composite indicator that includes the geographic coverage of accessible biodiversity data for all species[5]. In simple terms, success requires that geographically representative data is available across whole species' ranges.

There are also geographical data biases within individual protected areas. This inset map shows how occurrence records in a small part of Kruger National Park, South Africa, are concentrated around campsites, roads, and lookout points. Like in many protected areas on the continent, researchers and tourists to Kruger National Park may not leave their vehicles unless accompanied by armed rangers. This means that most sightings only occur in very specific areas; not because species prefer these areas, but because they are more accessible to scientists. In many protected areas, wildlife census counts are carried out from vehicles or fixed survey points and researchers must use sophisticated statistical techniques (e.g. distance-based sampling) to extrapolate their counts to areas not covered by their surveys.

Geographically representative data is essential for effective conservation and is prioritised by the Kunming-Montreal Global Biodiversity Framework (see Box). However, recent estimates suggest that at current rates of progress, it might take another two centuries before scientists have covered the whole continent at least once[4]. Unfortunately, just one visit might not be enough because it is unlikely that all species will be recorded during a single survey. Latest estimates suggest that more than 10 surveys will be needed to record half of all know species in an area[4].

Complete geographically representative biodiversity data will not be available for the foreseeable future. Therefore, the onus is on scientists, managers, and policymakers to acknowledge and accommodate imperfect information on species' distributions. In practice, this means confirming whether a high density of data in an area is due to there being high biodiversity there or merely because the area is easily accessible to researchers. It also means that the absence of data should be treated cautiously because, more often than not, it tells us more about important gaps in our knowledge than it does about the unsuitability of an area to plants and animals.

Scientists currently know more about the distributions of plants and animals than ever before, but information still tends to come from only a small subset of the whole planet. Data are predominantly from a few well-studied countries, while large parts of Africa are still unexplored. When analysing spatial biodiversity data, it is important to distinguish areas that truly have low biodiversity from areas that are simply understudied.

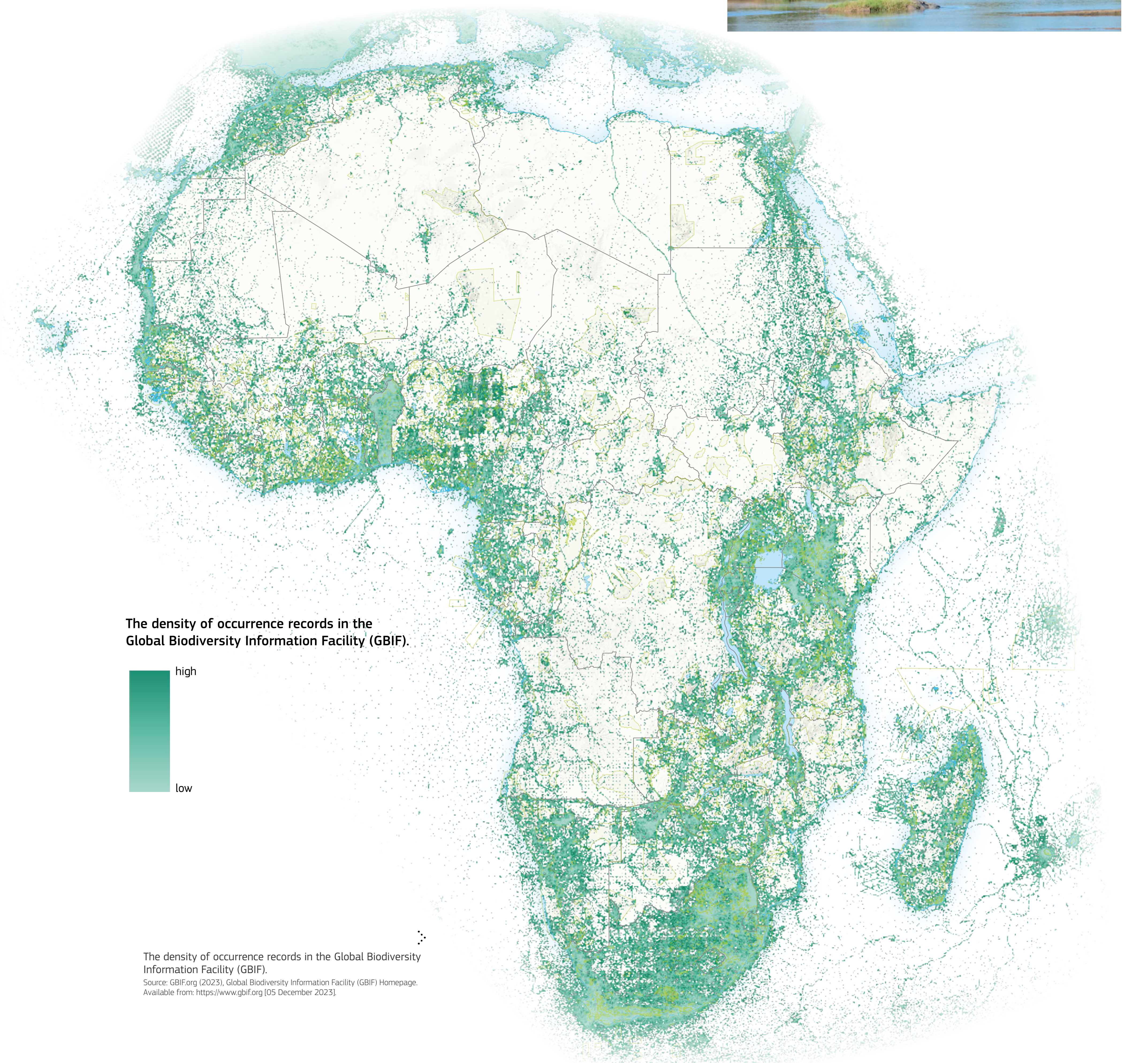


Letaba Rest Camp, Kruger National Park, South Africa.
The density of occurrence records from the Global Biodiversity Information Facility (GBIF) around Letaba Rest Camp in Kruger National Park, South Africa.
Source: GBIF.org (2023) Global Biodiversity Information Facility (GBIF) Homepage. Available from: https://www.gbif.org [05 December 2023].

Letaba Rest Camp, Kruger National Park, South Africa.
The Letaba Rest Camp in Kruger National Park has a viewing deck overlooking the Letaba River. Tourists can sip sundowners while watching animals on the wide-open sandbank. The ease of viewing means that there is a disproportionately high number of occurrence records at the campsite.
Source: Falko Buschke, with permission, all rights reserved.

Spotting species along the road network.
This lioness preferred resting on the warm road, rather than the dew-covered vegetation. Occurrence records are often concentrated along roadways because researchers and tourists are often not allowed to leave their vehicles and animals are much more visible on open routes compared to the dense roadside vegetation.
Source: Falko Buschke, with permission, all rights reserved.

Sightings at dedicated viewpoints.
This herd of elephants could be spotted easily from the bridge over the Letaba River, Kruger National Park, South Africa. Dedicated lookout points – like bridges, bird hides, or piers – tend to have a high concentration of occurrence records.
Source: Falko Buschke, with permission, all rights reserved.



The density of occurrence records in the Global Biodiversity Information Facility (GBIF).
Source: GBIF.org (2023), Global Biodiversity Information Facility (GBIF) Homepage. Available from: https://www.gbif.org [05 December 2023].

References

[1] Edwards, J. L. (2004) Research and societal benefits of the Global Biodiversity Information Facility. *BioScience*, 54, 485–486.

[2] Yesson, C. et al. (2007) How global is the Global Biodiversity Information Facility? *PLoS ONE*, 2, e1124.

[3] Akouèhou, S. G. (2010) *The Census of National Forests of Benin*. Ministry of the Environment, Benin.

[4] Farooq, H., et al. (2021) Mapping Africa's biodiversity: more of the same is just not good enough. *Systematic Biology*, 70, 623–633.

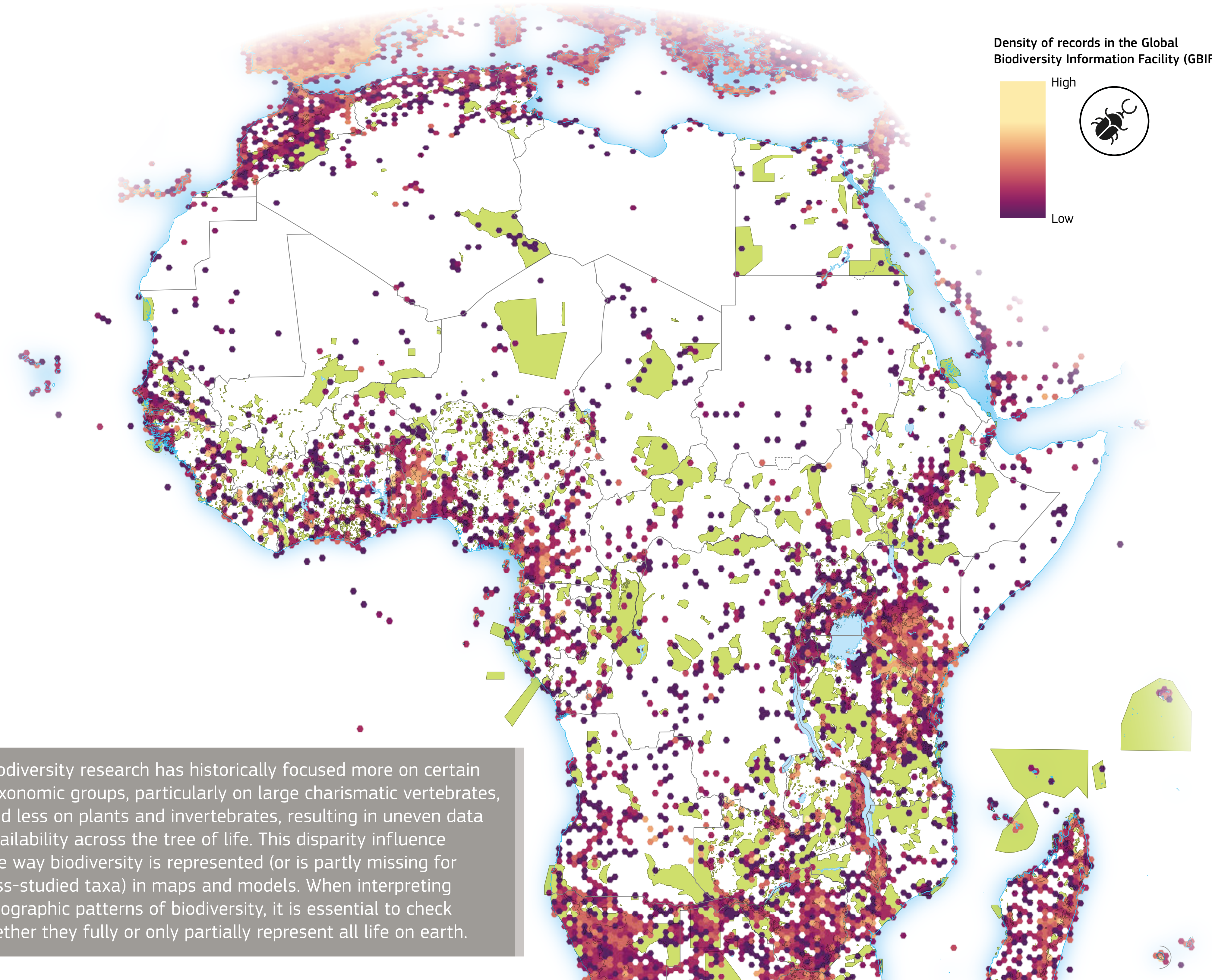
[5] Oliver, R. Y., et al. (2021) Global and national trends, gaps, and opportunities in documenting and monitoring species distributions. *PLoS Biology*, 19, e3001336.

1.3.4 Taxonomic bias issues

Biodiversity refers to the diversity of life on Earth at multiple levels, including genetic, species, and ecosystem diversity, as well as the ecological functions, interactions, and evolutionary processes that sustain them. But most of scientific evidence is from just a few well-studied taxonomic groups, with thousands of species yet to be discovered. For decades, charismatic vertebrate species received much more research focus than invertebrates and plants, even though the latter groups make up a much larger proportion of known biodiversity.

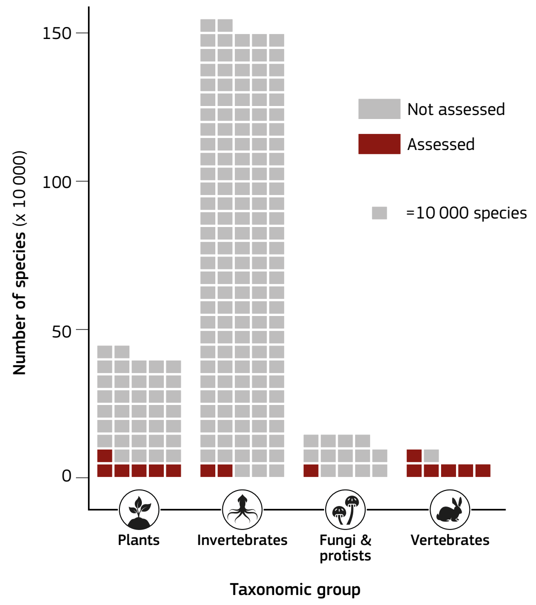
The map on the left shows the densities of nearly 650 000 known records from 1 346 beetle species in Africa represented in the Global Biodiversity Information Facility[1]. By contrast, the map on the right shown the densities of more than 48 million known records from 1 439 bird species from the same database. Even though there are an estimated 90 – 180 beetle species for every one bird species globally[2], there are 70 times fewer data records for this group of organisms.

This type of taxonomic bias is not unique to the Global Biodiversity Information Facility. For instance, the International Union for Conservation of Nature's (IUCN) Red List shows similar biases. Currently, the extinction threats of roughly the same amount of plant (~63 000) and vertebrate (~60 000) species have already assessed by the IUCN. However, there are about six plant species for every one vertebrate species, which equates to only 15 % of all known plants with extinction threat assessments compared to 81 % of vertebrates. The numbers look even worse for invertebrates (2 %) and fungi & protists (0.5 %), which are practically invisible in global threat assessments.



The density of records for beetles in the Global Biodiversity Information Facility (GBIF).
Source: GBIF.org (2024), Global Biodiversity Information Facility (GBIF) Homepage. Available from: https://www.gbif.org [18 January 2024].

Biodiversity research has historically focused more on certain taxonomic groups, particularly on large charismatic vertebrates, and less on plants and invertebrates, resulting in uneven data availability across the tree of life. This disparity influence the way biodiversity is represented (or is partly missing for less-studied taxa) in maps and models. When interpreting geographic patterns of biodiversity, it is essential to check wether they fully or only partially represent all life on earth.

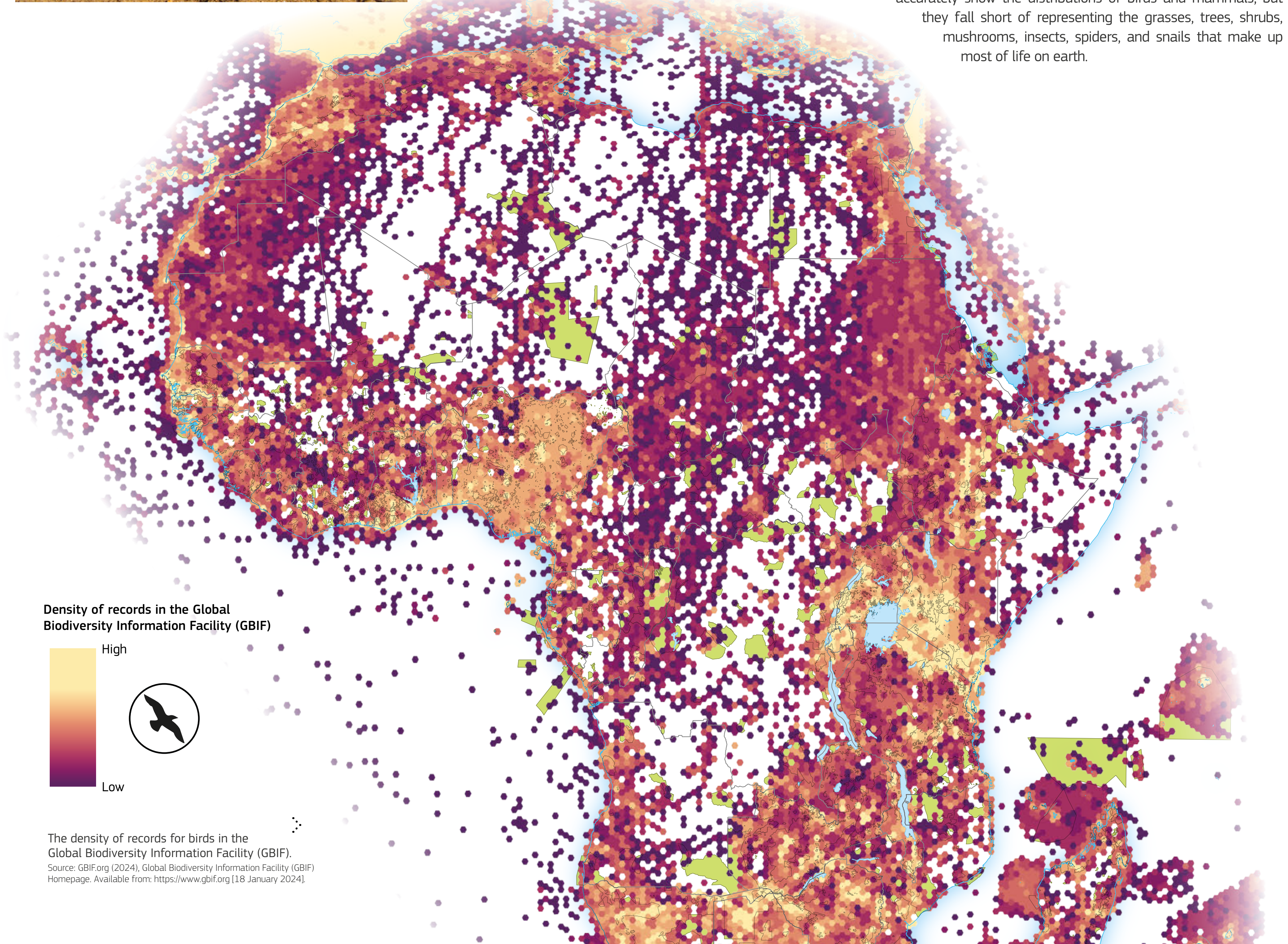


Taxonomic variation in the species evaluated by the IUCN Red List.
Each square represents 10 000 species. Red squares have had their extinction threat assessed by the IUCN, while grey squares have yet to be evaluated. Despite having the fewest species overall, vertebrates make up the bulk of species assessed by the IUCN Red List.
Source: Own calculations based on data from IUCN Red List version 2022-2 Summary Statistics (https://www.iucnredlist.org/resources/summary-statistics).

African dung beetles, one of more than a million known beetle species on the planet.
Dung beetles (*Scarabaeus* spp.) perform the essential role of processing animal waste and ensuring nutrient cycling and ecosystem functioning. Despite this thankless task, these beetles and others like them are mostly underrepresented in global biodiversity databases.
Source: Jochen Smolka on flickr CC BY-NC-SA 2.0.

Taxonomic biases can have major implications for policy. The use of wild plants and fungi for food, fibre, fuel, and medicine has shaped African societies over centuries. Similarly, some invertebrates have supported agriculture though pollination and pest control, while others have destroyed crops (e.g. locust swarms) and spread diseases (e.g. malaria vectors). Despite the social and cultural significance of these organisms, they tend to be overlooked in global biodiversity databases.

In practice, scientific knowledge covers only a small fraction of all species. Most biodiversity is still unknown. As a result, interpretations of maps that depict biodiversity need to take into account the limitations of the underlying data. In most cases, these maps might accurately show the distributions of birds and mammals, but they fall short of representing the grasses, trees, shrubs, mushrooms, insects, spiders, and snails that make up most of life on earth.



The density of records for birds in the Global Biodiversity Information Facility (GBIF).
Source: GBIF.org (2024), Global Biodiversity Information Facility (GBIF) Homepage. Available from: https://www.gbif.org [18 January 2024].

The common ostrich, an endemic African bird.
Birds, like this ostrich (*Struthio camelus*), are well represented in global biodiversity databases despite only making up a relatively small proportion of all known species.
Source: Axel Tschentscher on Wikimedia Commons CC BY-SA 4.0 DEED.

References

[1] GBIF.org (2024), Global Biodiversity Information Facility (GBIF) Homepage. Available from: https://www.gbif.org [18 January 2024].
[2] Stork, et al. (2015) New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proceeding of the National Academy of Sciences*, USA, 112, 7519-7523.

1.3.5 Methodological updates to underlying data

Measuring biodiversity for the whole continent is challenging, so scientists often use sophisticated models to predict biodiversity. These models rely on algorithms that process incomplete information from field surveys or satellite sensors, resulting in consistent and comparable predictions across large areas. However, improving and updating these algorithms can change their predictions dramatically, even when the actual biodiversity features stay the same. It is important to understand when the dynamics of a dataset are caused by real changes to biodiversity, and when they are simply the outcome of changes to the underlying model.

The Biodiversity Intactness Index is a perfect example of how biodiversity predictions can change due to improvements to the modelling approach. The Biodiversity Intactness Index is a measure of an area's species richness or average population abundance, presented as the proportion of what would be expected in an equivalent intact ecosystem[1]. So, an Index value of 0.5 means that the current biodiversity is half of what it was before any human interference. This is useful because it means that biodiversity can theoretically be compared across vastly different ecosystem types.

Even though the idea of the Biodiversity Intactness Index dates back to 2005, models for the whole globe only appeared a decade later[2,3]. These global estimates were based on a generalised linear mixed effect regression model, a statistical technique that finds patterns in comparable biodiversity surveys across multiple sites that differ in the nature or intensity of human impacts[4]. It then uses these patterns to predict biodiversity in unsurveyed sites based on land-cover. As with all statistical techniques, the predictions are only as good as the quality of the data used to train the model.

Edges of the Sahara desert.

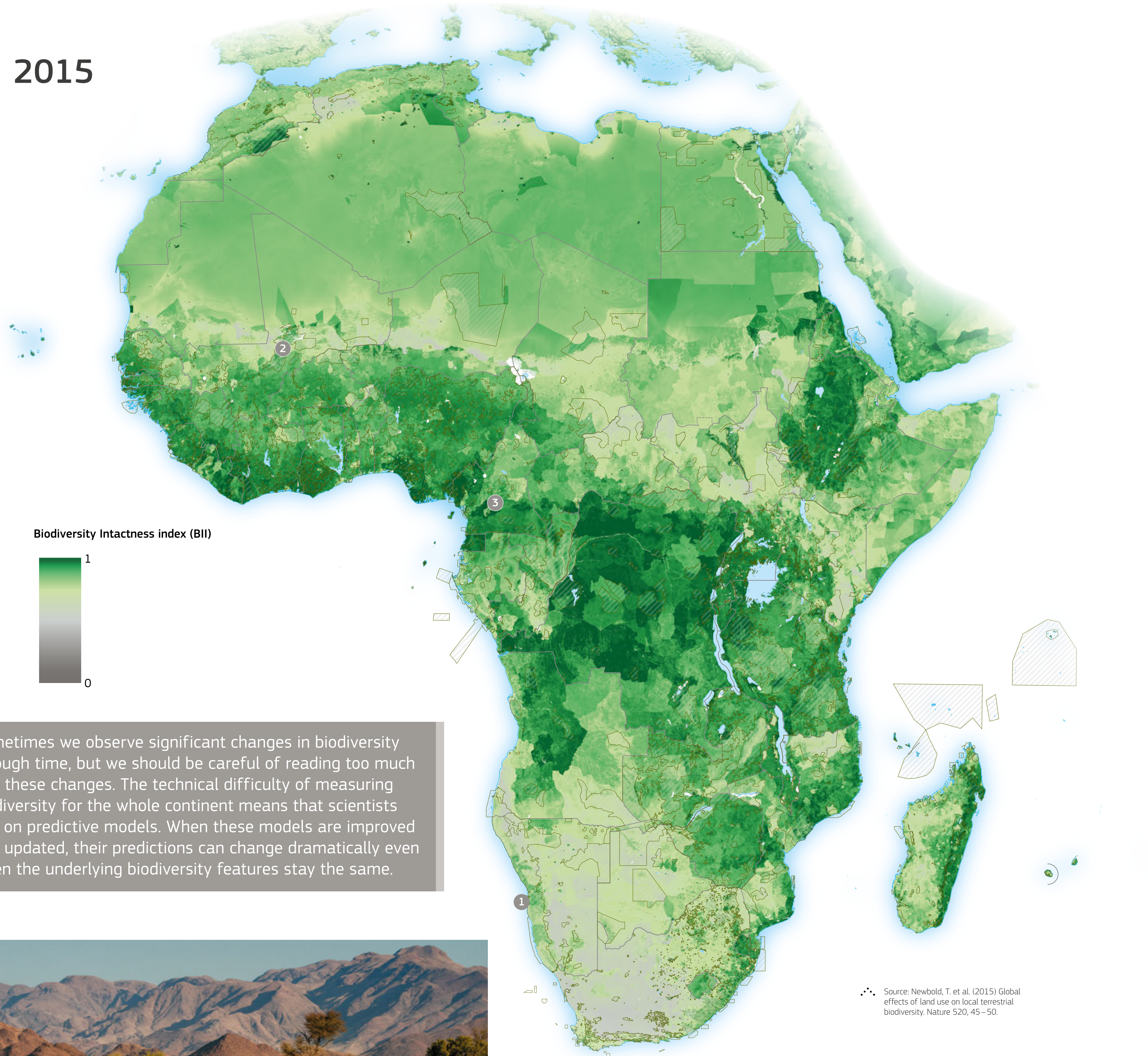
The updated Biodiversity Intactness Index now shows a clearer picture of areas of known ecological degradation, like the Sahel in Mali. The early model overestimated intactness in rangelands with intermediate human population pressure, but this has been fixed in the updated version.

Source: TREEAID on flickr CC BY 2.0.

Deforestation in Cameroon.

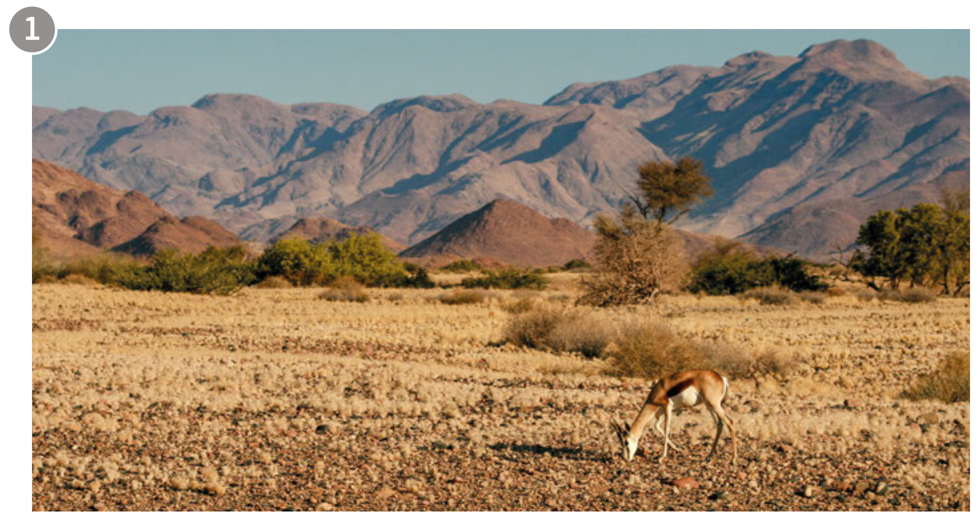
The original model for the Biodiversity Intactness Index predicted, perhaps counterintuitively, that biodiversity is higher in primary forests with more intense human use. The updated model does a better job of presenting biodiversity loss in heavily used forests, like these cleared forests in Cameroon.

Source: Baudouin Desclee, with permission, all rights reserved.



Source: Newbold, T. et al. (2015) Global effects of land use on local terrestrial biodiversity. Nature 520, 45 – 50.

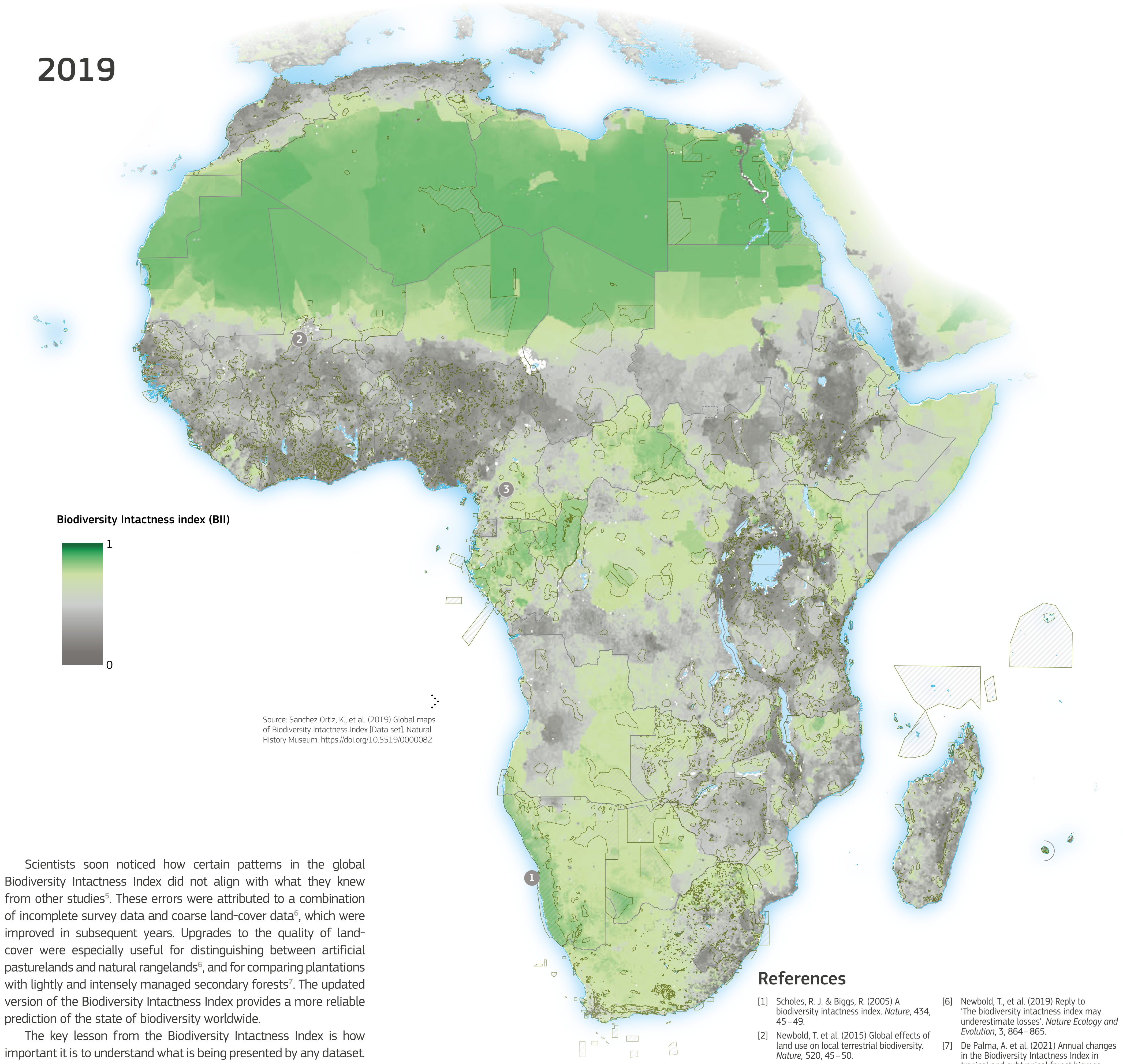
Sometimes we observe significant changes in biodiversity through time, but we should be careful of reading too much into these changes. The technical difficulty of measuring biodiversity for the whole continent means that scientists rely on predictive models. When these models are improved and updated, their predictions can change dramatically even when the underlying biodiversity features stay the same.



Springbok, Namib Desert, Namibia.

The original model of the Biodiversity Intactness Index underestimated biodiversity in arid rangelands with low human population densities, like the Namib Desert in Namibia. But such areas might have naturally low biodiversity due to their hot and dry climates, not because of human-driven biodiversity loss.

Source: Grégoire Dubois, with permission, all rights reserved.



Source: Sanchez Ortiz, K., et al. (2019) Global maps of Biodiversity Intactness Index [Data set]. Natural History Museum. https://doi.org/10.5519/0000082

Scientists soon noticed how certain patterns in the global Biodiversity Intactness Index did not align with what they knew from other studies[5]. These errors were attributed to a combination of incomplete survey data and coarse land-cover data[6], which were improved in subsequent years. Upgrades to the quality of land-cover were especially useful for distinguishing between artificial pasturelands and natural rangelands[6], and for comparing plantations with lightly and intensely managed secondary forests[7]. The updated version of the Biodiversity Intactness Index provides a more reliable prediction of the state of biodiversity worldwide.

The key lesson from the Biodiversity Intactness Index is how important it is to understand what is being presented by any dataset. It would be tempting – not to mention technically straightforward – to compare the Biodiversity Intactness Index between 2015 and 2019 and draw conclusions about biodiversity dynamics during that period. But this would be a major mistake. Therefore, scientists have a duty to guide policy officials to understand when data represent real changes in biodiversity and when they only represent improvements in data collection, monitoring, and modelling.

References

[1] Scholes, R. J. & Biggs, R. (2005) A biodiversity intactness index. *Nature*, 434, 45 – 49.

[2] Newbold, T. et al. (2015) Global effects of land use on local terrestrial biodiversity. *Nature,* 520, 45 – 50.

[3] Newbold, T. et al. (2016) Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science*, 353, 288 – 291.

[4] Hudson, L. N. et al. (2017) The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecology and Evolution*, 7, 145 – 188.

[5] Martin, P. A., et al. (2019) The biodiversity intactness index may underestimate losses. *Nature Ecology and Evolution*, 3, 862 – 863.

[6] Newbold, T., et al. (2019) Reply to 'The biodiversity intactness index may underestimate losses'. *Nature Ecology and Evolution*, 3, 864 – 865.

[7] De Palma, A. et al. (2021) Annual changes in the Biodiversity Intactness Index in tropical and subtropical forest biomes, 2001 – 2012. *Scientific Reports*, 11, 20249.