Received: 18 November 2021 | Revised: 18 November 2021 | Accepted: 23 December 2021

DOI: 10.1111/gcb.16114

Global Change Biology WILEY

OPINION

Sampling and modelling rare species: Conceptual guidelines for the neglected majority

Alienor Jeliazkov[1] | **Yoni Gavish**[2] | **Charles J. Marsh**[3,4] | **Jonas Geschke**[5] | **Neil Brummitt**[6] | **Duccio Rocchini**[7,8] | **Peter Haase**[9,10] | **William E. Kunin**[11] | **Klaus Henle**[12]

[1]University of Paris-Saclay, INRAE, UR HYCAR, Antony, France
[2]School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK
[3]Department of Plant Sciences, University of Oxford, Oxford, UK
[4]Department of Ecology and Evolution & Yale Center for Biodiversity and Global Change, Yale University, New Haven, Connecticut, USA
[5]Institute of Plant Sciences, University of Bern, Bern, Switzerland
[6]Department of Life Sciences, Natural History Museum, London, UK
[7]BIOME Lab, Department of Biological, Geological and Environmental Sciences, Alma Mater Studiorum University of Bologna, Bologna, Italy
[8]Department of Spatial Sciences, Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Praha - Suchdol, Czech Republic
[9]Department of River Ecology and Conservation, Senckenberg Research Institute and Natural History Museum Frankfurt, Gelnhausen, Germany
[10]Faculty of Biology, University of Duisburg-Essen, Essen, Germany
[11]University of Leeds, Leeds, UK
[12]Department of Conservation Biology & Social-Ecological Systems, UFZ – Helmholtz Centre for Environmental Research, Leipzig, Germany

Correspondence
Alienor Jeliazkov, University of Paris-Saclay, INRAE, UR HYCAR, Antony, France.
Email: alienor.jeliazkov@gmail.com

Funding information
Seventh Framework Programme, Grant/Award Number: 308454

Abstract

Biodiversity conservation faces a methodological conundrum: Biodiversity measurement often relies on species, most of which are rare at various scales, especially prone to extinction under global change, but also the most challenging to sample and model. Predicting the distribution change of rare species using conventional species distribution models is challenging because rare species are hardly captured by most survey systems. When enough data are available, predictions are usually spatially biased towards locations where the species is most likely to occur, violating the assumptions of many modelling frameworks. Workflows to predict and eventually map rare species distributions imply important trade-offs between data quantity, quality, representativeness and model complexity that need to be considered prior to survey and analysis. Our opinion is that study designs need to carefully integrate the different steps, from species sampling to modelling, in accordance with the different types of rarity and available data in order to improve our capacity for sound assessment and prediction of rare species distribution. In this article, we summarize and comment on how different categories of species rarity lead to different types of occurrence and distribution data depending on choices made during the survey process, namely the spatial distribution of samples (where to sample) and the sampling protocol in each selected

---

Alienor Jeliazkov and Yoni Gavish should be considered as joint first author.

Klaus Henle and William E. Kunin should be considered as joint senior author.

location (how to sample). We then clarify which species distribution models are suitable depending on the different types of distribution data (how to model). Among others, for most rarity forms, we highlight the insights from systematic species-targeted sampling coupled with hierarchical models that allow correcting for overdispersion and spatial and sampling sources of bias. Our article provides scientists and practitioners with a much-needed guide through the ever-increasing diversity of methodological developments to improve the prediction of rare species distribution depending on rarity type and available data.

KEYWORDS
bias, detectability, distribution change, methods, occupancy, rare species, sampling, spatial data, species distribution modelling, survey

1 | INTRODUCTION

Almost all international, national and local conservation planning activities flag biodiversity as a crucial environmental property (e.g. Aichi Targets, Sustainable Development Goals) (Butchart et al., 2016; Griggs et al., 2013) to be protected from the deleterious effects of habitat loss, exploitation, pollution and climate change (IPBES, 2019; Maxwell et al., 2016; Rands et al., 2010). However, biodiversity measurement often relies on species, most of which are rare at various scales (Enquist et al., 2019; Fontaine et al., 2007; Hartley & Kunin, 2003; Henle et al., 2010; Rabinowitz, 1981; Steege et al., 2013). Several initiatives to halt biodiversity loss have questioned whether current measures of biodiversity do actually sufficiently account for rare species (e.g. Fontaine et al., 2007). For example, one-third of plant species worldwide are too poorly known and have too few data for a Red List assessment (Brummitt et al., 2015; Enquist et al., 2019). At the same time, rare species are especially prone to extinction (Courchamp et al., 2006; Henle et al., 2004; Işik, 2011; Kunin & Gaston, 1993; McKinney, 1997). One way to assess extinction risk is to track the change in spatial distribution through time (Araújo et al., 2002; Benito et al., 2009; Gärdenfors et al., 2001; Thomas et al., 2004). Therefore, protecting species diversity directly implies protecting rare species, which requires understanding their distribution patterns.

Unfortunately, rarity causes considerable methodological difficulties in obtaining sufficient data from survey programmes or alternative sources (e.g. Roberts et al., 2016), which limits the ability of models to predict distribution patterns. For example, many studies using species distribution models (SDMs, defined in Box 1) need a minimum number of occurrences below which the models cannot be reliably trained and/or validated (e.g. van Proosdij et al., 2016). Thus, we are locked in the "rare-species modelling paradox" (Lomba et al., 2010): the majority of species that require the greatest protection also are the species we know least about and are most difficult to model.

However, rarity is an umbrella term used to describe various types of distribution patterns at various scales. Rabinowitz (1981) defined seven categories of rarity based on combinations of the range of a species, the distribution of populations within its range and the local density of the species when present (Figure 1a). Whatever measure used (e.g. range size, occupancy, abundance, relative cover, biomass), and ecosystem or scale of the study, a community is likely to include a handful of common species and a long tail of rare species (Fisher et al., 1943; Preston, 1948). The resulting pattern of species-abundance distributions, following a log-like curve in most natural systems (but also see (Magurran & Henderson, 2003)), is observed on local to global scales, with correspondingly fine abundance (McGill et al., 2007) to range size frequency (Gaston, 1998) data.

With the goal of mapping rare species' distribution ranges and changes for protection purposes, each of the seven types of rarity implies different problems in accumulating data for modelling. For example, two species A and B with similar prevalence are both found within an area: Species A has a narrow range with high local density (rarity category 2) and species B has a broad range with low local density (rarity category 4). Randomly distributed sampling in this area is likely to sample only a few sites where species A is present and many sites where species B is present; consequently, species B's distribution is likely to be better evaluated than species A's distribution. However, *a priori* knowledge on where species A is present may mean that species A is more often encountered than species B; consequently, the dataset contains more presences of species A than of species B. The type of rarity, the spatial distribution of samples and the protocol used to sample each location thus all affect the data generated, and the types of the model used to project the species' distribution range. Finally, with the perpetual changes in taxonomy (taxonomic revisions), the identification, assessment and conservation of rare species are constantly challenged (Ota, 2000; Schwartz & Simberloff, 2001; Standley, 1992) (but see also (Domínguez Lozano et al., 2007; Simkins et al., 2020)) and the expected increase in species number for some taxonomic groups (Morrison III et al., 2009) foresees an endless need to coping with rarity issues.

We, therefore, face a conundrum in which, although rarity is ubiquitous, it is particularly challenging to account for, sample and model, at all scales. Whilst some publications already provide a comprehensive overview on specific aspects of the different steps from

BOX 1 Glossary (of the terms underlined in the main text)

- Hierarchical Models (HM): or *multi-level models*. Statistical models of parameters that vary at more than one level of data organization (e.g. nested data, such as abundances of a given species located in different habitat types themselves located in different ecoregions) and thus allow accounting for the potential interdependence between the data points (for further details, see, e.g. (Gelman & Hill, 2007; Raudenbush & Bryk, 2002)).
- Mark-release-recapture (MRR): Mark-release-recapture, or *capture-mark-recapture*, is a sampling technique that consists in capturing, marking and releasing individuals of a species in a first capture session. In one or more follow-up capture sessions, the ratio of marked to unmarked specimens is taken to estimate population size (see, e.g. (Southwood & Henderson, 2009; Williams et al., 2002)).
- Occupancy: Occupancy can refer to two different notions (MacKenzie et al., 2017); (1) the probability of a site to be occupied by a given species, i.e. the *a priori* expectation that a particular site will be occupied by the species as determined by some underlying process (or occurrence probability), (2) the proportion of area or sites occupied, which results from the realization of the former process.
- Patchiness: The way habitat patches (and populations) are distributed through space. Habitat patches can be clumped (i.e. spatially aggregated in patches concentrated in a few places, potentially most at risk under environmental stochasticity), patchy (i.e. spatially aggregated according to irregular patterns, e.g. one, two, or five patches per group of patches), random and regular (i.e. uniformly distributed apart from each other).
- Spatially representative sample-set: Sample-set collected at a set of locations that are spatially distributed in a statistically unconstrained manner, e.g. by a stratified design, in which areas are stratified according to their environmental conditions and the number of samples in each stratum is proportional to the area of that stratum. Such sampling is spatially representative of the variability of these conditions over the whole study area and does not over-represent unusual but rare environmental conditions.
- Species Distribution Model (SDM): Here, used as a generic catch-all term to refer to any empirical model that allows spatially-explicit prediction of the current or future environmental suitability for a species (using presence-only, presence/absence and/or abundance data) based on predictors (such as climate, land-use, etc.) and, possibly, scenarios (e.g. IPCC's climate change scenarios) (Guisan & Thuiller, 2005). Depending on the objectives and underlying assumptions—but mostly using the same types of data and algorithms, these models are also called ecological niche models (ENMs), habitat suitability models (HSMs), niche-based models (NBMs), potential habitat distribution models (PHDMs) and when used only with climate variables, climate-envelope models (CEMs) or climate matching models (CMMs) (Guisan et al., 2013). For instance, ENM can give more focus to species niche quantification or requirements whilst strict SDMs focus more on getting spatial predictions of species distribution (Saupe et al., 2012). In the context of rare species modelling, models aim to predict either the probability/likelihood of occurrence or the probability of environmental suitability for the species, with the caution that these predictions may differ from the realized distribution because a location may be suitable but not reachable by the species.

sampling to modelling rare species (Cunningham & Lindenmayer, 2005; Green & Young, 1993; Hermoso et al., 2015; Kenkel et al., 1990; Milner-Gulland & Rowcliffe, 2007; Robinson et al., 2018; Thompson, 2013b), how to improve our prediction of rare species distribution changes remains a complete challenge to date (Aubry et al., 2017; Didham et al., 2020; Galante et al., 2018; Helmstetter et al., 2021). Our perspective is that all steps need to be integrated in the study design. In particular, their sequence needs to be adapted to the different types of rarity to improve our capacity for sound assessments and predictions of the distribution of the majority of biodiversity. However, the trade-offs faced when modelling the distribution of rare species and the decision path linking the form of the rarity with the sampling and modelling strategies have largely been neglected. Therefore, to help untangle the rarity conundrum and adapt modelling strategies to the rarity issues, we aim, for each of Rabinowitz's categories of rarity:

(i) To identify the main trade-offs involved in selecting adequate, cost-effective sampling strategies and how these affect the properties of the data.
(ii) To identify modelling frameworks that are potentially suitable for the type of data generated and to highlight gaps that require model development.

To address the first aim, we focus on the spatial distribution of samples ('where to sample') and on the protocols used to do the sampling ('how to sample'). For the second aim, we list and discuss the main modelling frameworks suitable for producing distribution maps for different types of rarity ('how to model'). We synthesize our findings, provide guidelines to optimize and integrate monitoring and modelling of rare species depending on their rarity characteristics and briefly discuss remaining challenges with respect to sampling and modelling rare species.

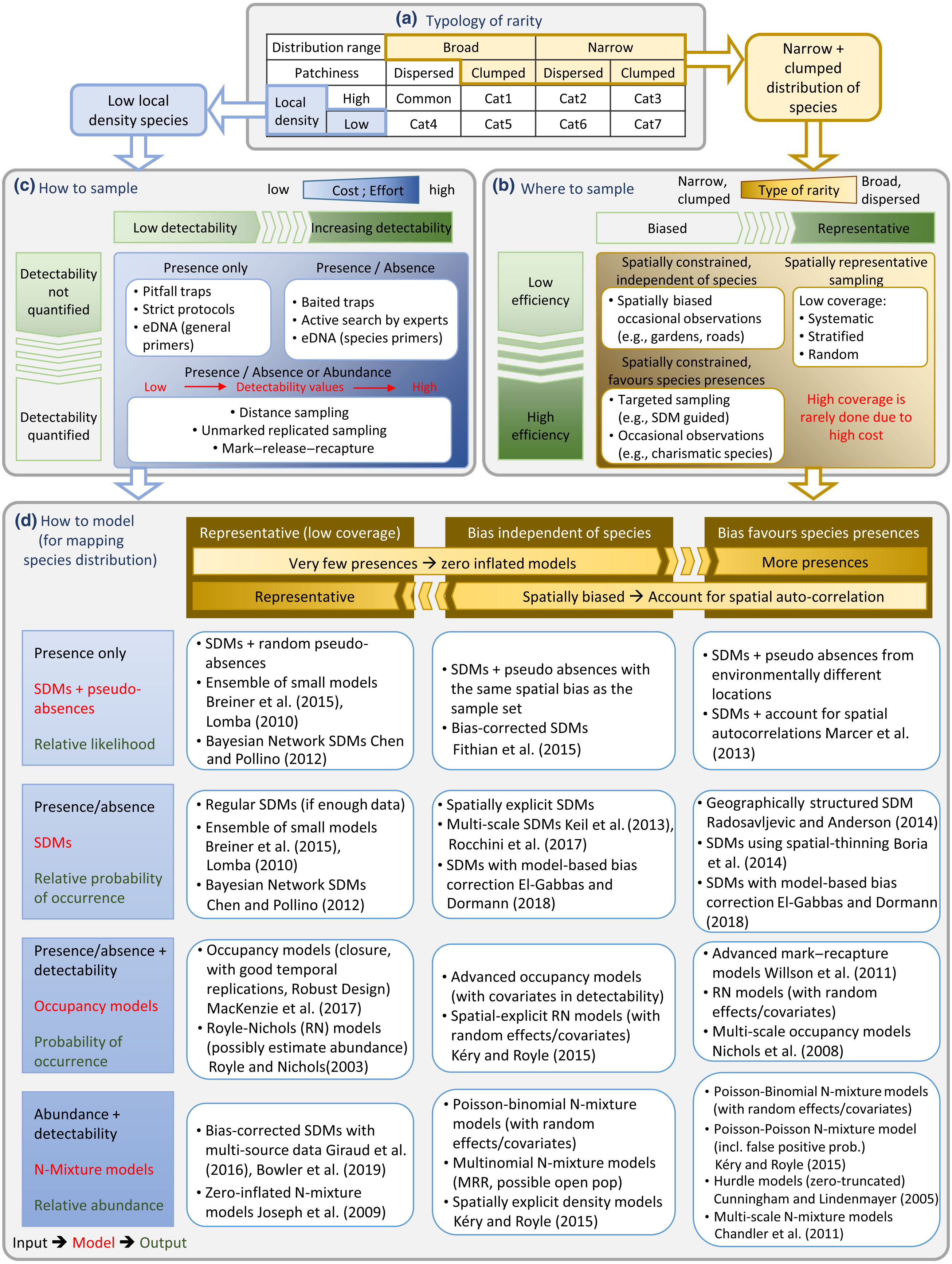
(a) Typology of rarity
Distribution range
Broad
Narrow
Patchiness
Dispersed
Clumped
Dispersed
Clumped
Local density
High
Common
Cat1
Cat2
Cat3
Low
Cat4
Cat5
Cat6
Cat7
Low local density species
Narrow + clumped distribution of species
(c) How to sample
low
Cost ; Effort
high
Low detectability
Increasing detectability
Detectability not quantified
Detectability quantified
Presence only
• Pitfall traps
• Strict protocols
• eDNA (general primers)
Presence / Absence
• Baited traps
• Active search by experts
• eDNA (species primers)
Presence / Absence or Abundance
Low
Detectability values
High
• Distance sampling
• Unmarked replicated sampling
• Mark–release–recapture
(b) Where to sample
Narrow, clumped
Type of rarity
Broad, dispersed
Biased
Representative
Low efficiency
High efficiency
Spatially constrained, independent of species
• Spatially biased occasional observations (e.g., gardens, roads)
Spatially constrained, favours species presences
• Targeted sampling (e.g., SDM guided)
• Occasional observations (e.g., charismatic species)
Spatially representative sampling
Low coverage:
• Systematic
• Stratified
• Random
High coverage is rarely done due to high cost
(d) How to model (for mapping species distribution)
Representative (low coverage)
Bias independent of species
Bias favours species presences
Very few presences → zero inflated models
More presences
Representative
Spatially biased → Account for spatial auto-correlation
Presence only
SDMs + pseudo-absences
Relative likelihood
• SDMs + random pseudo-absences
• Ensemble of small models Breiner et al. (2015), Lomba (2010)
• Bayesian Network SDMs Chen and Pollino (2012)
• SDMs + pseudo absences with the same spatial bias as the sample set
• Bias-corrected SDMs Fithian et al. (2015)
• SDMs + pseudo absences from environmentally different locations
• SDMs + account for spatial autocorrelations Marcer et al. (2013)
Presence/absence
SDMs
Relative probability of occurrence
• Regular SDMs (if enough data)
• Ensemble of small models Breiner et al. (2015), Lomba (2010)
• Bayesian Network SDMs Chen and Pollino (2012)
• Spatially explicit SDMs
• Multi-scale SDMs Keil et al. (2013), Rocchini et al. (2017)
• SDMs with model-based bias correction El-Gabbas and Dormann (2018)
• Geographically structured SDM Radosavljevic and Anderson (2014)
• SDMs using spatial-thinning Boria et al. (2014)
• SDMs with model-based bias correction El-Gabbas and Dormann (2018)
Presence/absence + detectability
Occupancy models
Probability of occurrence
• Occupancy models (closure, with good temporal replications, Robust Design) MacKenzie et al. (2017)
• Royle-Nichols (RN) models (possibly estimate abundance) Royle and Nichols(2003)
• Advanced occupancy models (with covariates in detectability)
• Spatial-explicit RN models (with random effects/covariates) Kéry and Royle (2015)
• Advanced mark–recapture models Willson et al. (2011)
• RN models (with random effects/covariates)
• Multi-scale occupancy models Nichols et al. (2008)
Abundance + detectability
N-Mixture models
Relative abundance
• Bias-corrected SDMs with multi-source data Giraud et al. (2016), Bowler et al. (2019)
• Zero-inflated N-mixture models Joseph et al. (2009)
• Poisson-binomial N-mixture models (with random effects/covariates)
• Multinomial N-mixture models (MRR, possible open pop)
• Spatially explicit density models Kéry and Royle (2015)
• Poisson-Binomial N-mixture models (with random effects/covariates)
• Poisson-Poisson N-mixture model (incl. false positive prob.) Kéry and Royle (2015)
• Hurdle models (zero-truncated) Cunningham and Lindenmayer (2005)
• Multi-scale N-mixture models Chandler et al. (2011)
Input → Model → Output

FIGURE 1 Synthesis infographic of (a) the Rabinowitz's seven categories of rarity, (b) examples of approaches to assess where to sample depending on the rarity category, (c) examples of approaches to assess how to sample depending on the rarity category and species local density and (d) examples of modelling approaches to predict and map species distribution depending on the type of data generated in previous steps (a) and (b). Note that most of the methods can be used in more than one situation, but for the simplicity of the figure, we did not systematically repeat them and rather highlighted the methods we considered as the most useful or relevant. The references (numbers in brackets) are listed below the figure. Breiner et al. (2015), Lomba (2010), Chen and Pollino (2012), Fithian et al. (2015), Marcer et al. (2013), Keil et al. (2013), Rocchini et al. (2017), El-Gabbas and Dormann (2018), Radosavljevic and Anderson (2014), Boria et al. (2014), MacKenzie et al. (2017), Royle and Nichols (2003), Kéry and Royle (2015), Willson et al. (2011), Nichols et al. (2008), Giraud et al. (2016), Bowler et al. (2019), Joseph et al. (2009), Cunningham and Lindenmayer (2005), Chandler et al. (2011)

2 | WHERE TO SAMPLE

When setting up a survey program there are multiple ways by which the spatial allocation of samples can be decided (Table 1; Figure 1b). Any choice made at this stage will affect the properties of the collected data. The main trade-off to consider is between sampling efficiency and spatial coverage.

Locally focused sampling targeting a particular species allows its population to be studied efficiently, but at the expense of a spatially non-representative sample of the species distribution. This conflicts with the aim of covering the realised niche of a species, an assumption of most modelling frameworks. For species whose distribution range is relatively wide and distribution pattern is dispersed (common species and rarity category 4), a spatially representative sample-set of the entire extent is more likely to provide the required occurrence data. Spatially representative sampling has several positive properties. First, data are comparable among species, allowing cost-effective monitoring of multiple species. Second, even if the location of samples is not constant, data remain comparable between years, allowing the detection of temporal changes in distribution (if sampling intensity is kept constant). Third, data on the focal species fit easily into most modelling frameworks, if enough are collected. This is usually done with a systematic sampling scheme on a grid, stratifying the sampling according to habitat or land cover (whilst ensuring proportional sampling in each stratum), or by randomly selecting the sampling locations (Table 1; Figure 1b).

However, for species with narrow and/or clumped and patchy distribution patterns (rarity categories 1, 2, 3, 5, 6, 7), a random sample-set of the entire extent is unlikely to capture sufficient information. For example, in the 2007 UK plants countryside survey, 591 one-km^2 locations were included in a stratified random design (Bunce et al., 2014; Carey et al., 2008): the survey recorded 880 species. As there are approximately 4000 plant species in the UK, the survey failed to detect 2400 rare species. In fact, the narrower and clumpier the distribution of a species, the larger the number of random sites needed to encounter the species in enough locations to make credible estimates of abundance or distributional status and changes. Thus, one may need to constrain the sampling towards the target species.

Various methods allow the distribution of samples to target locations more likely to contain a certain rare species (Table 1; Figure 1b). One such example is adaptive sampling (Thompson, 2013b; Yoccoz et al., 2001). Many programs periodically monitor rare species sample locations where the species is known to occur, but rarely look in new sites. Such adaptive sampling may be excellent in keeping track of known populations, but eventually leads to erroneous conclusions regarding distribution trends. Consider a species subject to metapopulation dynamics, experiencing local extinctions and colonization of patches: if sampling is in known locations only, one may identify all local extinctions (and a preceding gradual decrease in population size) but not identify the colonization of new patches. Thus, we might wrongly conclude that the species distribution is deteriorating whilst it may in fact be in an equilibrium state (Magurran et al., 2010) (but see (McRae et al., 2017)).

Another fruitful approach is to combine adaptive with SDM-guided sampling (Aizpurua et al., 2015; Chiffard et al., 2020; e.g. Lin et al., 2014) where one sampling session provides information to model and the following sessions allow adjusting the distribution of samples (Thompson, 2013a,b; Yoccoz et al., 2001). For example, a SDM with data sampled at a certain time can tag potentially unknown local populations for sampling the next year (e.g. Lin et al., 2014). Once the area is sampled and SDM parameters updated, the SDM is re-run and new locations targeted. Such a strategy may be very efficient at accumulating observations of rare species. However, it comes with the risk of estimating an over-optimistic occupancy trend, as the number of detected presences can increase over time whilst the distribution actually decreases (Table 1). Appropriately parametrized stacked SDMs, including rarity weighting, can further allow improving the sampling of multiple rare species and help prioritize sampling areas (Rosner-Katz et al., 2020). Any form of adaptive sampling, therefore, needs considerable manipulation and/or reliable complementary information for further species distribution modelling (Dorazio, 2014; Hefley et al., 2014; Phillips et al., 2009; Raes & ter Steege, 2007).

The transition from spatially representative sampling to species-targeted sampling also reflects a gradient of *a priori* knowledge (Table 1). Random sampling does not require specific knowledge. Adaptive sampling and SDM-guided approaches instead need considerable knowledge of the species and its requirements before designing the sampling scheme. Stratified schemes require knowledge about sampling sites and their habitats or environmental conditions across the full range of the target species. Additionally, stratified schemes depend on the quality of the original information used to guide the stratification that has its own uncertainty, due to potential spatial errors and classification issues (Rocchini et al., 2011).

To summarize, different strategies for defining the spatial distribution of samples reflect the compromise between sampling

TABLE 1 Non-exhaustive list of methods to assess where to sample rare species data, with their brief description, advantages and limits, the type of rarity for which they appear as most appropriate, and examples of references related (listed in the references section). Underlined words refer to the Glossary (Box 1)

Method	Brief description	Pros	Cons	Suitable for which rarity categories?	References
Accumulated opportunistic observations	Sampling locations are not chosen but emerge from the external contribution of various sources, e.g. data from citizen science programs free from any observation protocol	Depending on the species attractivity and ease of detection/identification, a large number of observations can be accumulated over time, with minimal investment of time and funds Can detect new populations and species May be used to create atlas data Rare species receive particular attention	Sample not representative of the entire extent Species-targeted Absences usually not reported, presence-only data Sampling effort varies through time Mainly done for charismatic taxa The risk of misidentification in the case of non-expert observations (particularly critical as even a small fraction of misidentified common species may swamp the true records of a rare species)	All	Chandler et al. (2017) (iNaturalist); Sullivan et al. (2017) (eBird); Deguines et al. (2012) (spipoll)
Simple random sampling	Random selection of the locations, i.e. all the locations of the study area have the same probability to be sampled	Spatially unbiased sample Objective and well-defined Sample representative of the study extent Temporally comparable samples No target species, multi-species sample	Ignores environmental/habitat variability Rare species are unlikely to be detected in sufficient numbers, even in huge samples	Cat4	Greig-Smith (1964); Diekmann et al. (2007); Hedgren and Weslien (2008)
Systematic sampling	Sampling according to a fixed spatial interval(s) that depends on the predefined total number of locations to be sampled in the study area, e.g. plots arranged along a regular grid or (equidistant) transects that cover the space evenly (hyper dispersed distribution of samples)	Simple to implement, no need for external information nor a priori species-specific knowledge More cost-efficient than simple random sampling as it guarantees even distribution of sites and good coverage of the study area Temporally comparable samples No target species, multi-species sample	Needs prior information on the total number of sites to be sampled Detection strongly depends on the choice of the spatial interval of the sampling and on the starting point of the sampling, e.g. in species with clumped populations; if the sampling interval is the same order of magnitude as the clumping interval, the sample will not be representative of the species distribution (will either under- or over-detect the species depending on the starting point)	Cat4 (and Cat5 if habitats are organised randomly)	Madow (1953); Fortin et al. (1989)
Stratified sampling	Sampling organised with respect to a categorisation deemed to be important for the community or species of interest, e.g. habitat type	Sample representative of the study extent with respect to the stratification factor	Depends on subjective a priori, or a priori ecological knowledge	Cat1, Cat3, Cat5, Cat7 (if we consider that for non-specialist species, habitat-stratified sampling would work worse)	Thompson (2013b)

TABLE 1 (Continued)

Method	Brief description	Pros	Cons	Suitable for which rarity categories?	References
Adaptive (cluster) sampling/ prior-informed sampling	Sampling design where site selection depends on previous sampling raw outcomes, either (a) from the overall survey, e.g. adaptive *cluster* sampling which consists in searching for a species in a given location and if the species is found, searches continue nearby (neighbourhood shape can vary according to the study needs), or (b) from other surveys, i.e. site selection depends on external source of information and/or belief on the species potential presence, e.g. atlas data	Accurate estimations of species abundances Appropriate for rare, clustered and unevenly distributed species	Not widely used in ecological studies Efficiency depends on the spatial distribution of the species Difficult to know the final sample size needed prior to the survey Data collection process is complicated Not fully adapted yet to mobile species, sensitive species and habitats (side-effects of intensive sampling) Resulting data biased towards the species of interest Sampling effort varies through time	Cat2, Cat3	Krebs (1989); Yoccoz et al. (2001); Thompson (1990, 2013a); Thompson (2002)
"SDM"-guided sampling	Sampling locations are drawn from a probability surface generated by modelling the know P/A of a species against environmental predictors and extrapolating the model in space and time, e.g. adaptive niche-based sampling, Direct Gradient Analyses	Sampling coverage optimisation Allows a systematic and exhaustive pre-selection of suitable locations	Time-consuming process Requires predictor layers (with a good spatial and thematic resolution for narrow range species) Subject to model error and uncertainty May work better for specialist species that are not too much dispersal limited (niche-based modelling)	Cat1, Cat3 (potentially Cat5, Cat 7 if clumping is not due to dispersal limitations)	Austin et al. (1984); Le Lay et al. (2010); Lin et al. (2014); Aizpurua et al. (2015); Chiffard et al. (2020)

efficiency and spatial representativeness (Figure 1b). Overall, three main types of data may be generated, each with implications for modelling: data can be spatially representative (of the species range, potentially for multiple species), spatially constrained independent of the species, or spatially constrained towards particular species.

3 | HOW TO SAMPLE

For assessing the distribution of species and changes therein, sampling should aim to collect the appropriate quantity of presence data, reduce the number of false absences, and account for the detectability of the sampled species (Table 2; Figure 1c). Locally rare as well as elusive (e.g. cryptic or trap-shy) species (Thompson, 2013b) both pose specific challenges. The probability of detecting a species depends on a range of factors, such as habitat type, time of the day and year, population density and methods employed to survey the species. Repeated sampling with methods targeting rare and elusive species reduce the probability of false absences and the latter may generate presence/absence data accounting for detection probability (MacKenzie et al., 2017).

Multiple methods increase the detectability of species; some are just a function of sampling effort (e.g. longer transects), others are more directly related to the known ecology of the target species (Table 2; Figure 1c). These latter methods include, for example, baited traps (e.g. Steyer et al., 2013), camera traps (e.g. Schüttler et al., 2017), species-specific markers in environmental DNA (eDNA) sampling (e.g. Carraro et al., 2018), expert knowledge of the species' habitat preference and/or behaviour or the use of detection dogs (Grimm-Seyfarth & Klenke, 2019; Grimm-Seyfarth et al., 2019; Hollerbach et al., 2018).

There are several points to consider. First, most of these methods increase the effort or costs required compared with simpler methods, especially when the sampling aims to detect several rare species simultaneously. Second, methods increase detection probability differently for different species, producing output less comparable between species unless methods are highly standardized. For example, a trap baited with pheromones of a specific species will attract more individuals of the focal species than baiting a trap with food utilized by many species (e.g. dung for dung beetles) (Marsh et al., 2013). However, recent advances in genetic monitoring, such as improved markers in eDNA detection of stream species (Carraro et al., 2021; e.g. Jerde et al., 2011; Leese et al., 2021), significantly increase the number of species detected, including many rare species, especially from rivers over several kilometres in length (Altermatt et al., 2020; e.g. Mächler et al., 2019)—but these methods still need further calibration works (Alsos et al., 2018; e.g. Beng & Corlett, 2020; Cristescu & Hebert, 2018). Third, highly standardized protocols are essential for comparisons among sites, although some variability in detectability between sites will remain; for example, bird songs are less audible in leaved deciduous forests than in mixed pine forests (Pacifici et al., 2008).

Some sampling methods generate presence/absence and even abundance data in sufficient quality and quantity to account for detection probability (with repeated sampling of selected sites during a specific period (Mackenzie & Royle, 2005)). Among others, such methods include distance sampling (Buckland et al., 2015) and capture-mark-recapture (Williams et al., 2002). For the latter, capture by camera traps coupled with image analysis is particularly promising for rare species (Schüttler et al., 2017; Table 2; Figure 1c). Although these data greatly increase the spectrum of models that can be applied, they require high effort and cost; hardly suitable for rare species except perhaps for those with high local density. However, combining such methods with occupancy surveys or opportunistic observations (e.g. atlas or citizen-science data) and the incorporation of environmental data as potential predictors of occupancy and/or abundance may allow the extrapolation of rare species distributions across large spatial scales (e.g. Bowler et al., 2019; Giraud et al., 2016).

4 | HOW TO MODEL

As discussed above, choices on the spatial distribution of samples eventually lead to three types of datasets: spatially representative, spatially constrained independent of the species (e.g. due to unrepresentative sampling of environments (see e.g. Bystriakova et al., 2012; Varela et al., 2014)), or spatially constrained towards target species. From a modelling perspective, this results in a trade-off between the number of presences and the need to account for spatial auto-correlation in the data. Similarly, sampling protocols affect the type of data obtained for modelling, be it presence-only, presence/absence, or presence/absence with detectability or estimates of abundances, and thus condition the type and quality of inference. Depending on the type of rarity, the 'where to sample' and 'how to sample' decisions, successful modelling of rare species require modelling tools that fall into all combinations of the cases above (Figure 1d).

When only presences are available, some methods produce pseudo-absences based on external information (e.g. habitat suitability (Barbet-Massin et al., 2012)). For some models, such as Maxent and Poisson point-process models (PPPMs), pseudo-absences are better interpreted as background points, not implying absences but rather samples of the available environment, where presences are compared against unsampled background locations (Merow et al., 2013; Phillips et al., 2009). They do not produce the probability of occurrence but relative occurrence rates (Guillera-Arroita et al., 2015) and can be appropriate for rare species modelling if proper bias correction is applied (Table 3; Figure 1d).

Where presence/absence data are available, developments in SDMs allow handling of data over-dispersion (e.g. negative-binomial and mixed effect models (Harrison, 2014; Molenberghs et al., 2007; O'Hara & Kotze, 2014)), spatial-autocorrelation (e.g. Dormann et al., 2007; Marcer et al., 2013), uncertainty in predictions (e.g. ensemble forecasting (Araújo & New, 2007; Guisan

TABLE 2 Non-exhaustive list of methods to assess how to sample rare species data, with their brief description, advantages and limits, the type of rarity for which they appear as most appropriate, and examples of references related (listed in the references section). Underlined words refer to the Glossary (Box 1)

Method	Brief description	Pros	Cons	Suitable for which rarity categories?	References
Standardized sampling	Sampling with commonly use methods following a standardized protocol (e.g. quadrats, transects, traps, etc.) without any adaptation to increase the probability of detecting rare species, e.g. biodiversity observatories	Detection of a large number of species Data comparable across locations Unbiased with respect to sampling effort	Rare species less likely to be detected when populations have a low local density	Cat1, Cat2, Cat3	Enquist et al. (2016); Bruelheide et al. (2019); Risely et al. (2010); Jiguet et al. (2012)
Occupancy sampling	Sampling that consists of repeated sampling following a standardized protocol within a period during which the targeted species remain available for detection	Multi-species; allows estimating detection probability that can be used to obtain unbiased presence/absence data	The effort required is high unless detection probability is high May require survey methods targeted to particular rare species, such as lures	All	MacKenzie and Royle (2005); MacKenzie et al. (2017)
Distance sampling	Sampling that consists in recording the distance from the observer to the organism when detected. This information can then be used to adjust the sampling strategy and to correct for detection probability in prediction models	Multi-species	Requires expert knowledge (able to identify species at different distances within a given radius) Locally rare species will not provide sufficient observations for reliable estimates of abundance	Common species, Cat2	Rosenstock et al. (2002); Buckland et al. (2015)
Species-targeted sampling (or species-specific sampling)	Sampling specifically designed for given locally rare species, based on fine information on the species' habits, to increase the encounter rate, e.g. traps with specific food items or pheromone baits	Highly efficient in detecting rare species of interest Fine resolution data	Intensive field work Cannot cover large spatial extent (but see promising methods such as detection dogs) Species-targeted	All	Grimm and Klenke (2019); Grimm et al. (2019)
Mark-Release-Recapture sampling	Sampling that consists in capturing, marking and releasing individuals of given species in order to keep track of their identity and be able to estimate capture rate and population parameters	Under particular assumptions, allows estimating population parameters, such as population size, fecundity, etc. Fine resolution data	Highly time-consuming and field-work intensive Cannot cover large spatial extents Species-targeted	Cat1, Cat2, Cat3	Williams et al. (2002)
Passive sampling	Sampling based on the setting up of devices that automatically record species passing within a certain radius, e.g. camera trapping, acoustic sampling	Allows large-scale surveys Multi-species	Non-specific, detects any species as well as noise Costly in terms of resources (to buy devices, process data, etc.)	Cat3, Cat7 (+ Cat2, Cat6 if devices can be set anywhere)	Fleschutz et al. (2016) (camera trapping) Schüttler et al. (2017) (radio-tracking) Jeliazkov et al. (2016) (acoustic sampling)
eDNA	Sampling based on DNA extraction from the environment (e.g. water, soil, sediments, snow) coming from cells of organisms that are and/or were present at some point in the environment. Specific or unspecific primers can be used to amplify eDNA samples, depending on whether the survey targets specific species or the whole community, respectively	Rapid survey at large scales, cost-effective Species-targeted as well as multi-species assessments High detection power Non-invasive method No licence constraints for protected species In some cases, can provide semi-quantitative estimation of abundances	Detectability depends on several parameters whose effects can be confounded with actual ecological responses, e.g. environmental conditions, such as UV light, temperature, and water flow, but also the activity and density of animals, their residence time, etc The importance of primer specificity	Cat1, Cat2, Cat3 (+Cat5, Cat7 if we consider that at low population density, habitat specificity may ensure higher eDNA concentrations than habitat unspecificity)	Bohmann et al. (2014); Rees et al. (2014); Jerde et al. (2011); Wilcox et al. (2013); Beng and Corlett (2020)

TABLE 3 Non-exhaustive list of methods to assess how to model rare species data, with their brief description, advantages and limits, the type of rarity for which they appear as most appropriate, and examples of references related (listed in the references section). Inputs/outputs can be P (presences only), lik (presence likelihood), PA (Presences/Absences), ab (abundance), det (detectability information), pocc (probability of occurrence). Underlined words refer to the Glossary (Box 1)

Method		Brief description	Pros
Data processing	Data processing	Different processing strategies can be applied to data prior to actual modelling which allows making data more appropriate, more powerful, or more in line with the assumptions of subsequent modelling; e.g. combine opportunistic observations with atlas data, correct biases in presence-only data, data transformations (e.g. abundances into rank abundance curves)	Data-saving, allows using the maximum of information available
Modelling methods commonly grouped under “SDMs”	Regular SDMs with absence data	SDMs with no particular correction effect nor sophistication when enough data are available and meet all modelling assumptions (rarely the case), e.g. GLM	Simple
	SDMs + pseudo-absences	SDMs where no absence data is unavailable. Models either attempt to generate absences where they believe the species to be absent (pseudo-absences) or sample environmental conditions available to the species (background points)	Simple Only requires readily-available presence data
	Bias-corrected SDMs	(Hierarchical) SDMs accounting for different, potential sources of biases due to spatial location, autocorrelation, observation effects, etc. Examples of models are mixed effect models with an observer random effect, models accounting for spatial auto-correlation, SDMs with model-based bias correction, zero-inflated models that allow modelling true and false absences separately	Accurate Particularly appropriate and flexible for rare species modelling Hypothesis-driven
	Multi-scale SDMs	Models incorporating distribution information at multiple grain sizes —Information from distribution data at multiple grain sizes constrain fine-grain predictions —Information on environmental conditions at multiple grain sizes used as inputs	Processes that operate at multiple spatial scales, and ones unrelated to environmental relationships, can be incorporated in to model predictions
	Geographically-structured SDMs	SDM procedure that: (1) splits evaluation data based on spatial clustering of the data; (2) using modelling data (e.g. creation of pseudo-absence/background data), incorporates spatial bias of presence data or taxonomic group	Can use most traditional SDM algorithms (only affects input data) Reduces the risk of overfitting data to spatial biases in sampling data
	Spatial-thinning SDMs	SDM procedure that consists in removing spatially clustered occurrence points to reduce the spatial autocorrelation in input data	Can use most traditional SDM algorithms (only affects input data) Reduces the spatial autocorrelation in input data Reduces the risk of overfitting data to spatial biases in sampling data

Cons	Suitable for which rarity categories?	Examples/references	Input data →Output calculated/ estimate
Often requires to take arbitrary decisions to select thresholds, correcting factors, etc	All	Fithian et al. (2015); Phillips (2009) (correct biases in presence-only data); Nekola et al. (2008) (data transformations)	PA→PA ab→ab
Requires absence data Often too simplistic, resulting in strongly biased results Can suffer overfitting if the number of predictors is too high compared to too few species occurrences Assumes that habitat suitability is the most limiting driver of species distribution Doesn't control for sampling biases or variable detectability	Common species	Guisan and Zimmermann (2000)	PA→relative pocc
Requires data and prior knowledge on habitat suitability Assumes that habitat suitability is the most limiting driver of species distribution	Common species	Barbet-Massin et al. (2012)	P (+background data)→relative lik
Interpretation sometimes difficult Hypothesis-driven Requires information on observational conditions	All	Dormann et al. (2007), Marcer et al. (2013) (models accounting for spatial auto-correlation); Fithian et al. (2015) (mixed effect models with an observer random effect); El-Gabbas and Dormann (2018) (SDMs with model-based bias correction); Zuur et al. (2009) (zero-inflated models)	P→relative lik PA→relative pocc ab + det→relative ab
Complicated fitting frameworks	Common species	Keil et al. (2013) (hierarchical models incorporating distribution information at multiple grain sizes); Rocchini et al. (2017)	PA→ relative pocc P→relative lik
Assumes that habitat suitability is the most limiting driver of species distribution Can cause nearly all data to be assigned to 1–2 folds, and other folds being constructed with v. few occurrence points	Common species	Radosavljevic and Anderson (2014); Phillips et al. (2009)	PA→relative pocc P→relative lik
Assumes that habitat suitability is the most limiting driver of species distribution Reduces the quantity of modelling data	Common species	Boria et al. (2014)	PA→relative pocc P→relative lik

(Continued)

TABLE 3 (Continued)

Method		Brief description	Pros
Ensemble of multiple SDMs	Ensemble SDMs	The procedure that takes outputs from several algorithms of SDMs, weights these outputs based on respective model performances (using e.g. AIC) and generates single 'consensus' predictions by model averaging methods	Does not rely on the single best model Ensemble predictions perform better compared to single modelling techniques Can use variance between models as an estimate of uncertainty
	Ensemble of Small Models (ESM)	Strategy that consists in modelling the distribution of rare species based on fitting a larger number of small (bivariate, trivariate, etc.) models, that is models with only two predictors at a time (although only one or three could also be used), and averaging them in an ensemble prediction using weights based on model performances (e.g. based on AUC score)	Circumvents the risk of overfitting when applying an SDM on too few occurrences data Excellent performance on species data with a low number of occurrences Allows structuring the modelling framework according to different scales of drivers of species distribution (e.g. local vs. climatic predictors)
Bayesian Belief Network SDMs	Bayesian Belief Network SDMs	(a.k.a. Bayesian networks, causal probability networks, acyclic directed graphs) Statistical tool derived from graph theory and Bayesian inference that predicts the probability of ecological responses to varying input assumptions such as habitat and population demography conditions and hypothesized causal relationships	All the pros related to Bayesian statistical frameworks: flexibility, accounting and quantification of uncertainties, integration of prior knowledge information on the rare species of interest, easily updatable with new data/information, etc. Integration, assessment and visualization of causal pathways to explain species distribution Due to its visual nature and relative ease of use, highly suitable for participatory modelling
Occupancy downscaling modelling	Occupancy downscaling modelling	Models that describe the OAR are fitted at large grain sizes to atlas data and then extrapolated to predict occupancy at fine grain sizes Occupancy-Area Relationship (OAR) (or scale-area curve or range-area relationship) (Harte & Kinzig, 1997; Kunin, 1998) is the relationship between the area occupied by a species and the sampling grain size. This relationship is positive and its shape is characteristic of the species distribution pattern (extent, patchiness, prevalence)	By aggregating data at large scales, overcomes sampling gaps (false absences in atlas data) and effects of sampling biases No need for covariates
Modelling methods commonly grouped under "site-occupancy models"	Mark-release-recapture modelling (robust design)	HM using mark-recapture histories to estimate population parameters (colonization, extinction, etc.), occurrence probability and detectability. Requires to fulfil the population closure assumption between the temporal replicates and to have relatively good temporal replication (robust design). Can use covariates to estimate detectability and other potential biases.	Provides accurate estimations of population parameters (e.g. population size, survivorship, fecundity) Provides accurate estimations of detectability (e.g. trap happiness/shyness effects, time-varying capture, sex-dependent detectability) Thanks to the robust design principle, if one has multiple visits that are separated by sufficiently short periods of time, one can consider each visit as fulfilling the assumption of population closure

Cons	Suitable for which rarity categories?	Examples/references	Input data →Output calculated/ estimate
All the cons of SDM approaches above Model averaging also has limitations (e.g. sensitivity to performance score and thresholds used) Predictive performance still questioned	Common species	Araújo and New (2007); Hao et al. (2019, 2020)	PA→relative pocc P→relative lik
Requires to choose thresholds of performance scores to decide which models are included in the ensemble Remains unclear how this method performs for the different forms of rarity, especially the spatially-biased ones, as it is mainly based on the number of occurrences and related IUCN status ESM performance (compared to both single-model Regular SDM and standard Ensemble SDMs) depends on the number of species occurrences available in the data	Cat4, Cat6 (low density but spatially dispersed)	Lomba (2010); Breiner et al. (2015)	P→relative lik
Requires to discretize input predictors with choices of thresholds which can lead to class edge effects (but see Aguilera et al., 2010) More appropriate for risk or conservation category assessment than for predicting or mapping species distribution Assumptions and reasoning behind the hypothesized influence diagram must be clearly documented/ justified as the latter strongly influences predictions	Potentially all (provided that enough prior knowledge and validation data are available)	Marcot et al. (2006a,b); Smith et al. (2007); Aguilera et al. (2010); Chen and Pollino (2012); MacCracken et al. (2012); Hamilton et al. (2015); Van Echelpoel et al. (2015)	P→relative lik PA→relative pocc ab→relative ab
Needs some atlas data Only determines occupancy in terms of proportion of sites or area occupied, i.e. not spatial-explicit May be subject to some errors/uncertainty from the models Requires to think carefully about how to fit the upscaling functions may not be possible to fit models for some species—-e.g. very rare, dispersed species, or very common widespread species—as the OAR reaches the scale of endemism or saturation	Cat1, Cat2, Cat4, Cat5, Cat6	Azaele et al. (2012); Barwell et al. (2014); Marsh et al. (2019)	PA (atlas data)→occupancy (as the proportion of sites or area occupied)
Hypothesis-driven Computationally intensive	All, especially for Cat4, Cat5, Cat6, Cat7 (low local density) but for low local density, it may be challenging to get enough data for reliable estimates	Pollock et al. (1990); MacKenzie et al. (2002); MacKenzie (2006); Willson et al. (2011)	PA + det→pocc

(Continued)

TABLE 3 (Continued)

Method		Brief description	Pros
	Multi-scale occupancy models	HM site-occupancy model that allows estimation of occupancy at different spatial scales to account for different scales of habitat, environmental, ecological or sampling influences; e.g. local habitat vs. landscape-scale effects. The approach accounts for the lack of independence of detections within a sampling occasion and uses this dependence to infer scale-specific occupancy, namely the study area scale and the site scale. This method is a variation of the classical site-occupancy model robust design, except that it does not model seasonal colonization and extinction, but simply presence or absence at the sample unit	Accounts for the scale-dependence of occupancy estimation
N-mixture models	Royle-Nichols models (RN) or Bernoulli-Poisson N-mixture models (for occurrences)	HM that estimates species occurrence probability using different submodels (and potentially different sets of predictors) for the "detection" and the "occurrence" processes. RN model provides the conceptual links between the N-mixture models for abundances and the classical site-occupancy models. RN can estimate abundances from spatio-temporally replicated measurements of presences/absences, can accommodate detection heterogeneity when focusing on occupancy and can link occupancy and abundance data in an integrated model. Some people consider RN as an occupancy model because the modelled data are identical. Can account for spatial autocorrelation using covariates as random or fixed effects	Provides two useful estimates: one for the detection probability and one for the occurrence probability
	N-mixture models for abundances	HM that estimate species abundances using different submodels (and potentially different sets of predictors) for the "detection" and the "abundance" processes. For instance, in "The N-mix" model, the detection probability can be estimated based on a binomial function of some predictors assumed as relevant to the detection process (e.g. vegetation density). This estimation is then incorporated in a (mixed) Poisson model that estimates species abundances (based on predictors relevant to the species ecology) whilst weighting by the imperfect detection (weighted likelihood). Examples of N-mixture models are: zero-inflated, Poisson-binomial, multinomial, Poisson-Poisson, multiscale N-mixture models, hurdle models, spatially-explicit density models	Provides two useful estimates: one for the detectability and one for the relative abundances Provides a fine estimation of species relative abundances With a sufficient amount of data and in some circumstances, some of these models can be used to relax the population closure assumption Zero-inflated and hurdle models are particularly interesting for rare species (due to the high risk of data overdispersion), quite intuitive to use and relatively easy to apply even in a likelihood framework
Occupancy or abundance modelling with multiple detection methods	Occupancy or abundance modelling with multiple detection methods	HM that permits simultaneous use of data from multiple detection methods for inference about method-specific detection probabilities. The approach accounts for the lack of independence of detections within a sampling campaign and uses this dependence to infer method-specific occupancy and detectability	Can be used with data that are produced by different sampling methods and devices (provides device-specific detection probability estimates for use in survey design)

Abbreviations: P, presence only; PA, presence/absence; ab, abundance; pocc, probability of occurrence; det, detection probability; lik, likelihood.

et al., 2017; Thuiller et al., 2019)), and biases due to sampling scales (Keil et al., 2013; Keil & Chase, 2019). Hierarchical models (HM) are especially useful due to their flexibility: they describe, on the one hand, the true state of nature that is not or only partly observable (e.g. variation in occurrence probability potentially due to variation in available resources), and on the other hand,

Cons	Suitable for which rarity categories?	Examples/references	Input data →Output calculated/ estimate
Hypothesis-driven Requires good data with sufficient spatial-temporal replicates and detections	All, providing that sufficient spatial-temporal replicates are available	Nichols et al. (2008); Mordecai et al. (2011) Pavlacky et al. (2012); Hagen et al. (2016)	PA + det→pocc
Requires a sufficient amount of spatio-temporal replications in the data Requires good sets of predictors for both the detection and the occurrence parts of the model	All, especially for Cat4, Cat5, Cat6, Cat7 (low local density)	Royle and Nichols (2003); Kéry and Royle (2015)	PA + det→pocc
Most of these models require good quality and a large amount of abundance data with both spatial and temporal replications (except zero-inflated and hurdle models) Computationally intensive Requires good sets of predictors for both the detection and the abundance parts of the model	All, especially for Cat4, Cat5, Cat6, Cat7 (low local density)	Welsh et al. (2000), Martin et al. (2005), Joseph et al. (2009) (zero-inflated N-mixture models); Royle (2004), Dénes et al. (2015) ("The N-mix" model); Kéry and Royle (2015) (Poisson-binomial/Poisson-Poisson/ multinomial/density models); Cunningham and Lindenmayer (2005), Fletcher et al. (2005), Zuur et al. (2009) (hurdle models); Chandler and Hepinstall-Cymerman (2016) (multiscale N-mixture models)	ab + det→relative ab
If the species of interest is locally rare or solitary, and one of the detection devices is a method that retains (a trap) or repels (a camera's flash) an individual upon detection, then the model needs to be extended to include different device-specific detection probabilities that differ based on whether or not the species was detected by one of the other devices at the immediate sampling site	All, especially for Cat4, Cat5, Cat6, Cat7 (low local density)	Nichols et al. (2008); Giraud et al. (2016); Bowler et al. (2019)	PA + det→pocc ab + det→relative ab

the measurement error (e.g. variation in detection probability potentially due to variable observer skills) (Kéry & Royle, 2015). Multi-scale hierarchical SDMs account for the fact that increasing the sampling extent increases the probability of detecting rare species (Rocchini et al., 2017). HMs thus allow imperfect detectability to be considered in the modelling procedure (Table 3). By integrating prior knowledge, Bayesian Belief Networks explicitly decompose causal pathways involved in the capture rate of species, including respective influences of detection and occupancy in small or incomplete datasets (Uusitalo, 2007): capture can be considered dependent on detectability, influenced by date and trapping effort, and by occupancy, influenced by the suitability of local habitat conditions (Marcot et al., 2006). Such methods have already proved useful for modelling species distributions (Van Echelpoel et al., 2015) and responses of rare and endangered species (Hamilton et al., 2015; Smith et al., 2007; Table 3).

When abundance data from a standardized survey or monitoring protocols are available, these can be used to fit rare species distribution models and track distribution changes (Howard et al., 2014). However, because such protocols usually do not detect most of the rare species, especially clumped and low local-density species (see 'how to sample' section), abundance-based SDMs are rarely possible for rare species.

If recapture data are available, distribution modelling can be done using classical site-occupancy models and different methods developed as mark-release-recapture analyses (MacKenzie et al., 2017; Pollock et al., 1990; Table 3).

For occurrence data from spatio-temporally replicated measurements of presences/absences, under the assumption of population closure (i.e. if the populations did not exchange propagules between the time steps under study), the Royle-Nichols model (Kéry & Royle, 2015; Royle & Nichols, 2003) allows occurrence probability to be estimated and detection heterogeneity accommodated (Table 3; Figure 1d). When 'unmarked' abundance data are available, N-mixture models can estimate both detectability and abundances used in large-scale species distribution modelling (Guélat & Kéry, 2018; Jakob et al., 2014; Kéry, 2018; Table 3; Figure 1d). When potential sources of measurement bias are known (e.g. type of observer, weather, vegetation density), these can be integrated as covariates in the latent state submodel (e.g. Cunningham & Lindenmayer, 2005).

When data are zero-inflated, as typical for rare species data, variants of Royle-Nichols or N-mixture models can be applied that allow extra parameters and account for data overdispersion. Variants of N-mixture models have further been developed that address spatial bias and scale dependence, such as variation of sampling grain size (Keil et al., 2018) or scales of environmental influence (Chandler & Hepinstall-Cymerman, 2016). However, the underlying assumptions are quite restrictive for species distribution modelling and further simulation studies are needed to assess their performance with rare species when assumptions are not met. This approach is also not necessarily the most cost-effective strategy when it comes to tracking species distribution changes over time compared with presence/absence data (Joseph et al., 2006).

When multiple types of data are available (presences, presence/absence, abundance), their combination within single modelling frameworks provides valuable insights into predicting species distributions, occupancy, even abundance (Table 3). Even if available over a restricted spatial extent, multiple sources of abundance data can be used together with more extensive data, such as occupancy surveys or opportunistic observations. HMs can include different submodels for the different sources of data, and potential detection biases, and incorporate environmental data as potential predictors of occupancy and/or abundance. Such methods allow extrapolation and even comparison of rare species' distributions across large spatial scales (e.g. Bowler et al., 2019; Giraud et al., 2016) and potentially for all categories of rarity if data sources are available and models well built (Figure 1d).

To summarize, the model choice will mainly depend on the nature of the data and the biases involved. From presence only, to presence/absence, to abundance, in Figure 1d, there is a change in the temporal comparability of SDMs, and thus their ability to track distributional changes. In the top row, the output is a relative likelihood, which is not comparable even for a given species over multiple time steps. Naïve presence/absence SDMs provide an estimate that does not separate the probability of occurrence from detectability, but if we assume detectability to be constant across time and space (including no drastic change in abundances), the resulting probability map is comparable for a given species over time. Finally, population size information allows the separate estimation of detectability and probability of occurrence, which is comparable over time, species and space. Comparability is important as it enables conservationists to assess changes in the environmental suitability, and ideally (see Dallas & Hastings, 2018; Jiménez-Valverde et al., 2021; Weber et al., 2017), in the distribution of rare species that could require revision of a species' status and protection needs.

More generally, whatever the type of rarity, several methodological aspects are to be considered to ensure SDM quality, including predictor selection (e.g. Le Rest et al., 2014; Saupe et al., 2012; Williams et al., 2012), model averaging (e.g. Burnham & Anderson, 2004; Dormann et al., 2018), spatial-explicit cross-validation (e.g. Roberts et al., 2017), optimization of model performance (e.g. Anderson & Gonzalez, 2011; Norberg et al., 2019; Radosavljevic & Anderson, 2014), and testing or improvement of the extrapolation abilities of the fitted models (e.g. Mesgaran et al., 2014; Owens et al., 2013; Qiao et al., 2019; Stohlgren et al., 2011; Zurell et al., 2012).

5 | CONCLUSION AND FUTURE PERSPECTIVES

Protecting species diversity implies protecting rare species. However, surveying and modelling rare species involve considerable methodological challenges. In this paper, we have identified how the main decisions on sampling strategy condition properties of the data, and how these in turn condition the range of appropriate modelling methods. With this perspective, we provide guidelines to optimize monitoring and modelling of rare species depending on their rarity characteristics and to ensure consistency between sampling methods, and modelling approaches (Figure 1).

Significant data on the occurrence of species is collected by citizen scientists (Amano et al., 2016; Chandler et al., 2017). It is highly valuable for monitoring biodiversity at different scales, but often biased and limited to specific areas. Whilst there are ways to correct biases in such data (Bird et al., 2014; Robinson et al., 2018), for monitoring "rarest" species (i.e. narrow distributional range, clumped population, low local density), a systematic species-targeted sampling design may be preferred. Significant advances are expected from advanced remote sensing techniques, genetic tools and using detection dogs, all with the potential to significantly increase the detection rate of rare species at comparatively low cost and with more or less bias towards the species. Above all, future research is still needed to integrate the type of rarity more explicitly into decisions on how and where to sample with the selection of appropriate models. Another challenge with respect to species conservation is that, although the rarity status is defined with respect to endemicity over a given period, it may be dynamic in the longer term, requiring constant adaptation of assessment strategies.

Considering most forms of rarity, our synthesis highlights the particular potential of HMs as a flexible tool to improve rarity modelling whilst accounting for spatial, observer and species-specific biases. Advances in zero-inflation modelling, in particular, have to be better integrated into rare species distribution modelling as both the conceptual and technical foundations of these approaches impact the rarity sampling and modelling issues. Considering the rarest forms of rarity, our synthesis suggests that recent HM developments to combine multiple sources of data are extremely promising (Figure 1).

Other promising perspectives have recently emerged, such as functional rarity modelling (Carmona et al., 2017; Violle et al., 2017) and the use of co-occurring species information (or the "neighbourly advice" (McInerny & Purves, 2011)) and of positive associations among rare species (Calatayud et al., 2019; Hines & Keil, 2020) as potentially valuable information to model rarity distribution. Other model developments include harnessing information from other sources that either directly informs a species' distribution at larger scales, such as incorporating expert-drawn range maps (Merow et al., 2017) or elevation ranges (Ellis-Soto et al., 2021) as model offsets. Joint species distribution models (JSDMs), which model multiple species simultaneously to infer the species' environmental response based on species co-occurrences (Ovaskainen & Soininen, 2011; Pollock et al., 2014), often incorporate ancillary information such as trait (Pollock et al., 2012) or phylogenetic similarity (Ovaskainen et al., 2017) and are promising further developments for rare species modelling (Tobler et al., 2019). Finally, machine-learning-based methods, including non-parametric methods, and methods tolerant of unstructured data, have shown promise for modelling and mapping rarity with a strong predictive ability (Pouteau et al., 2012; Robinson et al., 2018). Further research and sensitivity analyses are needed to assess the appropriateness of these methods in the workflow of rarity sampling and modelling, depending on the rarity type of the species.

ACKNOWLEDGEMENTS

We thank the EU BON project funded by the European Commission (EC) under the 7th Framework Programme (contract no. 308454) (Hoffmann et al., 2014) for making this study happen and for funding a part of the working group at an early stage. We thank an Anonymous Reviewer and Nigel G. Yoccoz for very constructive and useful comments on our manuscript and Anne-Christine Monnet for her help with SDM terminology.

AUTHOR CONTRIBUTION

YG and AJ have equally contributed to ideas, first drafting of the paper, production of tables and figures, and coordination of the writing process. WK and KH, at the origin of the project, have jointly supervised the work as project leaders (EU BON) and have equally contributed with the central ideas, orientation and deep revisions of the manuscript. All authors have contributed with ideas and revisions on all versions of the manuscript.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Alienor Jeliazkov https://orcid.org/0000-0001-5765-3721
Yoni Gavish https://orcid.org/0000-0002-6025-5668
Charles J. Marsh https://orcid.org/0000-0002-0281-3115
Jonas Geschke https://orcid.org/0000-0002-5654-9313
Neil Brummitt https://orcid.org/0000-0001-7769-4395
Duccio Rocchini https://orcid.org/0000-0003-0087-0594
Peter Haase https://orcid.org/0000-0002-9340-0438
William E. Kunin https://orcid.org/0000-0002-9812-2326
Klaus Henle https://orcid.org/0000-0002-6647-5362

REFERENCES

Aguilera, P. A., Fernández, A., Reche, F., & Rumí, R. (2010). Hybrid Bayesian network classifiers: Application to species distribution models. *Environmental Modelling & Software*, *25*, 1630–1639. https://doi.org/10.1016/j.envsoft.2010.04.016

Aizpurua, O., Paquet, J.-Y., Brotons, L., & Titeux, N. (2015). Optimising long-term monitoring projects for species distribution modelling: How atlas data may help. *Ecography*, *38*(1), 29–40. https://doi.org/10.1111/ecog.00749

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., & Edwards, M. E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. *PLoS One*, *13*(4), e0195403. https://doi.org/10.1371/journal.pone.0195403

Altermatt, F., Little, C. J., Mächler, E., Wang, S., Zhang, X., & Blackman, R. C. (2020). Uncovering the complete biodiversity structure in spatial networks: The example of riverine systems. *Oikos*, *129*(5), 607–618. https://doi.org/10.1111/oik.06806

Amano, T., Lamming, J. D. L., & Sutherland, W. J. (2016). Spatial gaps in global biodiversity information and the role of citizen science. *BioScience*, *66*(5), 393–400. https://doi.org/10.1093/biosci/biw022

Anderson, R. P., & Gonzalez, I. (2011). Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling*, *222*(15), 2796–2811. https://doi.org/10.1016/j.ecolmodel.2011.04.011

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, *22*(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010

Araújo, M. B., Williams, P. H., & Fuller, R. J. (2002). Dynamics of extinction and the selection of nature reserves. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1504), 1971–1980. https://doi.org/10.1098/rspb.2002.2121

Aubry, K. B., Raley, C. M., & McKelvey, K. S. (2017). The importance of data quality for generating reliable distribution models for rare, elusive, and cryptic species. *PLoS One*, *12*(6), e0179152. https://doi.org/10.1371/journal.pone.0179152

Austin, M. P., Cunningham, R. B., & Fleming, P. M. (1984). New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio*, *55*(1), 11–27. https://doi.org/10.1007/BF00039976

Azaele, S., Cornell, S. J., & Kunin, W. E. (2012). Downscaling species occupancy from coarse spatial scales. *Ecological Applications*, *22*, 1004–1014. https://doi.org/10.1890/11-0536.1

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo?absences for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, *3*(2), 327–338. https://doi.org/10.1111/j.2041-210X.2011.00172.x

Barwell, L. J., Azaele, S., Kunin, W. E., & Isaac, N. J. B. (2014). Can coarse-grain patterns in insect atlas data predict local occupancy? *Diversity and Distributions*, *20*, 895–907.

Beng, K. C., & Corlett, R. T. (2020). Applications of environmental DNA (eDNA) in ecology and conservation: Opportunities, challenges and prospects. *Biodiversity and Conservation*, *29*, 2089–2121. https://doi.org/10.1007/s10531-020-01980-0

Benito, B. M., Martínez-Ortega, M. M., Muñoz, L. M., Lorite, J., & Peñas, J. (2009). Assessing extinction-risk of endangered plants using species distribution models: A case study of habitat depletion caused by the spread of greenhouses. *Biodiversity and Conservation*, *18*(9), 2509–2520. https://doi.org/10.1007/s10531-009-9604-8

Bird, T. J., Bates, A. E., Lefcheck, J. S., Hill, N. A., Thomson, R. J., Edgar, G. J., Stuart-Smith, R. D., Wotherspoon, S., Krkosek, M., Stuart-Smith, J. F., Pecl, G. T., Barrett, N., & Frusher, S. (2014). Statistical solutions for error and bias in global citizen science datasets. *Biological Conservation*, *173*, 144–154. https://doi.org/10.1016/j.biocon.2013.07.037

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, *29*, 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, *275*, 73–77. https://doi.org/10.1016/j.ecolmodel.2013.12.012

Bowler, D. E., Nilsen, E. B., Bischof, R., O'Hara, R. B., Yu, T. T., Oo, T., Aung, M., & Linnell, J. D. C. (2019). Integrating data from different survey types for population monitoring of an endangered species: the case of the Eld's deer. *Scientific Reports*, *9*, 7766. https://doi.org/10.1038/s41598-019-44075-9

Breiner, F. T., Guisan, A., Bergamini, A., & Nobis, M. P. (2015). Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution*, *6*, 1210–1218. https://doi.org/10.1111/2041-210X.12403

Bruelheide, H., Dengler, J., Jiménez-Alfaro, B., Purschke, O., Hennekens, S. M., Chytrý, M., Pillar, V. D., Jansen, F., Kattge, J., Sandel, B., Aubin, I., Biurrun, I., Field, R., Haider, S., Jandt, U., Lenoir, J., Peet, R. K., Peyre, G., Sabatini, F. M., … Zverev, A. (2019). sPlot—A new tool for global vegetation analyses. *Journal of Vegetation Science*, *30*, 161–186. https://doi.org/10.1111/jvs.12710

Brummitt, N. A., Bachman, S. P., Griffiths-Lee, J., Lutz, M., Moat, J. F., Farjon, A., Donaldson, J. S., Hilton-Taylor, C., Meagher, T. R., Albuquerque, S., Aletrari, E., Andrews, A. K., Atchison, G., Baloch, E., Barlozzini, B., Brunazzi, A., Carretero, J., Celesti, M., Chadburn, H., … Nic Lughadha, E. M. (2015). Green plants in the red: A baseline global assessment for the IUCN sampled red list index for plants. *PLoS One*, *10*(8), e0135152. https://doi.org/10.1371/journal.pone.0135152

Buckland, S. T., Rexstad, E. A., Marques, T. A., & Oedekoven, C. S. (2015). *Distance sampling: Methods and applications*. Springer.

Bunce, R., Carey, P., Maskell, L., Norton, L., Scott, R., Smart, S., & Wood, C. (2014). Countryside Survey 2007 vegetation plot data.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Butchart, S. H. M., Marco, M. D., & Watson, J. E. M. (2016). Formulating smart commitments on biodiversity: Lessons from the Aichi targets. *Conservation Letters*, *9*(6), 457–468. https://doi.org/10.1111/conl.12278

Bystriakova, N., Peregrym, M., Erkens, R. H. J., Bezsmertna, O., & Schneider, H. (2012). Sampling bias in geographic and environmental space and its effect on the predictive power of species distribution models. *Systematics and Biodiversity*, *10*(3), 305–315. https://doi.org/10.1080/14772000.2012.705357

Calatayud, J., Andivia, E., Escudero, A., Melián, C. J., Bernardo-Madrid, R., Stoffel, M., Aponte, C., Medina, N. G., Molina-Venegas, R., Arnan, X., Rosvall, M., Neuman, M., Noriega, J. A., Alves-Martins, F., Draper, I., Luzuriaga, A., Ballesteros-Cánovas, J. A., Morales-Molino, C., Ferrandis, P., … Madrigal-González, J. (2019). Positive associations among rare species and their persistence in ecological assemblages. *Nature Ecology & Evolution*, *4*(1), 40–45. https://doi.org/10.1038/s41559-019-1053-5

Carey, P. D., Wallis, S., Emmett, B., Maskell, L., Murphy, J., Norton, L., Simpson, I., & Smart, S. (2008). Countryside Survey: UK headline messages from 2007.

Carmona, C. P., de Bello, F., Sasaki, T., Uchida, K., & Pärtel, M. (2017). Towards a common toolbox for rarity: A response to Violle et al. *Trends in Ecology & Evolution*, *32*(12), 889–891. https://doi.org/10.1016/j.tree.2017.09.010

Carraro, L., Hartikainen, H., Jokela, J., Bertuzzo, E., & Rinaldo, A. (2018). Estimating species distribution and abundance in river networks using environmental DNA. *Proceedings of the National Academy of Sciences*, *115*(46), 11724–11729. https://doi.org/10.1073/pnas.1813843115

Carraro, L., Stauffer, J. B., & Altermatt, F. (2021). How to design optimal eDNA sampling strategies for biomonitoring in river networks. *Environmental DNA*, *3*(1), 157–172. https://doi.org/10.1002/edn3.137

Chandler, M., See, L., Copas, K., Bonde, A. M. Z., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., Rosemartin, A., & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, *213*, 280–294. https://doi.org/10.1016/j.biocon.2016.09.004

Chandler, R., & Hepinstall-Cymerman, J. (2016). Estimating the spatial scales of landscape effects on abundance. *Landscape Ecology*, *31*, 1383–1394. https://doi.org/10.1007/s10980-016-0380-z

Chandler, R. B., Royle, J. A., & King, D. I. (2011). Inference about density and temporary emigration in unmarked populations. *Ecology*, *92*, 1429–1435. https://doi.org/10.1890/10-2433.1

Chen, S. H., & Pollino, C. A. (2012). Good practice in Bayesian network modelling. *Environmental Modelling & Software*, *37*, 134–145. https://doi.org/10.1016/j.envsoft.2012.03.012

Chiffard, J., Marciau, C., Yoccoz, N. G., Mouillot, F., Duchateau, S., Nadeau, I., Fontanilles, P., & Besnard, A. (2020). Adaptive niche-based sampling to improve ability to find rare and elusive species:

Simulations and field tests. *Methods in Ecology and Evolution*, *11*(8), 899–909. https://doi.org/10.1111/2041-210X.13399

Courchamp, F., Angulo, E., Rivalan, P., Hall, R. J., Signoret, L., Bull, L., & Meinard, Y. (2006). Rarity value and species extinction: The anthropogenic allee effect. *PLoS Biology*, *4*(12), e415. https://doi.org/10.1371/journal.pbio.0040415

Cristescu, M. E., & Hebert, P. D. N. (2018). Uses and misuses of environmental DNA in biodiversity science and conservation. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), 209–230. https://doi.org/10.1146/annurev-ecolsys-110617-062306

Cunningham, R. B., & Lindenmayer, D. B. (2005). Modeling count data of rare species: Some statistical issues. *Ecology*, *86*(5), 1135–1142. https://doi.org/10.1890/04-0589

Dallas, T. A., & Hastings, A. (2018). Habitat suitability estimated by niche models is largely unrelated to species abundance. *Global Ecology and Biogeography*, *27*(12), 1448–1456. https://doi.org/10.1111/geb.12820

Deguines, N., Julliard, R., de Flores, M., & Fontaine, C. (2012). The whereabouts of flower visitors: contrasting land-use preferences revealed by a country-wide survey based on citizen science. *PLoS One*, *7*, e45822. https://doi.org/10.1371/journal.pone.0045822

Dénes, F. V., Silveira, L. F., & Beissinger, S. R. (2015). Estimating abundance of unmarked animal populations: Accounting for imperfect detection and other sources of zero inflation. *Methods in Ecology and Evolution*, *6*, 543–556. https://doi.org/10.1111/2041-210X.12333

Didham, R. K., Basset, Y., Collins, C. M., Leather, S. R., Littlewood, N. A., Menz, M. H. M., Müller, J., Packer, L., Saunders, M. E., Schönrogge, K., Stewart, A. J. A., Yanoviak, S. P., & Hassall, C. (2020). Interpreting insect declines: Seven challenges and a way forward. *Insect Conservation and Diversity*, *13*(2), 103–114. https://doi.org/10.1111/icad.12408

Diekmann, M., Kühne, A., & Isermann, M. (2007). Randomvs non-random sampling: Effects on patterns of species abundance, species richness and vegetation-environment relationships. *Folia Geobot*, *42*, 179.

Domínguez Lozano, F., Moreno Saiz, J. C., Sainz Ollero, H., & Schwartz, M. W. (2007). Effects of dynamic taxonomy on rare species and conservation listing: Insights from the Iberian vascular flora. *Biodiversity and Conservation*, *16*(14), 4039–4050. https://doi.org/10.1007/s10531-007-9206-2

Dorazio, R. M. (2014). Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography*, *23*(12), 1472–1484. https://doi.org/10.1111/geb.12216

Dormann, C. F., Calabrese, J. M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., Beale, C. M., Ciuti, S., Elith, J., Gerstner, K., Guelat, J., Keil, P., Lahoz-Monfort, J. J., Pollock, L. J., Reineking, B., Roberts, D. R., Schröder, B., Thuiller, W., Warton, D. I., … Hartig, F. (2018). Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, *88*(4), 485–504. https://doi.org/10.1002/ecm.1309

Dormann, C., McPherson, J., Araújo, M., Bivand, R., Bolliger, J., Carl, G., Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Kühn, I., Ohlemüller, R., Peres-Neto, P., Reineking, B., Schröder, B., Schurr, F., & Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. *Ecography*, *30*, 609–628. https://doi.org/10.1111/j.2007.0906-7590.05171.x

El-Gabbas, A., & Dormann, C. F. (2018). Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. *Ecography*, *41*, 1161–1172. https://doi.org/10.1111/ecog.03149

Ellis-Soto, D., Merow, C., Amatulli, G., Parra, J. L., & Jetz, W. (2021). Continental-scale 1 km hummingbird diversity derived from fusing point records with lateral and elevational expert information. *Ecography*, *44*(4), 640–652. https://doi.org/10.1111/ecog.05119

Enquist, B. J., Condit, R., Peet, R. K., Schildhauer, M., & Thiers, B. M. (2016). Cyberinfrastructure for an integrated botanical information network to investigate the ecological impacts of global climate change on plant biodiversity. https://peerj.com/preprints/2615v2; doi:https://doi.org/10.7287/peerj.preprints.2615v2

Enquist, B. J., Feng, X., Boyle, B., Maitner, B., Newman, E. A., Jørgensen, P. M., Roehrdanz, P. R., Thiers, B. M., Burger, J. R., Corlett, R. T., Couvreur, T. L. P., Dauby, G., Donoghue, J. C., Foden, W., Lovett, J. C., Marquet, P. A., Merow, C., Midgley, G., Morueta-Holme, N., … McGill, B. J. (2019). The commonness of rarity: Global and future distribution of rarity across land plants. *Science Advances*, *5*(11), eaaz0414. https://doi.org/10.1126/sciadv.aaz0414

Fisher, R. A., Corbet, A. S., & Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, *12*(1), 42–58. https://doi.org/10.2307/1411

Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, *6*, 424–438. https://doi.org/10.1111/2041-210X.12242

Fleschutz, M. M., Gálvez, N., Pe'er, G., Davies, Z. G., Henle, K., & Schüttler, E. (2016). Response of a small felid of conservation concern to habitat fragmentation. *Biodiversity and Conservation*, *25*(8), 1447–1463. https://doi.org/10.1007/s10531-016-1118-6

Fletcher, D., MacKenzie, D., & Villouta, E. (2005). Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. *Environmental and Ecological Statistics*, *12*, 45–54. https://doi.org/10.1007/s10651-005-6817-1

Fontaine, B., Bouchet, P., Van Achterberg, K., Alonso-Zarazaga, M. A., Araujo, R., Asche, M., Aspöck, U., Audisio, P., Aukema, B., Bailly, N., Balsamo, M., Bank, R. A., Barnard, P., Belfiore, C., Bogdanowicz, W., Bongers, T., Boxshall, G., Burckhardt, D., Camicas, J.-L., … Willmann, R. (2007). The European union's 2010 target: Putting rare species in focus. *Biological Conservation*, *139*(1), 167–185. https://doi.org/10.1016/j.biocon.2007.06.012

Fortin, M.-J., Drapeau, P., & Legendre, P. (1989). Spatial autocorrelation and sampling design in plant ecology. *Vegetatio*, *83*, 209–222. https://doi.org/10.1007/BF00031693

Galante, P. J., Alade, B., Muscarella, R., Jansa, S. A., Goodman, S. M., & Anderson, R. P. (2018). The challenge of modeling niches and distributions for data-poor species: A comprehensive approach to model complexity. *Ecography*, *41*(5), 726–736. https://doi.org/10.1111/ecog.02909

Gärdenfors, U., Hilton-Taylor, C., Mace, G. M., & Rodríguez, J. P. (2001). The application of IUCN Red List criteria at regional levels. *Conservation Biology*, *15*(5), 1206–1212. https://doi.org/10.1111/j.1523-1739.2001.00112.x

Gaston, K. J. (1998). Species-range size distributions: Products of speciation, extinction and transformation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1366), 219–230. https://doi.org/10.1098/rstb.1998.0204

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Giraud, C., Calenge, C., Coron, C., & Julliard, R. (2016). Capitalizing on opportunistic data for monitoring relative abundances of species. *Biometrics*, *72*(2), 649–658. https://doi.org/10.1111/biom.12431

Green, R. H., & Young, R. C. (1993). Sampling to detect rare species. *Ecological Applications*, *3*(2), 351–356. https://doi.org/10.2307/1941837

Greig-Smith, P. (1964). *Quantitative plant ecology*. Butterworths.

Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., Steffen, W., Glaser, G., Kanie, N., & Noble, I. (2013). Sustainable development goals for people and planet. *Nature*, *495*(7441), 305–307. https://doi.org/10.1038/495305a

Grimm-Seyfarth, A., & Klenke, R. (2019). Wie findet man schwer zu erfassende Arten? Vorteile und Limitierungen von Artenspürhunden.

In C. Schüler, & P. Kaul (Eds.), *Faszinosum Spürhunde—Dem Geruch auf der Spur*. Tagungsergebnisse des 4. Symposiums für Odorologie im Diensthundewesen an der Hochschule Bonn-Rhein-Sieg.: Vol. Band 2 (Schriften der Arbeitsgemeinschaft Odorologie e.V., pp. 40–47).

Grimm-Seyfarth, A., Zarzycka, A., Nitz, T., Heynig, L., Weissheimer, N., Lampa, S., & Klenke, R. (2019). Performance of detection dogs and visual searches for scat detection and discrimination amongst related species with identical diets. *Nature Conservation*, *37*, 81. https://doi.org/10.3897/natureconservation.37.48208

Guélat, J., & Kéry, M. (2018). Effects of spatial autocorrelation and imperfect detection on species distribution models. *Methods in Ecology and Evolution*, *9*(6), 1614–1625. https://doi.org/10.1111/2041-210X.12983

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., & Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, *24*(3), 276–292. https://doi.org/10.1111/geb.12268

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*(9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: With applications in R*. Cambridge University Press.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. A., Elith, J., Schwartz, M. W., Wintle, B. A., Broennimann, O., Austin, M., Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*, 1424–1435.

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Hagen, C. A., Pavlacky, D. C., Adachi, K., Hornsby, F. E., Rintz, T. J., & McDonald, L. L. (2016). Multiscale occupancy modeling provides insights into range-wide conservation needs of Lesser Prairie-Chicken (*Tympanuchus pallidicinctus*). *The Condor American Ornithological Society*, *118*, 597–612.

Hamilton, S. H., Pollino, C. A., & Jakeman, A. J. (2015). Habitat suitability modelling of rare species using Bayesian networks: Model evaluation under limited data. *Ecological Modelling*, *299*, 64–78. https://doi.org/10.1016/j.ecolmodel.2014.12.004

Hao, T., Elith, J., Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2019). A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Diversity and Distributions*, *25*(5), 839–852. https://doi.org/10.1111/ddi.12892

Hao, T., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography*, *43*(4), 549–558. https://doi.org/10.1111/ecog.04890

Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, *2*, e616. https://doi.org/10.7717/peerj.616

Harte, J., & Kinzig, A. P. (1997). On the implications of species-area relationships for endemism, spatial turnover, and food web patterns. *Oikos*, *80*(3), 417–427. https://doi.org/10.2307/3546614

Hartley, S., & Kunin, W. E. (2003). Scale dependency of rarity, extinction risk, and conservation priority. *Conservation Biology*, *17*(6), 1559–1570. https://doi.org/10.1111/j.1523-1739.2003.00015.x

Hedgren, O., & Weslien, J. (2008). Detecting rare species with random or subjective sampling: A case study of red-listed saproxylic beetles in boreal Sweden: Random or subjective sampling. *Conservation Biology*, *22*, 212–215. https://doi.org/10.1111/j.1523-1739.2007.00848.x

Hefley, T. J., Baasch, D. M., Tyre, A. J., & Blankenship, E. E. (2014). Correction of location errors for presence-only species distribution models. *Methods in Ecology and Evolution*, *5*(3), 207–214. https://doi.org/10.1111/2041-210X.12144

Helmstetter, N. A., Conway, C. J., Stevens, B. S., & Goldberg, A. R. (2021). Balancing transferability and complexity of species distribution models for rare species conservation. *Diversity and Distributions*, *27*(1), 95–108. https://doi.org/10.1111/ddi.13174

Henle, K., Davies, K. F., Kleyer, M., Margules, C., & Settele, J. (2004). Predictors of species sensitivity to fragmentation. *Biodiversity & Conservation*, *13*(1), 207–251. https://doi.org/10.1023/B:BIOC.0000004319.91643.9e

Henle, K., Kunin, W., Schweiger, O., Schmeller, D. S., Grobelnik, V., Matsinos, Y., Pantis, J., Penev, L., Potts, S. G., Ring, I., Similä, J., Tzanopoulos, J., van den Hove, S., Baguette, M., Clobert, J., Excoffier, L., Framstad, E., Grodzińska-Jurczak, M., Lengyel, S., … Settele, J. (2010). Securing the conservation of biodiversity across administrative levels and spatial, temporal, and ecological scales—Research needs and approaches of the SCALES project. *GAIA—Ecological Perspectives for Science and Society*, *19*(3), 187–193. https://doi.org/10.14512/gaia.19.3.8

Hermoso, V., Kennard, M. J., & Linke, S. (2015). Evaluating the costs and benefits of systematic data acquisition for conservation assessments. *Ecography*, *38*(3), 283–292. https://doi.org/10.1111/ecog.00792

Hines, J., & Keil, P. (2020). Common competitors and rare friends. *Nature Ecology & Evolution*, *4*(1), 8–9. https://doi.org/10.1038/s41559-019-1071-3

Hoffmann, A., Penner, J., Vohland, K., Cramer, W., Doubleday, R., Henle, K., Kõljalg, U., Kühn, I., Kunin, W. E., Negro, J. J., & Häuser, C. L. (2014). *Improved access to integrated biodiversity data for science, practice, and policy-the European Biodiversity Observation Network (EU BON)* (pp. 49–66). Nature Conservation Pensoft Publishers.

Hollerbach, L., Heurich, M., Reiners, T. E., & Nowak, C. (2018). Detection dogs allow for systematic non-invasive collection of DNA samples from Eurasian lynx. *Mammalian Biology*, *90*, 42–46. https://doi.org/10.1016/j.mambio.2018.02.003

Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., & Willis, S. G. (2014). Improving species distribution models: The value of data on abundance. *Methods in Ecology and Evolution*, *5*(6), 506–513. https://doi.org/10.1111/2041-210X.12184

IPBES. (2019). *Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. .

Işik, K. (2011). Rare and endemic species: Why are they prone to extinction? *Turkish Journal of Botany*, *35*(4), 411–417.

Jakob, C., Ponce-Boutin, F., & Besnard, A. (2014). Coping with heterogeneity to detect species on a large scale: N-mixture modeling applied to red-legged partridge abundance. *The Journal of Wildlife Management*, *78*(3), 540–549. https://doi.org/10.1002/jwmg.686

Jeliazkov, A., Bas, Y., Kerbiriou, C., Julien, J.-F., Penone, C., & Le Viol, I. (2016). Large-scale semi-automated acoustic monitoring allows to detect temporal decline of bush-crickets. *Global Ecology and Conservation*, *6*, 208–218. https://doi.org/10.1016/j.gecco.2016.02.008

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, *4*(2), 150–157. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Jiguet, F., Devictor, V., Julliard, R., & Couvet, D. (2012). French citizens monitoring ordinary birds provide tools for conservation and ecological sciences. *Acta Oecologica*, *44*, 58–66. https://doi.org/10.1016/j.actao.2011.05.003

Jiménez-Valverde, A., Aragón, P., & Lobo, J. M. (2021). Deconstructing the abundance–suitability relationship in species distribution modelling. *Global Ecology and Biogeography*, *30*(1), 327–338. https://doi.org/10.1111/geb.13204

Joseph, L. N., Elkin, C., Martin, T. G., & Possingham, H. P. (2009). Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. *Ecological Applications*, *19*.

Joseph, L. N., Field, S. A., Wilcox, C., & Possingham, H. P. (2006). Presence-absence versus abundance data for monitoring threatened species. *Conservation Biology*, *20*(6), 1679–1687. https://doi.org/10.1111/j.1523-1739.2006.00529.x

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*(1), 82–94. https://doi.org/10.1111/j.2041-210x.2012.00264.x

Keil, P., & Chase, J. M. (2019). Global patterns and drivers of tree diversity integrated across a continuum of spatial grains. *Nature Ecology & Evolution*, *3*(3), 390–399. https://doi.org/10.1038/s41559-019-0799-0

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*(1), 2–13. https://doi.org/10.1111/geb.12669

Kenkel, N. C., Juhász-Nagy, P., & Podani, J. (1990). On sampling procedures in population and community ecology. In G. Grabherr, L. Mucina, M. B. Dale, & C. J. F. T. Braak (Eds.), *Progress in theoretical vegetation science* (pp. 195–207). Springer. https://doi.org/10.1007/978-94-009-1934-1_17

Kéry, M. (2018). Identifiability in N-mixture models: A large-scale screening test with bird data. *Ecology*, *99*(2), 281–288. https://doi.org/10.1002/ecy.2093

Kéry, M., & Royle, J. A. (2015). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 1: Prelude and static models*. Academic Press.

Krebs, C. J. (1989). *Ecological methodology*. Harper & Row.

Kunin, W. E. (1998). Extrapolating species abundance across spatial scales. *Science*, *281*(5382), 1513–1515. https://doi.org/10.1126/science.281.5382.1513

Kunin, W. E., & Gaston, K. J. (1993). The biology of rarity: Patterns, causes and consequences. *Trends in Ecology & Evolution*, *8*(8), 298–301. https://doi.org/10.1016/0169-5347(93)90259-R

Le Lay, G., Engler, R., Franc, E., & Guisan, A. (2010). Prospective sampling based on model ensembles improves the detection of rare species. *Ecography*, *33*, 1015–1027. https://doi.org/10.1111/j.1600-0587.2010.06338.x

Le Rest, K., Pinaud, D., Monestiez, P., Chadoeuf, J., & Bretagnolle, V. (2014). Spatial leave-one-out cross-validation for variable selection in the presence of spatial autocorrelation. *Global Ecology and Biogeography*, *23*(7), 811–820. https://doi.org/10.1111/geb.12161

Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P., & Zizka, V. M. A. (2021). Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification. *Environmental DNA*, *3*(1), 261–276. https://doi.org/10.1002/edn3.177

Lin, Y.-P., Lin, W.-C., Wang, Y.-C., Lien, W.-Y., Ding, T.-S., Lee, P.-F., Wu, T.-Y., Klenke, A., Schmeller, D. S., & Henle, K. (2014). An optimal spatial sampling approach for modelling the distribution of species. In I. K. Henle, S. Potts, W. Kunin, Y. Matsinos, J. Simila, J. Pantis, V. Grobelnik, L. Penev, & J. Settele (Eds.), *Scaling in ecology and biodiversity conservation*. Pensoft Publishers.

Lomba, A., Pellissier, L., Randin, C., Vicente, J., Moreira, F., Honrado, J., & Guisan, A. (2010). Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. *Biological Conservation*, *143*, 2647–2657. https://doi.org/10.1016/j.biocon.2010.07.007

MacCracken, J. G., Garlich-Miller, J., Snyder, J., & Meehan, R. (2012). Bayesian belief network models for species assessments: An example with the Pacific walrus. *Wildlife Society Bulletin*, *37*, 226–235. https://doi.org/10.1002/wsb.229

Mächler, E., Little, C. J., Wüthrich, R., Alther, R., Fronhofer, E. A., Gounand, I., Harvey, E., Hürlemann, S., Walser, J.-C., & Altermatt, F. (2019). Assessing different components of diversity across a river network using eDNA. *Environmental DNA*, *1*(3), 290–301. https://doi.org/10.1002/edn3.33

MacKenzie, D. I. (2006). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. Elsevier.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., & Hines, J. E. (2017). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. Elsevier.

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

Mackenzie, D. I., & Royle, J. A. (2005). Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology*, *42*(6), 1105–1114. https://doi.org/10.1111/j.1365-2664.2005.01098.x

Madow, W. G. (1953). On the theory of systematic sampling, III. Comparison of centered and random start systematic sampling. *The Annals of Mathematical Statistics*, *24*, 101–106. https://doi.org/10.1214/aoms/1177729087

Magurran, A. E., Baillie, S. R., Buckland, S. T., Dick, J. M. P., Elston, D. A., Scott, E. M., Smith, R. I., Somerfield, P. J., & Watt, A. D. (2010). Long-term datasets in biodiversity research and monitoring: Assessing change in ecological communities through time. *Trends in Ecology & Evolution*, *25*(10), 574–582. https://doi.org/10.1016/j.tree.2010.06.016

Magurran, A. E., & Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, *422*(6933), 714–716. https://doi.org/10.1038/nature01547

Marcer, A., Sáez, L., Molowny-Horas, R., Pons, X., & Pino, J. (2013). Using species distribution modelling to disentangle realised versus potential distributions for rare species conservation. *Biological Conservation*, *166*, 221–230. https://doi.org/10.1016/j.biocon.2013.07.001

Marcot, B. G., Hohenlohe, P. A., Morey, S., Holmes, R., Molina, R., Turley, M. C., Huff, M. H., & Laurence, J. A. (2006a). Characterizing species at risk II: Using Bayesian belief networks as decision support tools to determine species conservation categories under the northwest forest plan. *Ecology and Society*, *11*. https://doi.org/10.5751/ES-01781-110212

Marcot, B. G., Steventon, J. D., Sutherland, G. D., & McCann, R. K. (2006b). Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research*, *36*(12), 3063–3074. https://doi.org/10.1139/x06-135

Marsh, C. J., Gavish, Y., Kunin, W. E., & Brummitt, N. A. (2019). Mind the gap: Can downscaling area of occupancy overcome sampling gaps when assessing IUCN red list status? *Diversity and Distributions*, *25*, 1832–1845.

Marsh, C. J., Louzada, J., Beiroz, W., & Ewers, R. M. (2013). Optimising bait for pitfall trapping of amazonian dung beetles (Coleoptera: Scarabaeinae). *PLoS One*, *8*(8), e73147. https://doi.org/10.1371/journal.pone.0073147

Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., Tyre, A. J., & Possingham, H. P. (2005). Zero tolerance ecology: Improving ecological inference by modelling the source of zero observations: Modelling excess zeros in ecology. *Ecology Letters*, *8*, 1235–1246. https://doi.org/10.1111/j.1461-0248.2005.00826.x

Maxwell, S. L., Fuller, R. A., Brooks, T. M., & Watson, J. E. M. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature News*, *536*(7615), 143. https://doi.org/10.1038/536143a

McGill, B. J., Etienne, R. S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., Dornelas, M., Enquist, B. J., Green, J. L., He, F., Hurlbert, A. H., Magurran, A. E., Marquet, P. A., Maurer, B. A., Ostling, A., Soykan, C. U., Ugland, K. I., & White, E. P. (2007). Species abundance distributions: Moving beyond single prediction theories to integration within an ecological framework. *Ecology Letters*, *10*(10), 995–1015. https://doi.org/10.1111/j.1461-0248.2007.01094.x

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*(3), 248–257. https://doi.org/10.1111/j.2041-210X.2010.00077.x

McKinney, M. L. (1997). How do rare species avoid extinction? A paleontological view. In W. E. Kunin, & K. J. Gaston (Eds.), *The biology of rarity: Causes and consequences of rare—Common differences* (pp. 110–129). Springer. https://doi.org/10.1007/978-94-011-5874-9_7

McRae, L., Deinet, S., & Freeman, R. (2017). The diversity-weighted living planet index: Controlling for taxonomic bias in a global biodiversity indicator. *PLoS One*, *12*(1), 1–20. https://doi.org/10.1371/journal.pone.0169156

Merow, C., Smith, M. J., & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Merow, C., Wilson, A. M., & Jetz, W. (2017). Integrating occurrence data and expert maps for improved species range predictions. *Global Ecology and Biogeography*, *26*(2), 243–258. https://doi.org/10.1111/geb.12539

Mesgaran, M. B., Cousens, R. D., & Webber, B. L. (2014). Here be dragons: A tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity and Distributions*, *20*(10), 1147–1159. https://doi.org/10.1111/ddi.12209

Milner-Gulland, E. J., & Rowcliffe, J. M. (2007). *Conservation and sustainable use: A handbook of techniques*. Oxford University Press.

Molenberghs, G., Verbeke, G., & Demétrio, C. G. B. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, *13*(4), 513–531. https://doi.org/10.1007/s10985-007-9064-y

Mordecai, R. S., Mattsson, B. J., Tzilkowski, C. J., & Cooper, R. J. (2011). Addressing challenges when studying mobile or episodic species: Hierarchical Bayes estimation of occupancy and use. *Journal of Applied Ecology*, *48*, 56–66. https://doi.org/10.1111/j.1365-2664.2010.01921.x

Morrison, W. R. III, Lohr, J. L., Duchen, P., Wilches, R., Trujillo, D., Mair, M., & Renner, S. S. (2009). The impact of taxonomic change on conservation: Does it kill, can it save, or is it just irrelevant? *Biological Conservation*, *142*(12), 3201–3206. https://doi.org/10.1016/j.biocon.2009.07.019

Nekola, J. C., Šizling, A. L., Boyer, A. G., & Storch, D. (2008). Artifactions in the log-transformation of species abundance distributions. *Folia Geobot*, *43*, 259–268. https://doi.org/10.1007/s12224-008-9020-y

Nichols, J. D., Bailey, L. L., O'Connell, A. F. Jr, Talancy, N. W., Campbell Grant, E. H., Gilbert, A. T., Annand, E. M., Husband, T. P., & Hines, J. E. (2008). Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology*, *45*, 1321–1329. https://doi.org/10.1111/j.1365-2664.2008.01509.x

Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., … Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, *89*(3), e01370. https://doi.org/10.1002/ecm.1370

O'Hara, R. B., & Kotze, D. J. (2014). Do not log-transform count data. *Methods in Ecology and Evolution*, 118–122. https://doi.org/10.1111/j.2041-210X.2010.00021.x

Ota, H. (2000). Current status of the threatened amphibians and reptiles of Japan. *Population Ecology*, *42*(1), 5–9. https://doi.org/10.1007/s101440050003

Ovaskainen, O., & Soininen, J. (2011). Making more out of sparse data: Hierarchical modeling of species communities. *Ecology*, *92*(2), 289–295. https://doi.org/10.1890/10-1251.1

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*(5), 561–576. https://doi.org/10.1111/ele.12757

Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C. M., Myers, C. E., & Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, *263*, 10–18. https://doi.org/10.1016/j.ecolmodel.2013.04.011

Pacifici, K., Simons, T. R., & Pollock, K. H. (2008). Effects of vegetation and background noise on the detection process in auditory avian point-count surveys. *The Auk*, *125*(3), 600–607. https://doi.org/10.1525/auk.2008.07078

Pavlacky, D. C., Blakesley, J. A., White, G. C., Hanni, D. J., & Lukacs, P. M. (2012). Hierarchical multi-scale occupancy estimation for monitoring wildlife populations. *The Journal of Wildlife Management*, *76*, 154–162. https://doi.org/10.1002/jwmg.245

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, *19*(1), 181–197. https://doi.org/10.1890/07-2153.1

Pollock, K. H., Nichols, J. D., Brownie, C., & Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs*, *107*, 3–97.

Pollock, L. J., Morris, W. K., & Vesk, P. A. (2012). The role of functional traits in species distributions revealed through a hierarchical model. *Ecography*, *35*(8), 716–725. https://doi.org/10.1111/j.1600-0587.2011.07085.x

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*(5), 397–406. https://doi.org/10.1111/2041-210X.12180

Pouteau, R., Meyer, J.-Y., Taputuarai, R., & Stoll, B. (2012). Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecological Informatics*, *9*, 37–46. https://doi.org/10.1016/j.ecoinf.2012.03.003

Preston, F. W. (1948). The commonness, and rarity, of species. *Ecology*, *29*(3), 254–283. https://doi.org/10.2307/1930989

Qiao, H., Feng, X., Escobar, L. E., Peterson, A. T., Soberón, J., Zhu, G., & Papeş, M. (2019). An evaluation of transferability of ecological niche models. *Ecography*, *42*(3), 521–534. https://doi.org/10.1111/ecog.03986

Rabinowitz, D. (1981). Seven forms of rarity. In H. Synge (Ed.), *The biological aspects of rare plant conservation* (pp. 205–217). Riley.

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography*, *41*(4), 629–643. https://doi.org/10.1111/jbi.12227

Raes, N., & ter Steege, H. (2007). A null-model for significance testing of presence-only species distribution models. *Ecography*, *30*(5), 727–736. https://doi.org/10.1111/j.2007.0906-7590.05041.x

Rands, M. R. W., Adams, W. M., Bennun, L., Butchart, S. H. M., Clements, A., Coomes, D., Entwistle, A., Hodge, I., Kapos, V., Scharlemann, J. P. W., Sutherland, W. J., & Vira, B. (2010). Biodiversity conservation: Challenges beyond 2010. *Science*, *329*(5997), 1298–1303. https://doi.org/10.1126/science.1189138

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. SAGE.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, *51*, 1450–1459. https://doi.org/10.1111/1365-2664.12306

Risely, K., Baillie, S., Eaton, M., Joys, A., Musgrove, A., Noble, D., Renwick, A., & Wright, L. (2010). *Breeding bird survey 2009* (p. 559). BTO.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Roberts, D. L., Taylor, L., & Joppa, L. N. (2016). Threatened or data deficient: Assessing the conservation status of poorly known species. *Diversity and Distributions*, *22*(5), 558–565. https://doi.org/10.1111/ddi.12418

Robinson, O. J., Ruiz-Gutierrez, V., & Fink, D. (2018). Correcting for bias in distribution modelling for rare species using citizen science data. *Diversity and Distributions*, *24*(4), 460–472. https://doi.org/10.1111/ddi.12698

Rocchini, D., Garzon-Lopez, C. X., Marcantonio, M., Amici, V., Bacaro, G., Bastin, L., Brummitt, N., Chiarucci, A., Foody, G. M., Hauffe, H. C., He, K. S., Ricotta, C., Rizzoli, A., & Rosà, R. (2017). Anticipating species distributions: Handling sampling effort bias under a Bayesian framework. *Science of the Total Environment*, *584–585*, 282–290. https://doi.org/10.1016/j.scitotenv.2016.12.038

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J. M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G., & Chiarucci, A. (2011). Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography: Earth and Environment*, *35*(2), 211–226. https://doi.org/10.1177/0309133311399491

Rosenstock, S. S., Anderson, D. R., Giesen, K. M., Leukering, T., & Carter, M. F. (2002). Landbird counting techniques: current practices and an alternative. *The Auk*, *119*, 46–53. https://doi.org/10.1093/auk/119.1.46

Rosner-Katz, H., McCune, J. L., & Bennett, J. R. (2020). Using stacked SDMs with accuracy and rarity weighting to optimize surveys for rare plant species. *Biodiversity and Conservation*, *29*(11), 3209–3225. https://doi.org/10.1007/s10531-020-02018-1

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, *60*, 108–115.

Royle, J. A., & Nichols, J. D. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology*, *84*(3), 777–790.

Saupe, E. E., Barve, V., Myers, C. E., Soberón, J., Barve, N., Hensz, C. M., Peterson, A. T., Owens, H. L., & Lira-Noriega, A. (2012). Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. *Ecological Modelling*, *237–238*, 11–22. https://doi.org/10.1016/j.ecolmodel.2012.04.001

Schüttler, E., Klenke, R., Galuppo, S., Castro, R. A., Bonacic, C., Laker, J., & Henle, K. (2017). Habitat use and sensitivity to fragmentation in America's smallest wildcat. *Mammalian Biology*, *86*(1), 1–8. https://doi.org/10.1016/j.mambio.2016.11.013

Schwartz, M. W., & Simberloff, D. (2001). Taxon size predicts rates of rarity in vascular plants. *Ecology Letters*, *4*, 464–469.

Simkins, A. T., Buchanan, G. M., Davies, R. G., & Donald, P. F. (2020). The implications for conservation of a major taxonomic revision of the world's birds. *Animal Conservation*, *23*(4), 345–352. https://doi.org/10.1111/acv.12545

Smith, C. S., Howes, A. L., Price, B., & McAlpine, C. A. (2007). Using a Bayesian belief network to predict suitable habitat of an endangered mammal—The Julia Creek dunnart (Sminthopsis douglasi). *Biological Conservation*, *139*(3), 333–347. https://doi.org/10.1016/j.biocon.2007.06.025

Southwood, T. R. E., & Henderson, P. A. (2009). *Ecological methods*. John Wiley & Sons.

Standley, L. A. (1992). Taxonomic issues in rare species protection. *Rhodora*, *94*(879), 218–242.

Steyer, K., Simon, O., Kraus, R. H. S., Haase, P., & Nowak, C. (2013). Hair trapping with valerian-treated lure sticks as a tool for genetic wildcat monitoring in low-density habitats. *European Journal of Wildlife Research*, *59*(1), 39–46. https://doi.org/10.1007/s10344-012-0644-0

Stohlgren, T. J., Jarnevich, C. S., Esaias, W. E., & Morisette, J. T. (2011). Bounding species distribution models. *Current Zoology*, *57*(5), 642–647. https://doi.org/10.1093/czoolo/57.5.642

Sullivan, B. L., Phillips, T., Dayer, A. A., Wood, C. L., Farnsworth, A., Iliff, M. J., Davies, I. J., Wiggins, A., Fink, D., Hochachka, W. M., Rodewald, A. D., Rosenberg, K. V., Bonney, R., & Kelling, S. (2017). Using open access observational data for conservation action: A case study for birds. *Biological Conservation*, *208*, 5–14. https://doi.org/10.1016/j.biocon.2016.04.031

ter Steege, H., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomão, R. P., Guevara, J. E., Phillips, O. L., Castilho, C. V., Magnusson, W. E., Molino, J.-F., Monteagudo, A., Núñez Vargas, P., Montero, J. C., Feldpausch, T. R., Coronado, E. N. H., Killeen, T. J., Mostacedo, B., Vasquez, R., Assis, R. L., … Silman, M. R. (2013). Hyperdominance in the Amazonian Tree Flora. *Science*, *342*(6156). https://doi.org/10.1126/science.1243092

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L., & Williams, S. E. (2004). Extinction risk from climate change. *Nature*, *427*(6970), 145–148. https://doi.org/10.1038/nature02121

Thompson, S. K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, *85*, 1050–1059. https://doi.org/10.1080/01621459.1990.10474975

Thompson, S. K. (2013a). Adaptive web sampling in ecology. *Statistical Methods and Applications*, *22*, 33–43. https://doi.org/10.1007/s10260-012-0222-3

Thompson, W. L. (2002). Towards reliable bird surveys: Accounting for individuals present but not detected. *The Auk*, *119*, 18–25. https://doi.org/10.1093/auk/119.1.18

Thompson, W. (2013b). *Sampling rare or elusive species: Concepts, designs, and techniques for estimating population parameters*. Island Press.

Thuiller, W., Guéguen, M., Renaud, J., Karger, D. N., & Zimmermann, N. E. (2019). Uncertainty in ensembles of global biodiversity scenarios. *Nature Communications*, *10*(1), 1–9. https://doi.org/10.1038/s41467-019-09519-w

Tobler, M. W., Kéry, M., Hui, F. K. C., Guillera-Arroita, G., Knaus, P., & Sattler, T. (2019). Joint species distribution models with species correlations and imperfect detection. *Ecology*, *100*(8), e02754. https://doi.org/10.1002/ecy.2754

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, *203*(3), 312–318. https://doi.org/10.1016/j.ecolmodel.2006.11.033

Van Echelpoel, W., Boets, P., Landuyt, D., Gobeyn, S., Everaert, G., Bennetsen, E., Mouton, A., & Goethals, P. L. M. (2015). Chapter 6 - Species distribution models for sustainable ecosystem management. In Y.-S. Park, S. Lek, C. Baehr, S. E. Jørgensen (Eds.), *Developments in environmental modelling* (pp. 115–134). Elsevier. https://www.sciencedirect.com/science/article/pii/B9780444635365000089

van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J., & Raes, N. (2016). Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, *39*(6), 542–552. https://doi.org/10.1111/ecog.01509

Varela, S., Anderson, R. P., García-Valdés, R., & Fernández-González, F. (2014). Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. *Ecography*, *37*(11), 1084–1091. https://doi.org/10.1111/j.1600-0587.2013.00441.x

Violle, C., Thuiller, W., Mouquet, N., Munoz, F., Kraft, N. J. B., Cadotte, M. W., Livingstone, S. W., & Mouillot, D. (2017). Functional rarity: The ecology of outliers. *Trends in Ecology & Evolution*, *32*(5), 356–367. https://doi.org/10.1016/j.tree.2017.02.002

Weber, M. M., Stevens, R. D., Diniz-Filho, J. A. F., & Grelle, C. E. V. (2017). Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? *A meta-analysis. Ecography*, *40*(7), 817–828. https://doi.org/10.1111/ecog.02125

Welsh, A. H., Cunningham, R. B., & Chambers, R. L. (2000). Methodology for estimating the abundance of rare animals: Seabird nesting on North East Herald Cay. *Biometrics*, *56*, 22–30. https://doi.org/10.1111/j.0006-341X.2000.00022.x

Wilcox, T. M., McKelvey, K. S., Young, M. K., Jane, S. F., Lowe, W. H., Whiteley, A. R., & Schwartz, M. K. (2013). Robust detection of rare species using environmental DNA: The importance of primer specificity. *PLoS One*, *8*, e59520. https://doi.org/10.1371/journal.pone.0059520

Williams, B. K., Nichols, J. D., & Conroy, M. J. (2002). *Analysis and management of animal populations*. Elsevier Science.

Williams, K. J., Belbin, L., Austin, M. P., Stein, J. L., & Ferrier, S. (2012). Which environmental variables should I use in my biodiversity model? *International Journal of Geographical Information Science*, *26*(11), 2009–2047. https://doi.org/10.1080/13658816.2012.698015

Willson, J. D., Winne, C. T., & Todd, B. D. (2011). Ecological and methodological factors affecting detectability and population estimation in elusive species. *The Journal of Wildlife Management*, *75*, 36–45. https://doi.org/10.1002/jwmg.15

Yoccoz, N. G., Nichols, J. D., & Boulinier, T. (2001). Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution*, *16*(8), 446–453. https://doi.org/10.1016/S0169-5347(01)02205-4

Zurell, D., Elith, J., & Schröder, B. (2012). Predicting to new environments: Tools for visualizing model behaviour and impacts on mapped distributions. *Diversity and Distributions*, *18*(6), 628–634. https://doi.org/10.1111/j.1472-4642.2012.00887.x

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. (2009). Zero-truncated and zero-inflated models for count data. Mixed effects models and extensions in ecology with R 261-293.

How to cite this article: Jeliazkov, A., Gavish, Y., Marsh, C. J., Geschke, J., Brummitt, N., Rocchini, D., Haase, P., Kunin, W. E., & Henle, K. (2022). Sampling and modelling rare species: Conceptual guidelines for the neglected majority. *Global Change Biology*, 00, 1–24. https://doi.org/10.1111/gcb.16114