Type of paper: Review

Problems of geometry, sampling, and scale in gridded biodiversity data, and proposed solutions

Authors: Petr Keil[1,*], Friederike Wölke[1,2], Gabriel Ortega[1], Florencia Grattarola[1], Francois Leroy[1,3], Florian Jansen[4], Carmen Soria[1]

[1] Faculty of Environmental Sciences, Czech University of Life Sciences in Prague, Kamýcká 129, 165 00 Praha – Suchdol, Czech Republic

[2] Department of Ecology, Environment and Plant Sciences, Stockholm University, Svante Arrhenius väg 20 A, 106 91 Stockholm, Sweden

[3] Department of Evolution, Ecology and Organismal Biology, The Ohio State University, Columbus, Ohio, 43210, USA

[4] Faculty for Agriculture, Civil and Environmental Engineering, University of Rostock, Justus-von-Liebig-Weg 6, 18059 Rostock, Germany

* Correspondence: keil@fzp.czu.cz

Highlights

- Standardized gridded biodiversity data are a valuable resource for ecology, yet they come with problems of geometry, sampling, and scale.
- Here we review such problems, and we provide a selection of solutions.
- Some of our advice relates to steps during the creation of gridded data, particularly with regard to geometry.
- Other solutions involve manipulating existing grids or using specialized statistical models.
- We hope that this review will serve as a useful resource, particularly for researchers who are new to gridded data.

Abstract

Grids, and gridded biodiversity data such as regional or country-level atlases, play a prominent role in ecology, particularly in the study of spatial patterns of species occupancy, geographic ranges, biodiversity, and their drivers and temporal dynamics. However, managing, exploring, and analyzing data in grids comes with problems. Here, we review the problems with gridded data, and the existing solutions. We focus on grid-specific problems of sampling (e.g. varying sampling method and effort in space and time, imperfect detection), geometry (e.g. varying grid cell area and shape, positional errors), and scale (e.g. spatial grain and temporal extent). A first group of solutions can be implemented prior to gridding of the data. This includes the selection of an appropriate geographic projection, grid grain, and grid cell shape. The second type of solution involves the manipulation and processing of the gridded data. Examples include aggregating cells to coarser grains or removing cells that fail to meet certain quality criteria. The third type of solution is implemented during the analysis of the data. The most important is the quantification of the problem for use in statistical models or machine learning algorithms as a covariate. We hope to provide guidance particularly to early-career ecologists who may otherwise struggle to make sense of the various solutions scattered through the literature.

Key words: biogeography, scale, macroecology, squares, species-area, bias, observation error, imperfect detection

1 Introduction

In biogeography, macroecology, and applied ecology, data on species distributions are often arranged in a grid (Fig. 1, Box 1). Prominent examples are gridded atlases (Asher, 2001; Preston, 2013; Keller *et al.*, 2020; Šťastný *et al.*, 2021) (Table 1). Compared to local and spatially discrete data from local observations as in most data in GBIF(www.gbif.org) or in USGS North American Breeding Bird Survey (Ziolkowski *et al.*, 2022), grids have the advantage that their cells are directly adjacent. This enables mapping the data with no blank spots, and it also ensures a thorough, continuous coverage of environments—useful when, for instance, modelling distributions of species that are absent in some suitable locations (Svenning & Skov, 2004), or when assessing occupancy and species ranges (Warren *et al.*, 2001). Further, because grids are contiguous, we can assess not only occupancy, but also spatial patterns such as aggregation or fractality (Gaston, 2003). Importantly, adjacent cells can be aggregated to create coarser cells (Rahbek, 2005; Hurlbert & Jetz, 2007), or can be analyzed using moving windows (Lennon *et al.*, 2001). This is necessary for analyses of scale-dependent phenomena such as species-area relationships (Storch, 2016), effects of abiotic environment on species-environment associations (Belmaker & Jetz, 2011), or occupancy-area relationship (Kunin, 1998), including their change in time (Leroy *et al.*, 2024). Gridded biodiversity data are thus a vital ecological resource.

Managing, exploring, and analysing data on species distributions in a grid, however, comes with problems. Perhaps the most discussed in the literature is the problem of incomplete or biased sampling (Isaac *et al.*, 2014; Boyd *et al.*, 2021, 2024; Rapacciuolo *et al.*, 2021). Less discussed are problems with geometry (Kunin, 1997) and scale (Hurlbert & Jetz, 2007; Boyd *et al.*, 2024). Furthermore, there may be problems, decisions, and solutions that are not part of methodological overviews, but are described in specific biodiversity analyses in methods sections, appendices, protocols, and workflows. These may be scattered, and motivations for decisions are not always well described or justified: we may see what people did, but we don't know why.

In this paper, we review the problems with gridded data, summarize their causes, and propose solutions. We focus specifically on problems of geometry, sampling, and scale, as these are closest to our expertise. The solutions that we offer are all aimed at (i) making the raw gridded data more informative in visualizations and exploratory analyses, and (ii) making the data more useful for further statistical modelling. Rather than offering a unified solution to all the problems, we provide a toolbox of many existing solutions, hoping that the reader will be able to choose one. We should also add that we review problems and solutions which apply equally to existing grids and gridded atlases, and when designing new grids. Ultimately, these solutions are proposed with the goal of having accurate maps and adequate models of species distributions and diversity which represent reality and provide good predictions and inference.

Box 1 – Basic definitions

Here we use the term **grid** (also called a mesh) for a structure consisting of multiple grid **cells** (polygons) of non-zero area, arranged in a **contiguous** way so that each grid cell is adjacent to another (Fig. 1); we thus leave out point meshes. In our view, grid cells may have varying area and shape even within one grid, and thus we also consider a set of adjacent countries (Essl *et al.*, 2013) or a set of smaller adjacent administrative units (Craven *et al.*, 2020) to be grids.

Gridded biodiversity data describe the distribution of one or more species, or their diversity, across the grid. A typical example is a gridded **atlas** (Table 1). Observations in the data are always defined per grid cell, i.e. one row in the data table is a record from a single grid cell. These can be detections or non-detections of species, their presence and/or absences, some metrics or proxies for abundance, species richness, and other indices of diversity.

The average area of the cell is the grid's **spatial grain** (resolution), and the distance or area between the grid's outer boundaries is the grid's **spatial extent**. For some ecological quantities (biodiversity and related concepts such as beta diversity, species associations, or species composition), there is no "correct" or "best" grain of a grid, and fine-grain grids are not better than coarse ones, they simply carry different information (Rahbek, 2005). This may come as a surprise to researchers from remote sensing, where the finer the raster resolution, the better. Finally, a **grid origin** is the exact geographic location of a selected fixed point of the grid; if we keep the grain and extent constant, but we change the origin, we are effectively moving the grid around the map.

Structured vs semi-structured vs unstructured gridded data (Table 1). The best quality gridded data derive from **structured** surveys with standardized sampling protocols, with pre-specified sampling method, effort, site selection, and temporal replication. A second case of gridded dataset derives from **semi-structured** surveys. These do not follow standardized sampling efforts, survey site selection, nor have a regularly specified replication. However, the metadata associated with the data collection hold information on the survey methods and effort. Finally, gridded data can be created using **unstructured** sampling processes (i.e., opportunistic observations) by overlapping point observations with the grid. In this case, the observation process is unknown, and there is little to no metadata on the biases inherent in the data collection. Many museum collections and most citizen science initiatives (e.g., eBird, iNaturalist, Observation.org) gather unstructured point occurrences (Pocock *et al.*, 2017) widely available through GBIF.

Table 1. Examples of gridded atlas datasets, ordered from most structured (top rows) to least structured (bottom rows).

Dataset	Type	Spatial extent	Spatial grain	Temporal extent	Sampling effort	Effort units per cell
Birds of Japan (植田 *et al.*, 2021)	Structured	Japan	20×20 km^2	2016-2021	Known, fixed	
Czech bird atlas (Šťastný *et al.*, 2021)	Semi-structured	Czech Republic	11.2×12 km^2	2014-2017	Known but varies	# cards, # observers
SABAP2 - South African bird atlas (Brooks *et al.*, 2022)	Semi-structured	South African countries	5'×5'	2007-present	Known but varies	# cards, hours
Mammals of Hamburg (Schäfers *et al.*, 2016)	Unstructured	Hamburg, Germany	2×2 km^2	1997-2015	No	

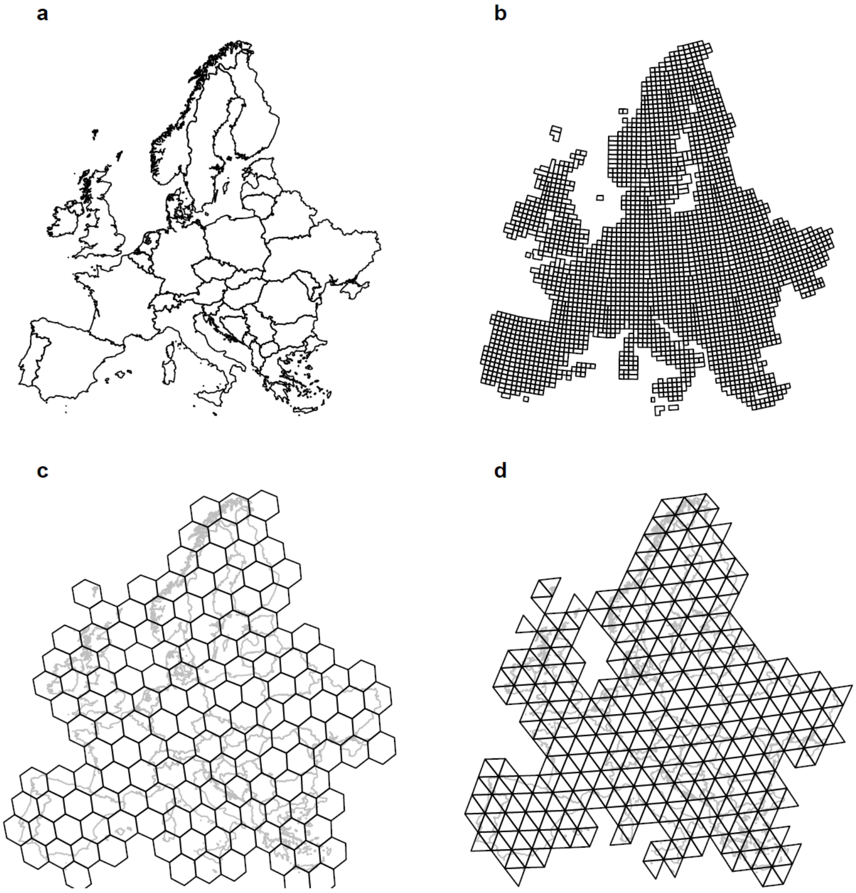


Figure 1. Examples of spatially contiguous grids. (a) A grid where cells are European countries, and they vary considerably in their area and shape. (b) A rectangular UTM grid used in (Keller *et al.*, 2020); note the irregular non-square shapes of some cells. (c, d) Hexagonal and triangular grids; their cells are only minimally distorted by geographic projections. Note that we are not dealing with grids of spatially discrete points in this review.

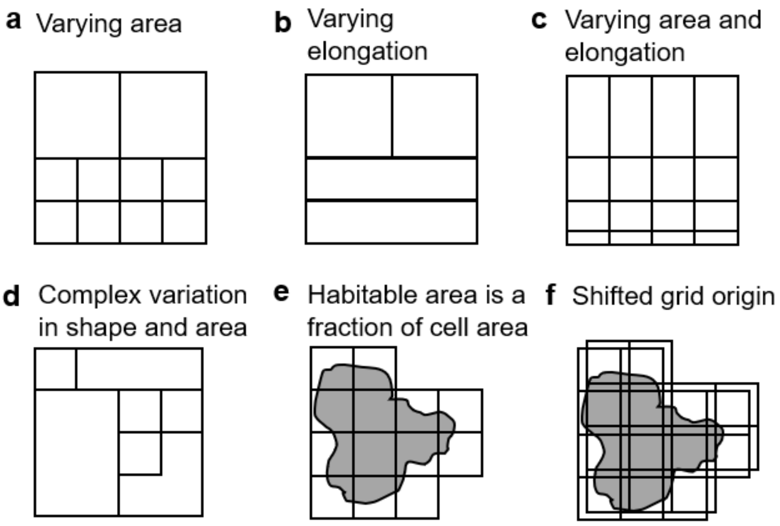


Figure 2. Common geometrical problems of gridded biodiversity data.

2 Geometrical problems

Problems of geometry affect almost every grid to some degree. These problems include variation of grid cell shape, or variation of the shape of the habitable area within cells (Fig. 2). Many of these are caused by geographic projections. For example, in rectangular grids with global extent and equal-area projections, there is increasing north-south elongation of cells towards the poles. An example is the Behrmann projection of global vertebrate grids (Faurby *et al.*, 2018; Coelho *et al.*, 2023), or the European 50 x 50 km^2 UTM grid (European bird census council, 2020) with increasing fraction of small grid cells northwards. Second, cells may have an irregular shape by definition; these can be countries, biogeographic regions, or counties (e.g. Craven *et al.,* 2020). Finally, the problem often arises when grids are created by aggregation ("coarsening") of other grids, when coarse-grain grid cells are not fully filled by fine-grain cells, particularly around the grid's outer boundary (Keil *et al.*, 2011).

2.1 Overview of geometrical problems

Varying cell area. Some cells in a grid can be larger than others (Fig. 2a). This severely affects prevalence (occupancy) of every species, since it follows a positive occupancy-area relationship (Kunin, 1998); in other words, the larger the cell, the larger the probability that a species is in it. Consequently, variation of cell area can mask, or distort, geographic patterns and other drivers of distributions. The same problem applies to species diversity, which follows the species-area relationship (Arrhenius, 1921; Storch, 2016): the larger the grid cell, the more species are in it. Furthermore, variation in cell area distorts beta diversity, species co-occurrences, and species composition in grids, since they are all derived from species distributions and/or from diversity patterns.

Varying cell elongation. Some cells can be elongated more than others (Fig. 2b-d). Given a constant grid cell area, elongated grid cells have more species than square or circular cells (Kunin, 1997), affecting patterns of species occupancy, diversity, and beta diversity. This is because of Tobler's law (i.e. spatial autocorrelation), which states that closer locations have more similar species composition than distant ones (Nekola & White, 1999). One explanation is that the more elongated the shape, the larger the variety of environments in it (but note that the law also emerges in homogeneous environments, Rosindell *et al.*, 2011). Consequently, more elongated cells can accommodate longer distances between sites, and thus larger variation in environments.

Habitable area is a fraction of cell area. The world is full of boundaries and barriers that delineate habitable areas for species (Fig. 2e). An example is the coastline delimiting land for terrestrial species. When we overlay a grid cell over the boundary, we end up with only a fraction of the cell covered by the habitable area; for example, in the grid of the European bird atlas (Keller *et al.*, 2020) the only land in a ~2,500 km^2 cell is Ibiza, an island of 573 km^2, which is 22.9% of the cell area. Because of this small inhabitable area, the number of terrestrial species in the cell is likely lower than in an inland cell of continental Spain with 100% of mainland. Consequently, even in a regular grid with constant cell area and shape, variation of the fraction of habitable area may distort patterns of species occupancy, distributions, and diversity. Indeed, sharp habitat or biome boundaries may cause the same problem.

Shifted grid origin. When we shift the whole grid (we move the origin of the grid, Fig. 2f) by less than a grid cell length, we affect occupancy and apparent distribution of species (Witte *et al.*, 2008), and consequently also patterns of diversity. This problem emerges during creation of new grids, or in already existing older and newer grids in the same region, e.g. in the Czech atlases (Šťastný *et al.*, 2021). The problem of shifted origin gets more serious the coarser the grain of the grid (Witte *et al.*, 2008) because (i) there are fewer cells, and thus smaller sample size for statistical analyses, and (ii) because rare species located close to cell boundaries can completely reverse the spatial pattern.

Small number of grid cells at coarse resolutions or small extents. Some grids have a small number of grid cells natively, others because they have been created by aggregation (coarsening) of finer grids (Rahbek, 2005; Craven *et al.*, 2020). If the number of cells becomes too low, it may compromise statistical estimators that need sufficient sample size.

2.2 Solutions to geometrical problems

Use equal-area projection. A popular way to avoid varying area caused by geographic projections is to use a regular grid in an equal-area projection; these projections can be custom tailored for any region using projectionwizard.org. However, note that, for rectangular grids, equal-area grids cause variation in elongation, which also affects diversity and distributions (Kunin, 1997).

Use triangular or hexagonal grids. Triangular and hexagonal grids don't require special projections, and they keep area, distance between cell centroids, and shape constant. In other words, they can be wrapped around a sphere while avoiding shape distortion. An example use of a global hexagonal grid is in Keil & Chase (2019). The disadvantage of hexagons is that they look unfamiliar, they cannot be easily converted to rasters, and are challenging to be made nested, i.e. aggregated to coarser resolutions [but see Jurasinski & Beierkuhnlein (2006)]. Triangular grids can be coarsened and made nested, but we are unaware of any use of a triangular grid in biogeography.

Record the area and shape for each cell. If the previous solutions can't be applied (e.g., because the grid is given and we have no control over its geometry), or when the habitable area is just a fraction of grid cell area, then the geometrical problems should be quantified and noted. We suggest that, for each cell in the grid, metadata are provided that specify: cell area, fraction of habitable area within the cell (e.g. fraction of mainland if terrestrial organisms are mapped), fraction of cell area covered by smaller grid cells (if the grid is an aggregate of a finer grid), and some metrics of shape deformation, e.g. circularity, linearity, or elongation ratio (Stojmenovic & Nayak, 2008). These metrics can then be used as covariates in statistical models, such as in Leroy *et al.* (2024), or they can be used to filter out problematic cells that don't meet certain criteria (see below).

Remove cells with undesirable geometry. These can be cells with excessive shape distortion (too elongated), cells with a small fraction of land, or cells that are too small or too large. Examples of this are cells overlapping coastal areas or the thin cells — sometimes less than 1 km wide but 10 km long — created when two or more contiguous UTM zones are used to build a grid. A simple and brute-force solution is to remove such cells, which usually requires a threshold to be specified, for example a threshold of a minimum area of land within a grid cell. Threshold values for this removal can vary from strict [e.g. only include cells with 100% of land (Keil *et al.*, 2012)] to subjective, and there will likely be a tradeoff between sample size (number of cells) and the threshold strictness.

Test sensitivity of patterns to grid origin. This can be wise particularly for coarse grids with few grid cells. When a pattern changes considerably with grid origin, avoid general conclusions. When a fine-grain grid of the same extent as the coarse-grain grid is available, we can use moving windows (Lennon *et al.*, 2001), i.e. a large sampling cell is moved continuously over the fine grain grid, and values such as diversity are calculated per fine-grain cell in the center of the sampling cell. This is equivalent to overlaying many coarse-grain grids with different origins.

Coarsen up the grid. Some problems of varying cell area and shape can be fixed by aggregation of smaller cells to larger units, as done for example in (Craven *et al.*, 2020). This can be also done to ameliorate problems of positional errors (Moudrý & Šímová, 2012), although there is some controversy around that (Gábor *et al.*, 2022).

Use area and shape in statistical models. When cells vary in their area (both, the cells themselves, and the proportion of the cell covered by habitable land), there are ways to include the effect of area in statistical models. First, a mathematical transformation of area values may be necessary, usually log- or square root transformations. Second, the transformed area can be used as a predictor in models of diversity and distribution, ideally also as an interaction term with other predictors [e.g. see (Keil & Chase, 2019)]. The latter stems from the well-established observation that sample area not only affects diversity and distributions per se (Storch *et al.*, 2007), but it also modulates the effects of other variables (Belmaker & Jetz, 2011). Furthermore, some types of statistical models assume a very specific effect of area so that it is forced into models as an offset (Isaac *et al.*, 2019). Regarding shape attributes, elongation is probably the most important for common, square cells. Elongation can vary due to spatial distortion in grids covering a large extent (i.e. a continent) or because of procedural artifacts (i.e. merging adjacent UTM zones). Compared to area, there is much less research on how (and if) elongation should be transformed, and how it interacts with other predictors. Here, we recommend standard model selection procedures to assess the effect of elongation, as well as investigating the statistical interaction between shape and area.

Avoid generalizations from a few cells. When the number of cells is small, for example at very large grains, it may still be useful to calculate some quantities per grid cell; for example, we may be interested in the total number of species (species pool) in a large region corresponding to a large grid cell. However, statistical generalizations based on these small samples are not advisable. If generality is desirable, use a finer grid with more grid cells. Alternatively, Bayesian statistics can quantify the uncertainty caused by small sample size (Gelman *et al.*, 2013).

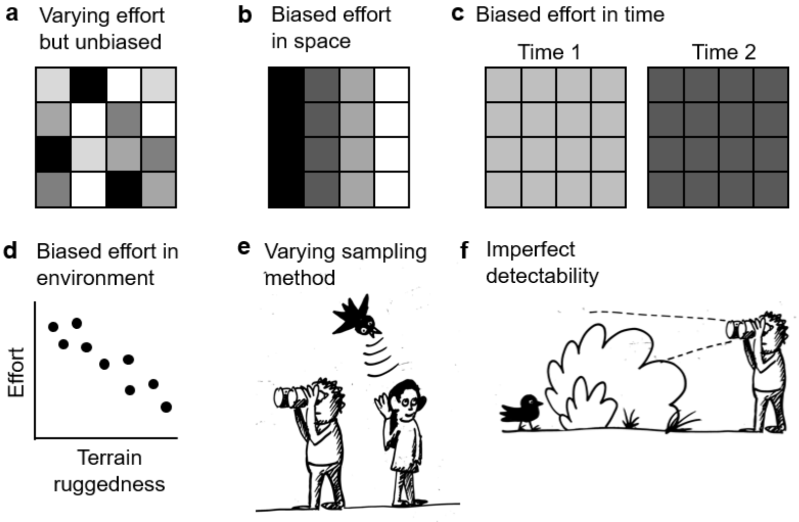


Figure 3. Common sampling problems of gridded biodiversity data.

3 Sampling problems

Sampling problems (Fig. 3), specifically those related to variation in sampling effort and methods, are perhaps the most common, and potentially the most influential, in gridded biodiversity data. They were also a subject of several reviews (Robertson *et al.*, 2010; Botts *et al.*, 2012; Isaac *et al.*, 2014; Boyd *et al.*, 2023).

3.1 Overview of sampling problems

Varying sampling effort. Sampling effort can vary geographically, or temporally, which can render some grid cells under-sampled. This may affect some ecological quantities, e.g. occupancy of species, average richness, or regional richness. However, if the variation is random in space (Fig. 3a), time, or environment, it may not be such a problem for geographic, temporal, or environmental gradients of diversity and distributions. However, this randomness will likely be rare.

Biased sampling effort in space, time, or environment. This is a special case of the abovementioned problem, when sampling not only varies, but is also systematically biased in space (Fig. 3b), for instance when some regions are better accessible or attractive. Temporal biases (Bowler *et al.*, 2022) (Fig. 3c) can emerge when the weather (or climate) in certain time periods is better for sampling, or with the general increase in biodiversity records towards the present (https://www.gbif.org). This can lead to biases in environmental space (Xu *et al.*, 2024), e.g. when some environments are less accessible or attractive to observers (Fig. 3d).

Sampling effort or method unknown. Sometimes we may have no information on the sampling effort. This can be further complicated when sampling methods are unknown, or insufficiently described. In such cases it may seem impossible to determine if non-detections of species are real absences.

Varying skill among observers. Ecological data collection usually involves many observers with various levels of expertise, impacting data quality (Isaac *et al.*, 2014). Less-skilled observers may work slowly, are more likely to miss cryptic species, and tend to misidentify species. Consequently, even if sampling effort in observer hours per grid cell is constant, the real sampling may vary, i.e. it is lower in cells surveyed by unskilled observers (Prendergast et al. 1993). Since identification relies on visual and auditory cues, the ability of observers to detect the species around has been shown to decay with age (Farmer *et al.*, 2014).

Imperfect detectability. Detecting all individuals in a grid cell is not guaranteed due to combination of imperfect detectability and insufficient sampling effort, resulting in false negatives and inaccurate distributions (MacKenzie *et al.*, 2002). The probability of detection at a grid cell depends on species abundance, conspicuity, and behavior, which are influenced by the cell characteristics (habitat, disturbance level), environmental conditions (including season, phenology, weather, time of the day), and survey method, timing, effort, and observer skill (Guillera-Arroita, 2017). As perfect surveys are unlikely, most raw gridded biodiversity data tend to have a lower number of detected presences than true occurrences of species.

3.2 Solutions to sampling problems

Get data on sampling methods and effort. We advise seeking data on sampling effort and methods in gridded biodiversity data whenever possible; if unavailable publicly, they may still be possible to get from the authors. These data can take many forms, e.g. number of survey cards or observers per grid cell, number of observer hours, number of traps, time of trap exposure, number of transects and their length, or surveyed point locations. Sampling effort may be further standardized to a relative measure of sampling *completeness* (Keller *et al.*, 2020).

Get data on recorder identity. Knowing the identity of recorders can be used to determine *recorder syndromes* (Isaac & Pocock, 2015), i.e., typical sets of recording behaviors which can be estimated from species list lengths, coverage of rare or difficult-to-identify species, frequency of recording etc. These can then be used in statistical models that account for sampling variation (Isaac & Pocock, 2015).

Use external proxies of sampling effort. In unstructured or semi-structured data, sampling effort often correlates with accessibility, e.g. modelled as distance to human settlements (Monsarrat *et al.*, 2019) or distance to roads and research institutions (Meyer *et al.*, 2015). These data can be extracted from public GIS layers and used as proxies for sampling effort in statistical occupancy models (see below).

Derive proxies of effort from occurrences. Some methods use detections and non-detections of species to infer proxies of sampling effort. For instance, Szabo *et al.* (2010) proposed to use the number of species present in a grid cell as an indicator of sampling effort; we see this method as questionable, since it confuses effort with one of the fundamental ecological quantities of interest. In a better alternative, (Hill, 2012) proposed to use the frequency of the so called “benchmark species” that are present in the neighboring cells as an indicator of sampling effort. Encouragingly, we found that this frequency correlates well with some real metrics of sampling effort (*in preparation*). Hill (2012) then demonstrates how this can be used, in what is now known as the *Frescalo algorithm*, to infer species occupancy trends over time or to identify unexpected presences or absences. Studies that use the algorithm are (Eichenberg *et al.*, 2021; Montràs-Janer *et al.*, 2024). There is also a practical guide on how to use the algorithm (Goury *et al.*, 2025).

Remove (filter out) under-sampled cells. Once we have sampling effort, or its proxy, a crude but simple solution is to remove sites with sampling effort below a certain threshold. This comes at the expense of data loss, since even well-sampled cells in one atlas period may need to be removed if the same cells were under-sampled in another period. A good overview of filtering approaches is provided by (Hughes *et al.*, 2024).

Coarsen up the grid. In macroecology, it has been recognized (Hurlbert & Jetz, 2007) that imperfectly sampled fine-scale data converge to true presences and absences if they are aggregated to a sufficiently coarse grids. Coarser grids also give more accurate estimates of species occupancy, i.e. the proportion of occupied grid cells (Boyd *et al.*, 2024). In other words, there is a minimum reliable grid resolution (Keil *et al.*, 2014) above which we are safe to assume true presence and absence. This has been demonstrated to work for spatially random variation in sampling (Fig. 3a) (Figure S1 in Keil *et al.*, 2014). This is also the reason why global biodiversity grids have grains of 100 x 100 km^2 and coarser (Orme *et al.*, 2006; Faurby *et al.*, 2018) and why finer global grids (Jenkins *et al.*, 2013) are problematic.

Use rarefaction. In a set of studies related through their data and authors (Biesmeijer *et al.*, 2006; Keil *et al.*, 2011; Carvalheiro *et al.*, 2013), a variant of *individual-based rarefaction* (Gotelli & Colwell, 2001; Hsieh *et al.*, 2016) was used to account for variation in sampling effort in gridded data, where records in each grid cell were down-sampled to a standard number to obtain standardized number of species per grid cell. Another example of the use of rarefaction over a country-wide grid is Grattarola *et al.* (2020). However, we still lack validation of this approach in the gridded setting either on simulated data, or on data with known true richness.

Ad-hoc corrections of metrics of change. (Isaac *et al.*, 2014) provides an overview of simple statistical tricks that have been proposed to control for varying effort directly in the calculation of metrics of temporal change. Examples of these include (Telfer *et al.*, 2002)’s estimation of change in species occupancy as a deviation from the overall occupancy trend of all species, or (Ball *et al.*, 2011)’s proposition to model species status in a cell as the proportion of total records in the grid cell. Although inspirational, these are specific solutions linked to specific metrics and data types.

Use sampling effort as a covariate. An easy-to-apply method of controlling for sampling problems is to use additional data on sampling effort, method, or their proxies, in simple statistical models such as GLM or in machine learning. The formula of this approach is $Y = f(X, effort)$, where Y is the ecological quantity of interest (e.g. occupancy or species richness) and X are predictors not related to sampling. This approach is valid, but it has a key limitation: It does not estimate the true occurrence of species (or species richness); it only estimates frequency of detections and richness of detected species, given a fixed value of sampling effort. We thus need to assume that frequency of detections corresponds to the actual species occurrence, which may not hold (MacKenzie *et al.*, 2002). Further, the effect of effort on ecological quantities can be non-linear (Gotelli & Colwell, 2001) and effort also interacts in its effect with other variables [sometimes those affecting species detection; (Kéry & Royle, 2015)], making such simple statistical approach problematic.

Use occupancy models. There is a class of models--hereafter *occupancy models* (MacKenzie *et al.*, 2002; Guillera-Arroita, 2017) which separate the observation (detection) process from the true but unobserved ecological quantity such as occurrence of a species. For gridded data, convenient software packages are available (Doser *et al.*, 2022). Occupancy models can be used to account for varying effort and imperfect detection when estimating the probability of a species' presence. These models typically rely on repeated site visits, but these are not always available. To address single-visit versions of occupancy models have been developed (Lele *et al.*, 2012) and successfully applied to gridded atlas data (Sadoti *et al.*, 2013; Peach *et al.*, 2017; Klinkovská *et al.*, 2024), they either rely on external predictor of the observation process (it needs to be orthogonal to predictors of the ecological process), or on space-for-time substitution where adjacent grid cells are aggregated, which reduces the effective sampling size. Importantly, (Pescott, 2026) presents a modelling approach in which occupancy models can be combined with a sampling effort derived from the Frescalo algorithm.

Leverage external data with integrated models. Sometimes there can be other data types available for the region of the gridded atlas, and these can improve models that are based on the atlas data with varying sampling effort. Specifically, there are species distribution models which can calculate joint likelihood of multiple types of biodiversity data (Dorazio, 2014; Isaac *et al.*, 2019), and we call these the *integrated models*. These have been used to integrate gridded records with local surveys to improve predictions of population trends (Pagel *et al.*, 2014) or to integrate gridded records with expert range maps and camera traps to predict large-scale dynamics of geographic ranges (Grattarola *et al.*, 2024). For an overview of the uses of integrated models in gridded citizen science data see (Johnston *et al.*, 2023).

Borrow strength from other taxa in joint models. Joint species distribution models [JSDM; (Pollock *et al.*, 2014; Ovaskainen *et al.*, 2017)] are another tool that can strengthen the inference and predictions of true species distributions from data affected by the sampling issues. In the case of gridded biodiversity data affected by sampling issues, the idea is that different species may be affected by the same sampling issues, and thus data on other species can improve estimation of the sampling (observation) process (Rapacciuolo *et al.*, 2021). This is closely related to the abovementioned (Hill, 2012)'s Frescalo algorithm, which also assumes a sampling process shared across species. Furthermore, thanks to niche conservatism, phylogenetically or functionally similar species may also have similar environmental requirements; modelling this in JSDMs can improve estimation of the ecological process (McInerny & Purves, 2011; Ovaskainen *et al.*, 2017) even if the target species is undersampled. A related idea is to model ecological quantities at multiple taxonomic levels (Adjei *et al.*, 2024).

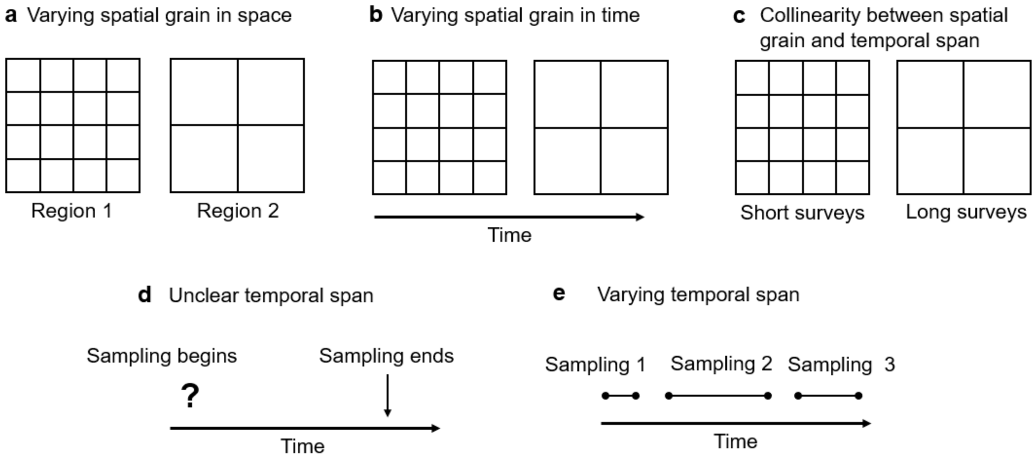


Figure 4. Common problems of gridded biodiversity data related to spatial and temporal scale.

4 Problems of scale

Scale has several meanings. Here, we deal with two aspects of scale: (1) spatial and temporal *grain* or resolution, which is an average area of a grid cell or average temporal duration of an observation, and (2) spatial and temporal *extent*, which is the total area of a region covered by a gridded dataset, or the total temporal period covered by a series of gridded atlases in the same region. For instance, the last three editions of Czech bird atlases (Šťastný *et al.*, 2021) has spatial grain of 11.2×12 km, spatial extent is the whole Czech Republic, temporal grain is roughly 4 years (the average time covered by one edition), and a temporal extent of 30 years (from 1985 to 2015). Most scaling problems emerge when comparing atlases, or atlas editions, with different scales, or when one wants to generalize results from a gridded dataset outside of its extent and grain.

4.1 Overview of problems of scale

Varying spatial grain in space or time. Spatial grain of the grid affects species commonness and rarity (Hartley & Kunin, 2003), patterns of species diversity (Rahbek, 2005), turnover (Keil *et al.*, 2012), aggregation (Soria *et al.*, 2026), and also associations of these with the environment (Belmaker & Jetz, 2011). This also includes species-distribution models (SDM), which can lead to different inference and predictions depending on the grain of the data (Mertes & Jetz, 2018). The effect of temporal grain on patterns of species diversity and distribution is much less studied, but there are good theoretical and empirical reasons for temporal grain being similarly important as spatial grain, mostly due to the species-time-area relationship (Adler *et al.*, 2005). All this means that analyses of gridded data at one grain may be hard to compare to analyses using different grains.

Collinearity between spatial grain and temporal span. We expect that the larger the spatial grain of gridded data, the longer the temporal span (period) over which the data were aggregated (Fig. 4c). For instance, the 50×50 km European gridded plant atlas (Lahti & Lampinen, 1999) covers a long period of time of more than 100 years, while smaller per-country gridded data [e.g. (Stroh *et al.*, 2023b; Klinkovská *et al.*, 2024)] cover intervals of past decades. When comparing analyses based on data with different spatial grains, one may then observe differences that may in fact be caused by different temporal spans.

Unclear temporal span. Some gridded data are aggregations of all historical records that are available, with unspecified start year of the aggregation (Fig. 4d). Examples are the per-country butterfly and plant species lists (Van Swaay & Warren, 1999; Essl *et al.*, 2013) or the abovementioned European plant atlas (Stroh *et al.*, 2023a). This can pose a challenge when assessing rates of change between two periods where the first period is undefined (Keil *et al.*, 2018).

Varying temporal span. Sometimes multiple editions of an atlas in a region can vary in the temporal extent that they cover. For instance, the four editions of the Czech bird atlas (Šťastný *et al.*, 2021) cover 5, 5, 3, and 4 years respectively. If used for assessments of temporal trends, a declining trend may be observed over the four editions simply because the first editions have longer temporal extent.

4.2 Solutions to problems of scale

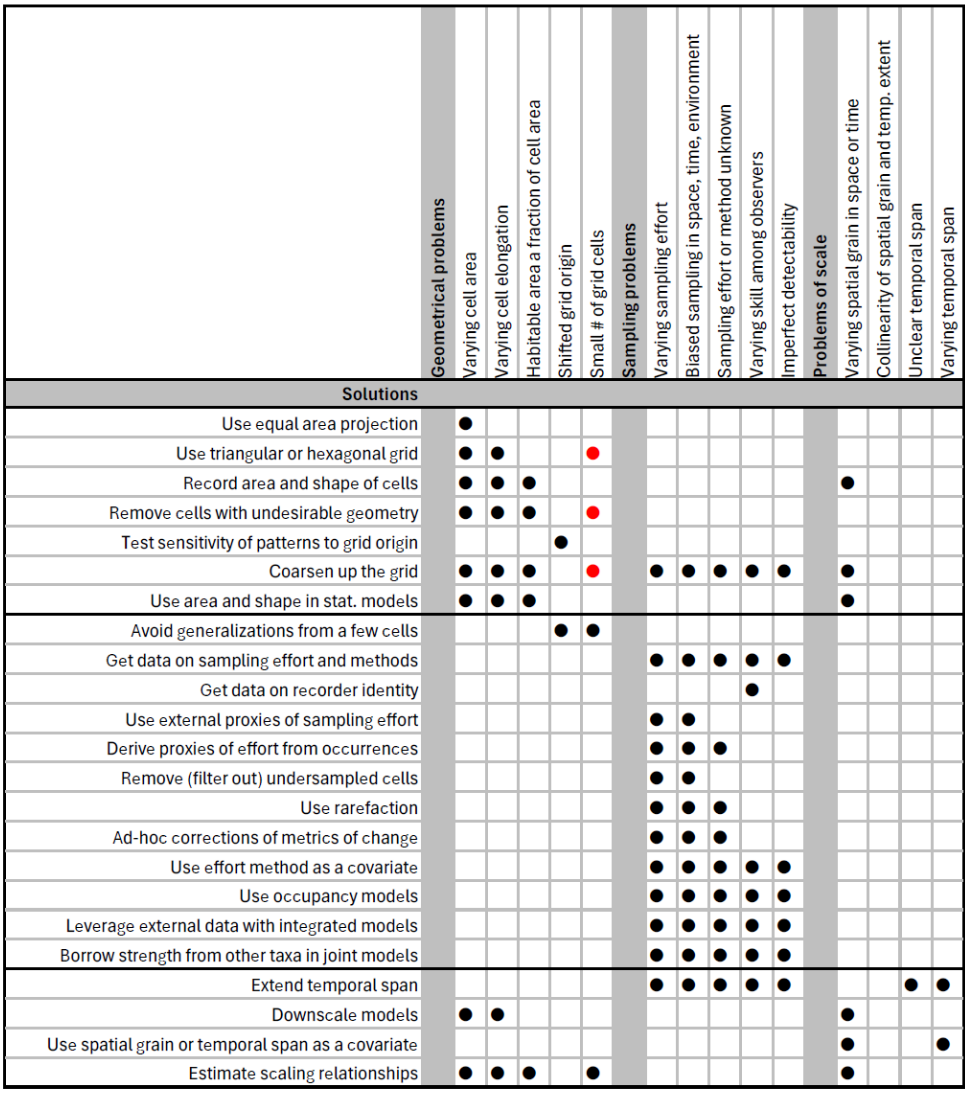
Coarsen up the grid. The simplest solution to the variation in spatial grain is to aggregate some data to a coarse grid resolution to make them comparable with other coarse-grain data. This may also ameliorate some of the sampling issues (as we described in the previous section). The downside is that analyses of coarsened data will capture ecological processes that may differ from fine-grain processes. That does not mean they are incorrect, just different.

Extend temporal span. An analogical solution to the coarsening up the grid in space is merging multiple temporal replications to one, i.e. extending the temporal span. This can fix the issue of varying temporal span (Fig. 4e), but comes at a cost of losing temporal detail, and it is only possible when many temporal replications are available, which can be a rare commodity. To see how extending the temporal span affects patterns and drivers of biodiversity see Mellado-Mansilla *et al.* (2026).

Downscale models. A more advanced and challenging approach is to infer fine-grain distributions or diversity, or ecological processes, from the coarse grain data. This is typically done through hierarchical statistical models where likelihood of fine-grain distributions and parameters are conditional on coarse-grain data (Keil *et al.*, 2013; Keil & Jetz, 2013), or when we assume a scale-free underlying point pattern intensity behind the grid (Isaac *et al.*, 2019). The latter approach also allows integration of gridded data with other types of data for improved inference and predictions (Pagel *et al.*, 2014; Grattarola *et al.*, 2023).

Use spatial grain or temporal span as a covariate. When combining entire datasets (atlases) with different grid resolutions in a single analysis, we can include the grain as a covariate, or as an interaction term in statistical models (Keil & Chase, 2019; Midolo *et al.*, 2025). In the temporal context, this has been done by (Mellado-Mansilla *et al.*, 2026). However, using grain as a covariate is only sensible if differences in grain between atlases are not collinear with other properties (such as geographic or environmental gradients).

Estimate scaling relationships, not single-scale effects. Finally, when faced with variation in grain, it is always useful to consider that ecological quantities such as species richness (Storch, 2016), their patterns such as latitudinal or altitudinal gradients (Rahbek, 2005), and their environmental drivers and coefficients all interact with area, and hence with grain (Field *et al.*, 2009; Belmaker & Jetz, 2011; Keil & Chase, 2019). Let's call this relationship between area (or grain) and ecological quantity, or model coefficient, a scaling relationship. These scaling relationships put both the statistical inference and predictions in context of a given grain, enabling comparisons across datasets of different grains. Here, gridded biodiversity data are especially suitable for estimation of these scaling relationships, as they are spatially contiguous and can be coarsened to a set of nested scales. For examples of this practice see (Hartley & Kunin, 2003; Hurlbert & Jetz, 2007; Keil *et al.*, 2013).



Solutions	**Geometrical problems**	Varying cell area	Varying cell elongation	Habitable area a fraction of cell area	Shifted grid origin	Small # of grid cells	**Sampling problems**	Varying sampling effort	Biased sampling in space, time, environment	Sampling effort or method unknown	Varying skill among observers	Imperfect detectability	**Problems of scale**	Varying spatial grain in space or time	Collinearity of spatial grain and temp. extent	Unclear temporal span	Varying temporal span
Use equal area projection		●															
Use triangular or hexagonal grid		●	●			🔴											
Record area and shape of cells		●	●	●										●			
Remove cells with undesirable geometry		●	●	●		🔴											
Test sensitivity of patterns to grid origin					●												
Coarsen up the grid		●	●	●		🔴		●	●	●	●	●		●			
Use area and shape in stat. models		●	●	●										●			
Avoid generalizations from a few cells					●	●											
Get data on sampling effort and methods								●	●	●	●	●					
Get data on recorder identity											●						
Use external proxies of sampling effort								●	●								
Derive proxies of effort from occurrences								●	●	●							
Remove (filter out) undersampled cells								●	●								
Use rarefaction								●	●	●							
Ad-hoc corrections of metrics of change								●	●	●							
Use effort method as a covariate								●	●	●	●	●					
Use occupancy models								●	●	●	●	●					
Leverage external data with integrated models								●	●	●	●	●					
Borrow strength from other taxa in joint models								●	●	●	●	●					
Extend temporal span								●	●	●	●	●				●	●
Downscale models		●	●											●			
Use spatial grain or temporal span as a covariate														●			●
Estimate scaling relationships		●	●	●		●								●			

Figure 5. Connections between the problems and solutions. Black cells indicate that a solution (row) can help with a given problem (column). Red cells indicate that the solution would make the problem worse. The collinearity between spatial grain and temporal grain is the only problem for which we do not see a solution.

5 Conclusion

We've covered problems that we know from our work. However, there are whole families of important issues that we've missed. Examples are problems of taxonomic standardization and taxonomic scale, problems of data ownership and licensing, or biases in geography and socio-economic history (e.g., most gridded data come from wealthy countries). Each of these would deserve their own section or even a paper.

Among the practices recommended in this paper, we see some as particularly straightforward, or as fixing several problems at a time (Fig. 5): (1) Measuring geometry of cells such as their area or elongation, or within-cell land proportion and elongation, is a smart step which will likely pay off later during analyses. (2) Coarsening the spatial grain of grids is an amelioration of several problems, including sampling and scaling problems. (3) Seeking direct measures of sampling effort, or their proxies [through Frescalo algorithm of Hill (2012)] or external variables such as accessibility or distance to institutions/cities] is always desirable. (4) There are statistical tools which account for problems of geometry, sampling, and scale, and these are hierarchical occupancy models, integrated models, joint species distribution models, and cross-scale models.

In contrast, some of the problems likely don't have a solution at all. These are collinearities of all sorts (Dormann *et al.*, 2013) and lack of information about spatial or taxonomic variation in methods or abrupt changes in methodology between editions of atlases. In such cases, we can only hope to get some additional data or information.

And indeed, none of our proposed solutions are perfect. We expect that there will always be hidden variability and biases, proposed solutions will have unmet assumptions, and some problems have no solution at all. On the other hand, compared to other types of observational data in ecology, gridded biodiversity data are still among the best, and so even limited effort to address their problems can improve the picture of true species diversity, distributions, and their dynamics.

6 Acknowledgements

The research was supported by the European Union (ERC, BEAST, 101044740). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. We thank Oliver L. Pescott and Melanie Tietje for helpful comments.

7 References

Adjei, K.P., Carvell, C., Isaac, N.J.B., Mancini, F. & O'Hara, R.B. (2024) Integrating data from different taxonomic resolutions to better estimate community alpha diversity. *Ecography*, **2024**, e07182.

Adler, P.B., White, E.P., Lauenroth, W.K., Kaufman, D.M., Rassweiler, A. & Rusak, J.A. (2005) Evidence for a general species–time–area relationship. *Ecology*, **86**, 2032–2039.

Arrhenius, O. (1921) Species and area. *Journal of Ecology*, **9**, 95–99.

Asher, J. ed. (2001) *The millennium atlas of butterflies in Britain and Ireland*, Oxford University Press, New York.

Ball, S.G., Morris, R. K. A., Rotheray, G.E. & Watt, K. (2011) *Atlas of the hoverflies of Great Britain (diptera, syrphidae)*, Centre for Ecology & Hydrology.

Belmaker, J. & Jetz, W. (2011) Cross-scale variation in species richness–environment associations. *Glob. Ecol. Biogeogr.*, **20**, 464–474.

Biesmeijer, J.C., Roberts, S.P.M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., Schaffers, A.P., Potts, S.G., Kleukers, R., Thomas, C.D., Settele, J. & Kunin, W.E. (2006) Parallel Declines in Pollinators and Insect-Pollinated Plants in Britain and the Netherlands. *Science*, **313**, 351–354.

Botts, E.A., Erasmus, B.F.N. & Alexander, G.J. (2012) Methods to detect species range size change from biological atlas data: A comparison using the South African Frog Atlas Project. *Biological Conservation*, **146**, 72–80.

Bowler, D.E., Callaghan, C.T., Bhandari, N., Henle, K., Benjamin Barth, M., Koppitz, C., Klenke, R., Winter, M., Jansen, F., Bruelheide, H. & Bonn, A. (2022) Temporal trends in the spatial bias of species occurrence records. *Ecography*, **2022**, e06219.

Boyd, R.J., August, T.A., Cooke, R., Logie, M., Mancini, F., Powney, G.D., Roy, D.B., Turvey, K. & Isaac, N.J.B. (2023) An operational workflow for producing periodic estimates of species occupancy at national scales. *Biological Reviews*, **98**, 1492–1508.

Boyd, R.J., Bowler, D.E., Isaac, N.J.B. & Pescott, O.L. (2024) On the trade-off between accuracy and spatial resolution when estimating species occupancy from geographically biased samples. *Ecological Modelling*, **493**, 110739.

Boyd, R.J., Powney, G.D., Carvell, C. & Pescott, O.L. (2021) occAssess: An R package for assessing potential biases in species occurrence data. *Ecology and Evolution*, **11**, 16177–16187.

Brooks, M., Rose, S., Altwegg, R., Lee, A.T., Nel, H., Ottosson, U., Retief, E., Reynolds, C., Ryan, P.G., Shema, S., Tende, T., Underhill, L.G. & Thomson, R.L. (2022) The African Bird Atlas Project: a description of the project and BirdMap data-collection protocol. *Ostrich*, **93**, 223–232.

Carvalheiro, L.G., Kunin, W.E., Keil, P., Aguirre-Gutiérrez, J., Ellis, W.N., Fox, R., Groom, Q., Hennekens, S., Van Landuyt, W., Maes, D., Van de Meutter, F., Michez, D., Rasmont, P., Ode, B., Potts, S.G., Reemer, M., Roberts, S.P.M., Schaminée, J., WallisDeVries, M.F. & Biesmeijer, J.C. (2013) Species richness declines and biotic homogenisation have slowed down for NW-European pollinators and plants. *Ecology Letters*, **16**, 870–878.

Coelho, M.T.P., Barreto, E., Rangel, T.F., Diniz-Filho, J.A.F., Wüest, R.O., Bach, W., Skeels, A., McFadden, I.R., Roberts, D.W., Pellissier, L., Zimmermann, N.E. & Graham, C.H. (2023) The geography of climate and the global patterns of species diversity. *Nature*, **622**, 537–544.

Craven, D., Sande, M.T., Meyer, C., Gerstner, K., Bennett, J.M., Giling, D.P., Hines, J., Phillips, H.R.P., May, F., Bannar-Martin, K.H., Chase, J.M. & Keil, P. (2020) A cross-scale assessment of productivity–diversity relationships. *Global Ecology and Biogeography*, **29**, 1940–1955.

Dorazio, R.M. (2014) Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, **23**, 1472–1484.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D. & Lautenbach, S. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.

Doser, J.W., Finley, A.O., Kéry, M. & Zipkin, E.F. (2022) spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. *Methods in Ecology and Evolution*, **13**, 1670–1678.

Eichenberg, D., Bowler, D.E., Bonn, A., Bruelheide, H., Grescho, V., Harter, D., Jandt, U., May, R., Winter, M. & Jansen, F. (2021) Widespread decline in Central European plant diversity across six decades. *Global Change Biology*, **27**, 1097–1110.

Essl, F., Moser, D., Dirnböck, T., Dullinger, S., Milasowszky, N., Winter, M. & Rabitsch, W. (2013) Native, alien, endemic, threatened, and extinct species diversity in European countries. *Biological Conservation*, **164**, 90–97.

European bird census council ed. (2020) *European breeding bird atlas*, Lynx Edicions, Barcelona.

Farmer, R.G., Leonard, M.L., Mills Flemming, J.E. & Anderson, S.C. (2014) Observer aging and long-term avian survey data quality. *Ecology and Evolution*, **4**, 2563–2576.

Faurby, S., Davis, M., Pedersen, R.Ø., Schowanek, S.D., Antonelli, A. & Svenning, J. (2018) PHYLACINE 1.2: The phylogenetic atlas of mammal macroecology. *Ecology*, **99**, 2626–2626.

Field, R., Hawkins, B.A., Cornell, H.V., Currie, D.J., Diniz-Filho, J.A.F., Guégan, J., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdorff, T., O'Brien, E.M. & Turner, J.R.G. (2009) Spatial species-richness gradients across scales: a meta-analysis. *Journal of Biogeography*, **36**, 132–147.

Gábor, L., Jetz, W., Lu, M., Rocchini, D., Cord, A., Malavasi, M., Zarzo-Arias, A., Barták, V. & Moudrý, V. (2022) Positional errors in species distribution modelling are not overcome by the coarser grains of analysis. *Methods in Ecology and Evolution*, **13**, 2289–2302.

Gaston, K.J. (2003) *The structure and dynamics of geographic ranges*, Oxford Univ. Press.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2013) *Bayesian data analysis*, 3 edition. CRC Press.

Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Goury, R., Bowler, D., Harrower, C., Münkemüller, T., Vallet, J., Yearsley, J., Thuiller, W. & Pescott, O. (2025) A practical guide to species trend detection with unstructured data using local frequency scaling (Frescalo).

Grattarola, F., Bowler, D.E. & Keil, P. (2023) Integrating presence-only and presence–absence data to model changes in species geographic ranges: An example in the Neotropics. *Journal of Biogeography*, **50**, 1561–1575.

Grattarola, F., Martínez-Lanfranco, J.A., Botto, G., Naya, D.E., Maneyro, R., Mai, P., Hernández, D., Laufer, G., Ziegler, L., González, E.M., da Rosa, I., Gobel, N., González, A., González, J., Rodales, A.L. & Pincheira-Donoso, D. (2020) Multiple forms of hotspots of tetrapod biodiversity and the challenges of open-access data scarcity. *Scientific Reports*, **10**, 22045.

Grattarola, F., Tschernosterová, K. & Keil, P. (2024) A continental-wide decline of occupancy and diversity in five Neotropical carnivores. *Global Ecology and Conservation*, **55**, e03226.

Guillera-Arroita, G. (2017) Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, **40**, 281–295.

Hartley, S. & Kunin, W.E. (2003) Scale dependency of rarity, extinction risk, and conservation priority. *Conservation Biology*, **17**, 1559–1570.

Hill, M.O. (2012) Local frequency as a key to interpreting species occurrence data when recording effort is not known. *Methods in Ecology and Evolution*, **3**, 195–205.

Hsieh, T.C., Ma, K.H. & Chao, A. (2016) iNEXT: an R package for rarefaction and extrapolation of species diversity (H ill numbers). *Methods in Ecology and Evolution*, **7**, 1451–1456.

Hughes, A.C., Dorey, J.B., Bossert, S., Qiao, H. & Orr, M.C. (2024) Big data, big problems? How to circumvent problems in biodiversity mapping and ensure meaningful results. *Ecography*, **2024**, e07115.

Hurlbert, A.H. & Jetz, W. (2007) Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. *Proceedings of the National Academy of Sciences*, **104**, 13384–13389.

Isaac, N.J.B., Jarzyna, M.A., Keil, P., Dambly, L.I., Boersch-Supan, P.H., Browning, E., Freeman, S.N., Golding, N., Guillera-Arroita, G., Henrys, P.A., Jarvis, S., Lahoz-Monfort, J., Pagel, J., Pescott, O.L., Schmucki, R., Simmonds, E.G. & O’Hara, R.B. (2019) Data integration for large-scale models of species distributions. *Trends in Ecology & Evolution*, **35**, 56–67.

Isaac, N.J.B. & Pocock, M.J.O. (2015) Bias and information in biological records: Bias and information in biological records. *Biological Journal of the Linnean Society*, **115**, 522–531.

Isaac, N.J.B., Van Strien, A.J., August, T.A., De Zeeuw, M.P. & Roy, D.B. (2014) Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, **5**, 1052–1060.

Jenkins, C.N., Pimm, S.L. & Joppa, L.N. (2013) Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences*, **110**.

Johnston, A., Matechou, E. & Dennis, E.B. (2023) Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, **14**, 103–116.

Jurasinski, G. & Beierkuhnlein, C. (2006) Spatial patterns of biodiversity - assessing vegetation using hexagonal grids. *Biology and Environment: Proceedings of the Royal Irish Academy*, **106B**, 401–411.

Keil, P., Belmaker, J., Wilson, A.M., Unitt, P. & Jetz, W. (2013) Downscaling of species distribution models: a hierarchical approach. *Methods in Ecology and Evolution*, **4**, 82–94.

Keil, P., Biesmeijer, J.C., Barendregt, A., Reemer, M. & Kunin, W.E. (2011) Biodiversity change is scale-dependent: an example from Dutch and UK hoverflies (Diptera, Syrphidae). *Ecography*, **34**, 392–401.

Keil, P. & Chase, J.M. (2019) Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution*, **3**, 390–399.

Keil, P. & Jetz, W. (2013) Downscaling the environmental associations and spatial patterns of species richness. *Ecological Applications*, **24**, 823–831.

Keil, P., Pereira, H.M., Cabral, J.S., Chase, J.M., May, F., Martins, I.S. & Winter, M. (2018) Spatial scaling of extinction rates: Theory and data reveal nonlinearity and a major upscaling and downscaling challenge. *Global Ecology and Biogeography*, **27**, 2–13.

Keil, P., Schweiger, O., Kuehn, I., Kunin, W.E., Kuussaari, M., Settele, J., Henle, K., Brotons, L., Pe'er, G., Lengyel, S., Moustakas, A., Steinicke, H. & Storch, D. (2012) Patterns of beta diversity in Europe: the role of climate, land cover and distance across scales. *Journal of Biogeography*, **39**, 1473–1486.

Keil, P., Wilson, A.M. & Jetz, W. (2014) Uncertainty, priors, autocorrelation and disparate data in downscaling of species distributions. *Diversity and Distributions*, **20**, 797–812.

Keller, V., Herrando, S., Voříšek, P., Franch, M., Kipson, M., Milanesi, P., Martí, D., Anton, M., Klvaňová, A., Kalyakin, M.V., Bauer, H.G. & Foppen, R.P.B. (2020) *European breeding bird atlas 2: distribution, abundance and change*, European Bird Census Council & Lynx Edicions, Barcelona.

Kéry, M. & Royle, J.A. (2015) *Applied hierarchical modeling in ecology*, Academic Press.

Klinkovská, K., Glaser, M., Danihelka, J., Kaplan, Z., Knollová, I., Novotný, P., Pyšek, P., Řezníčková, M., Wild, J. & Chytrý, M. (2024) Dynamics of the Czech flora over the last 60 years: Winners, losers and causes of changes. *Biological Conservation*, **292**, 110502.

Kunin, W.E. (1998) Extrapolating species abundance across spatial scales. *Science*, **281**, 1513–1515.

Kunin, W.E. (1997) Sample shape, spatial scale and species counts: Implications for reserve design. *Biological Conservation*, **82**, 369–377.

Lahti, T. & Lampinen, R. (1999) From dot maps to bitmaps: Atlas Florae Europae goes digital. *Acta Botanica Fennica*, **162**, 5–9.

Lele, S.R., Moreno, M. & Bayne, E. (2012) Dealing with detection error in site occupancy surveys: what can we do with a single survey? *Journal of Plant Ecology*, **5**, 22–31.

Lennon, J.J., Koleff, P., GreenwooD, J.J.D. & Gaston, K.J. (2001) The geographical structure of British bird distributions: diversity, spatial turnover and scale. *Journal of Animal Ecology*, **70**, 966–979.

Leroy, F., Reif, J., Vermouzek, Z., Šťastný, K., Trávníčková, E., Bejček, V., Mikuláš, I. & Keil, P. (2024) Decomposing biodiversity change to processes of extinction, colonization, and recurrence across scales. *Ecography*, **2024**, e06995.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less then one. *Ecology*, **83**, 2248–2255.

McInerny, G.J. & Purves, D.W. (2011) Fine-scale environmental variation in species distribution modelling: regression dilution, latent variables and neighbourly advice. *Methods in Ecology and Evolution*, **2**, 248–257.

Mellado-Mansilla, D., Midolo, G., Ortega-Solís, G., Reif, J., Grattarola, F., Craven, D., Leroy, F., Perrone, M., Šťastný, K., Bejček, V. & Keil, P. (2026) Temporal span of biodiversity monitoring mediates the effects of area and environment.

Mertes, K. & Jetz, W. (2018) Disentangling scale dependencies in species environmental niches and distributions. *Ecography*, **41**, 1604–1615.

Meyer, C., Kreft, H., Guralnick, R. & Jetz, W. (2015) Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, **6**, 8221.

Midolo, G., Clark, A.T., Chytrý, M., Essl, F., Dullinger, S., Jandt, U., Bruelheide, H., Argagnon, O., Biurrun, I., Chiarucci, A., Ćušterevska, R., De Frenne, P., De Sanctis, M., Dengler, J., Divíšek, J., Dziuba, T., Ejrnæs, R., Garbolino, E., Illa, E., Jentsch, A., Jiménez-Alfaro, B., Lenoir, J., Moeslund, J.E., Napoleone, F., Pielech, R., Rumpf, S.B., Sanz-Zubizarreta, I., Silva, V., Svenning, J., Swacha, G., Večeřa, M., Vynokurov, D. & Keil, P. (2025) Six decades of losses and gains in alpha diversity of European plant communities. *Ecology Letters*, **28**, e70248.

Monsarrat, S., Boshoff, A.F. & Kerley, G.I.H. (2019) Accessibility maps as a tool to predict sampling bias in historical biodiversity occurrence records. *Ecography*, **42**, 125–136.

Montràs-Janer, T., Suggitt, A.J., Fox, R., Jönsson, M., Martay, B., Roy, D.B., Walker, K.J. & Auffret, A.G. (2024) Anthropogenic climate and land-use change drive short- and long-term biodiversity shifts across taxa. *Nature Ecology & Evolution*, **8**, 739–751.

Moudrý, V. & Šímová, P. (2012) Influence of positional accuracy, sample size and scale on modelling species distributions: a review. *International Journal of Geographical Information Science*, **26**, 2083–2095.

Nekola, J.C. & White, P.S. (1999) Special Paper: The Distance Decay of Similarity in Biogeography and Ecology. *Journal of Biogeography*, **26**, 867–878.

Orme, C.D.L., Davies, R.G., Olson, V.A., Thomas, G.H., Ding, T.-S., Rasmussen, P.C., Ridgely, R.S., Stattersfield, A.J., Bennett, P.M., Owens, I.P.F., Blackburn, T.M. & Gaston, K.J. (2006) Global Patterns of Geographic Range Size in Birds. *PLoS Biol*, **4**, e208.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T. & Abrego, N. (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, **20**, 561–576.

Pagel, J., Anderson, B.J., O'Hara, R.B., Cramer, W., Fox, R., Jeltsch, F., Roy, D.B., Thomas, C.D. & Schurr, F.M. (2014) Quantifying range-wide variation in population trends from local abundance surveys and widespread opportunistic occurrence records. *Methods in Ecology and Evolution*, **5**, 751–760.

Peach, M.A., Cohen, J.B. & Frair, J.L. (2017) Single-visit dynamic occupancy models: an approach to account for imperfect detection with Atlas data. *Journal of Applied Ecology*, **54**, 2033–2042.

Pescott, O.L. (2026) Unifying occupancy-detection and local frequency scaling (Frescalo) models. *Ecological Modelling*, **511**, 111367.

Pocock, M.J.O., Tweddle, J.C., Savage, J., Robinson, L.D. & Roy, H.E. (2017) The diversity and evolution of ecological and environmental citizen science. *PLOS ONE*, **12**, e0172579.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, **5**, 397–406.

Preston, C.D. (2013) Following the BSBI's lead: the influence of the *Atlas of the British flora* , 1962–2012. *New Journal of Botany*, **3**, 2–14.

Rahbek, C. (2005) The role of spatial scale and the perception of large-scale species-richness patterns. *Ecology Letters*, **8**, 224.

Rapacciuolo, G., Young, A. & Johnson, R. (2021) Deriving indicators of biodiversity change from unstructured community-contributed data. *Oikos*, **130**, 1225–1239.

Robertson, M.P., Cumming, G.S. & Erasmus, B.F.N. (2010) Getting the most out of atlas data. *Diversity and Distributions*, **16**, 363–375.

Rosindell, J., Hubbell, S.P. & Etienne, R.S. (2011) The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. *Trends in Ecology & Evolution*, **26**, 340–348.

Sadoti, G., Zuckerberg, B., Jarzyna, M.A. & Porter, W.F. (2013) Applying occupancy estimation and modelling to the analysis of atlas data. *Diversity and Distributions*, **19**, 804–814.

Schäfers, G., Ebersbach, H., Reimers, H., Körber, P., Janke, K., Borggräfe, K. & Landwehr, F. (2016) *Atlas der Säugetiere Hamburgs. Artenbestand, Verbreitung, Rote Liste, Gefährdung und Schutz*, Behörde für Umwelt und Energie, Amt f. Naturschutz, Grünplanung und Energie, Abteilung Naturschutz, Hamburg.

Soria, C.D., Ortega-Solís, G.R., Wölke, F.J.R., Barták, V., Leroy, F., Tschernosterová, K., Bejček, V., Herrando, S., Mikuláš, I., Šťastný, K., Ueta, M., Voříšek, P. & Keil, P. (2026) Spatial autocorrelation of species diversity and distributions in time and across spatial scales. *Global Ecology and Biogeography*, **35**, e70221.

Šťastný, K., Bejček, V., Mikuláš, I. & Telenský, Tomáš. (2021) *Atlas hnízdního rozšíření ptáků v České republice 2014-2017*, 1. vydání. Aventinum, Praha.

Stojmenovic, M. & Nayak, A. (2008) *Measuring the Related Properties of Linearity and Elongation of Point Sets*. *Progress in Pattern Recognition, Image Analysis and Applications* Lecture Notes in Computer Science. (ed. by J. Ruiz-Shulcloper) and W.G. Kropatsch), pp. 102–111. Springer Berlin Heidelberg, Berlin, Heidelberg.

Storch, D. (2016) The theory of the nested species–area relationship: geometric foundations of biodiversity scaling. *Journal of Vegetation Science*, **27**, 880–891.

Storch, D., Marquet, P.A. & Brown, J.H. (2007) *Scaling Biodiversity*, Cambridge University Press.

Stroh, P.A., Walker, K.J., Humphrey, T.A., Pescott, O.L. & Burkmar, R.J. (2023a) *Plant atlas 2020: mapping changes in the distribution of the British and Irish Flora*, Botanical Society of Britain and Ireland ; Princeton University Press, Durham; Princeton, New Jersey.

Stroh, P.A., Walker, K.J., Humphrey, T.A., Pescott, O.L. & Burkmar, R.J. (2023b) *Plant atlas 2020: mapping changes in the distribution of the British and Irish flora: volume 1-2*, Princeton University Press, Princeton, New Jersey Oxford.

Svenning, J.-C. & Skov, F. (2004) Limited filling of the potential range in European tree species: Limited range filling in European trees. *Ecology Letters*, **7**, 565–573.

Szabo, J.K., Vesk, P.A., Baxter, P.W.J. & Possingham, H.P. (2010) Regional avian species declines estimated from volunteer-collected long-term data using List Length Analysis. *Ecological Applications*, **20**, 2157–2169.

Telfer, M.G., Preston, C.D. & Rothery, P. (2002) A general method for measuring relative change in range size from biological atlas data. *Biological Conservation*, **107**, 99–109.

Van Swaay, C.A.M. & Warren, M.S. (1999) *Red data book of European butterflies*, Council of Europe Publishing, Strasbourg.

Warren, M.S., Hill, J.K., Thomas, J.A., Asher, J., Fox, R., Huntley, B., Roy, D.B., Telfer, M.G., Jeffcoate, S., Harding, P., Jeffcoate, G., Willis, S.G., Greatorex-Davies, J.N., Moss, D.

& Thomas, C.D. (2001) Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature*, **414**, 65–69.

Witte, J.M., He, F. & Groen, C.L.G. (2008) Grid origin affects scaling of species across spatial scales. *Global Ecology and Biogeography*, **17**, 448–456.

Xu, Q., Wang, X., Yi, J. & Wang, Y. (2024) Bias correction in species distribution models based on geographic and environmental characteristics. *Ecological Informatics*, **81**, 102604.

Ziolkowski, D., Lutmerding, M., Aponte, V. & Hudson, M.-A. (2022) 2022 Release - North American Breeding Bird Survey Dataset (1966-2021).

植田睦之, 植村慎吾, 葉山政治, 荒哲平, 高川晋一, 川路則友, 梶田学, 仲村昇, 田畑早紀 & 内山優奈 (2021) 全国鳥類繁殖分布調査で収集された20km メッシュの鳥類の繁殖分布のデータ.

8 Declaration of generative AI and AI-assisted technologies

During an early stage of this work in 2023, we used the AI tool Elicit to scan the literature on the topic and extract information about taxa, temporal replication, data type, geographic extent, spatial scale, and metrics from a selection of > 200 studies that were compiled by a manual search in Scopus and Web of Science. This information was then used to focus our reading efforts: after reviewing the output, we read all relevant literature individually. However, no generative AI was used for writing the manuscript, we wrote it manually.