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

Published Version:
Sampling and modelling rare species: Conceptual guidelines for the neglected majority / Jeliazkov, Alienor; Gavish, Yoni; Marsh, Charles J.; Geschke, Jonas; Brummitt, Neil; Rocchini, Duccio; Haase, Peter; Kunin, William E.; Henle, Klaus. - In: GLOBAL CHANGE BIOLOGY. - ISSN 1354-1013. - STAMPA. - 28:12(2022), pp. 3754-3777. [10.1111/gcb.16114]

Availability:
This version is available at: https://hdl.handle.net/11585/885525 since: 2022-11-18

Published:
DOI: http://doi.org/10.1111/gcb.16114

Title page

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

Journal target

Nature Ecology and Evolution – Perspective section

3,000-4,000 words long + 4-6 display items

➔ Our current version: ~4,513 words (excluding References) + 5 display items.

Authors

Alienor Jeliazkov[*$a,]	alienor.jeliazkov@gmail.com
Yoni Gavish[$b]	gavishyoni@gmail.com
Charles J. Marsh[c,l]	charliem2003@gmail.com
Jonas Geschke[d]	jonas.geschke@ips.unibe.ch
Klaus Henle[e]	klaus.henle@ufz.de
Neil Brummitt[f]	n.brummitt@nhm.ac.uk
Duccio Rocchini[g,h]	duccio.rocchini@unibo.it
Peter Haasei[i,j]	peter.haase@senckenberg.de
William E. Kunin[k]	w.e.kunin@leeds.ac.uk

Affiliations

[*] Corresponding author

[$] Co-first authors

[a] University of Paris-Saclay, INRAE, UR HYCAR, Antony, France

[b] School of Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK

[c] Department of Plant Sciences, University of Oxford, UK

[d] Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland
[e] UFZ – Helmholtz Centre for Environmental Research, Department of Conservation Biology, Permoserstrasse 15, 04318 Leipzig, Germany
[f] Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK
[g] Alma Mater Studiorum University of Bologna, Department of Biological, Geological and Environmental Sciences, via Irnerio 42, 40126, Bologna, Italy
[h] Czech University of Life Sciences Prague, Faculty of Environmental Sciences, Department of Applied Geoinformatics and Spatial Planning, Kamycka 129, Praha - Suchdol, 16500, Czech Republic
[i] Department of River Ecology and Conservation, Senckenberg Research Institute and Natural History Museum Frankfurt, Gelnhausen, Germany
[j] Faculty of Biology, University of Duisburg-Essen, Essen, Germany
[k] University of Leeds, Leeds, UK
[l] Department of Ecology and Evolution & Yale Center for Biodiversity and Global Change, Yale University, New Haven, CT, USA

Abstract

The majority of species are rare, yet they also are the most challenging to sample and model. Predicting the distribution of rare species using conventional species distribution models is challenging because rare species are hardly captured by most survey systems. When enough data is available, predictions are usually spatially biased toward locations where the species is most likely to occur, violating the assumptions of many modelling frameworks. Workflows to predict and eventually map rare species distribution implies important trade-offs between data quantity, quality, representativeness, and model complexity that need to be considered prior to survey and analysis. In this synthesis, we summarize 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 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 and hierarchical models that allow correcting for overdispersion, spatial and sampling biases.

Keywords
rare species, sampling, survey, species distribution modeling, occupancy, bias, detectability

Box 1. Glossary

- **Hierarchical Models (HM):** a.k.a. *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., Raudenbush and Bryk 2002, Gelman et al. 2007).
- **Mark-release-recapture (MRR):** Mark-release-recapture, also known as capture-mark-recapture, is a sampling technique that consists in capturing, marking and releasing individuals of a species in a first capture session. Thereafter, in a second capture session, the ratio of marked to unmarked specimens is taken to estimate the population size of the species. The process can extend to more than two sessions to produce estimates that are more precise (see e.g., Williams et al. 2002, Southwood and Henderson 2009).
- **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 (a.k.a. occurrence probability), (2) the proportion of area or sites occupied, which results from the realization of the former process.
- **Occupancy-Area Relationship (OAR):** a.k.a. "scale-area curve" or "range-area relationship" (Harte and Kinzig 1997, Kunin 1998); 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).
- **Patchiness:** The way habitat patches (and populations) are distributed through space. Habitat patches can be clumped (i.e., spatially aggregated according to regular patterns with many patches aggregated in 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 spatially unbiased 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 so that the sampling is 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):** generic term to refer to any niche model that allows predicting the current or future distribution of a species (using occurrence and/or abundance data) based on predictors (such as climate, land-use, etc.) and, possibly, scenarios (e.g., IPCC's climate change scenarios).

Introduction

Almost all international, national, and local conservation planning activities flag biodiversity as a crucial environmental property (e.g. Aichi Targets, Sustainable Development Goals; Griggs et al. 2013, Butchart et al. 2016)) to be protected from the deleterious effects of habitat loss, exploitation, pollution and climate change (Rands et al. 2010, Maxwell et al. 2016, IPBES 2019). However, part of biodiversity relies on species, most of which are rare at various scales (Rabinowitz 1981, Hartley and Kunin 2003, Fontaine et al. 2007, Henle et al. 2010). Several initiatives that aim halting 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). At the same time, in context of global change, rare species are especially prone to extinction (Kunin and Gaston 1993, McKinney 1997, Henle et al. 2004, Courchamp et al. 2006, Işik 2011). One way to assess extinction risk is to track the change in spatial distribution through time (Gärdenfors et al. 2001, Araújo et al. 2002, Thomas et al. 2004, Benito et al. 2009). Therefore, protecting species diversity directly implies protecting rare species and this aim requires understanding their distribution patterns.

Unfortunately, rarity causes considerable methodological difficulties in obtaining sufficient data from survey programs 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)* 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 a depressing loop, also called the 'rare-species modelling paradox' (Lomba et al. 2010): the majority of species that require the greatest protection also are the species we know the least about and the most difficult to model, due to data deficiency or since the data we have violate the basic assumptions of the models.

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 all combinations of the distributional range of a species, the distribution pattern of populations within the range and the local density of the species when present (**Figure 1a**). Whatever the measure used (e.g., range size, occupancy, abundance, relative cover, biomass), and the ecosystem or scale of the study: the community is likely to include a handful of common species, and a much longer 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 and Henderson 2003), is observed on a local to global scale, with correspondingly fine abundance (McGill et al. 2007) to range size frequency (Gaston 1998) data. With the ultimate goal of e.g. mapping a rare species' distribution range for protection purposes, each of the seven types of rarity implies different problems in accumulating data for modelling. For example, having two species A and B with similar prevalence that are both dispersed in their range 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, if applying a sampling that is oriented by *a priori* knowledge on where species A is present, species A is more likely to be encountered than species B. Consequently, the dataset of species A contains more presences than the dataset of species B. The type of rarity, the spatial distribution of samples, and the protocol used to sample each location thus all affect the characteristics of the data generated, and the types of model we can use to project the species’ distribution range.

We therefore face a conundrum in which, although rarity is ubiquitous, it is particularly challenging to account for, sample and model it at all scales. To address this challenge and help untangle the conundrum, we aim for each of Rabinowitz's categories of rarity:

1. To identify the main trade-offs involved in the search for both adequate and cost-effective sampling strategies when designing a survey program, and how these decisions affect the properties of the data,
2. 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 types of modelling frameworks suitable for producing distribution maps for different types of rarity ('how to model'). We synthesize our findings and briefly discuss some remaining challenges to be addressed with respect to sampling and modelling rare species.

Where to sample

When setting up a monitoring scheme, 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 designed to target a particular species allows studying its population efficiently, yet at the expense of producing a representative sample of the species distribution. This conflicts with the aim of covering the fundamental niche of a species, thus with the assumptions of many modelling frameworks. For those species whose distribution range is relatively wide and whose distribution pattern is dispersed (common species and Rarity category 4), a representative sample-set of the entire extent is more likely to provide the required occurrence data. Representative sampling has several positive properties. First, the data it generates is comparable among species, allowing cost-effective monitoring of multiple species. Second, even if the location of samples is not constant, the data remains comparable between years, allowing the tracking of temporal changes in distribution (if sampling intensity kept constant). Third, the data generated can easily fit into most modelling frameworks if enough data on the focal species is collected. This is usually done with a systematic sampling scheme on a grid, stratifying the sampling according to habitat or land covers (while ensuring proportional sampling in each strata), or by randomly selecting the locations of the samples (**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 basic random sample-set of the entire extent is unlikely to capture sufficient information. Most current monitoring schemes fail to capture the required information for the majority of species as most tend to be rare (Preston 1948, Magurran and Henderson 2003). For example, in the 2007 UK plants countryside survey, 591 locations with one km^2 each were chosen to be included using a stratified random design (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 one will need to encounter the species in enough locations to make credible estimates of abundance or distributional status and changes. Thus, to survey rare species, one needs to use methods that increase the probability of encounter beyond that expected at random. To do so, one may need to bias the sampling towards the species of interest.

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

Another directed, fruitful approach is to combine adaptive with SDM-guided sampling (e.g., Lin et al. 2014, Aizpurua et al. 2015, Chiffard et al. 2020) where one sampling session provides information to model and the following sessions allow adjusting the distribution of samples (Yoccoz et al. 2001, Thompson 2013a, 2013b). For example, a SDM performed on data that was sampled at a certain time can tag potentially unknown local populations for sampling in the next year (e.g., Lin et al. 2014). Once the area is sampled and the SDM parameters are updated, the SDM is re-run and new locations are targeted. Such a strategy may be very efficient at accumulating observations of rare species. However, it comes with the risk to estimate an overoptimistic trend of occupancy*, as the number of detected presences may increase with time while the actual distribution decreases (**Table 1**). Any form of adaptive sampling therefore needs considerable manipulation and/or reliable complementary information to be used in further species distribution modelling (Raes and ter Steege 2007, Phillips et al. 2009, Dorazio 2014, Hefley et al. 2014).

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. In addition, stratified schemes depend on the quality of the original information used to guide the stratification (e.g., habitat and land-use maps) that also 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 relate to different elements of the compromise between sampling efficiency and representativeness (**Figure 1b**). Overall, depending on the sampling strategy used, three main types of data may be generated, each having implications for modelling. Data can be either spatially representative (of the species range for potentially multiple species), spatially biased independent of the species, or spatially biased towards particular species.

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 detectability of the sampled species (**Table 2**; **Figure 1c**). Locally rare as well as elusive species (e.g. cryptic or trap-shy species) be they rare or common (Thompson 2013b), both pose specific challenges for achieving these goals of sampling. The probability of detecting a species that is present depends on a range of factors, such as habitat type, time of the day and year, population density and the methods employed to survey the species. Methods that target rare and elusive species and repeated sampling will reduce the probability of false absences and the latter may allow generating presence/absence data that account for detection probability.
There are multiple methods that increase the detectability of species. Some are just a function of effort (e.g., more pitfall traps or longer transects), while others are more directly related to the known ecology of the target species (**Table 2**; **Figure 1c**). The latter methods include, for example, baiting traps with materials that attract individuals (e.g. valerian-treated lure sticks for wildcat detection, 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), resorting to expert knowledge on the species' habitat preference and/or behavior to actively detect the species on site, or the use of detection dogs (Grimm & Klenke 2019, Grimm et al. 2019, Hollerbach et al. 2018).

There are several points to consider when applying methods targeted to rare species to increase detection probability. First, most of them increase the effort required or costs compared to simpler methods, especially when the sampling aims at detecting several rare species simultaneously rather than only one. Second, methods that increase detection probability increase it differently for different species, making the output less comparable between species unless the method is highly standardized. For example, baiting a trap with pheromones (or mating calls) of a specific species will attract more individuals from the focal species whereas baiting a trap with a food source utilized by many species (e.g. dung for dung beetles) may retain sufficient levels of comparability between species in a given site. However, recent advances in genetic monitoring tools, such as improved markers in eDNA detection of stream species (e.g. Jerde et al. 2019; Leese et al. 2020; Carraro et al. 2020), significantly increase the number of detected species including many rare species. This is particularly true when using water samples from rivers that often integrate over several kilometers in river length (e.g. Mächler et al. 2019; Altermatt et al. 2020). Third, highly standardized protocols are also essential for comparisons among sites, although some variability in detectability between sites is always likely to remain; for example, bird songs are less audible in leaved deciduous forests than in mixed pine forests (e.g. Pacifici et al. 2008) and ungulates less visible in dense vegetation habitats (e.g. Bukombe et al. 2016).

Some sampling methods allow generating presence/absence, and even abundance data, in sufficient quality and quantity to account for detection probability (based on 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 (e.g. Schüttler et al. 2017) (**Table 2**; **Figure 1c**). However, although these data greatly increase the spectrum of models that can be applied, they require high efforts, are rarely applicable except for a spatially limited area, and thus hardly available for rare species except perhaps for the ones clumped with high local density. They may only allow developing SDMs for small regions because of the efforts and costs involved. However, we will see in the next section that the combination of 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 extrapolation of rare species distribution across large spatial scales (e.g. Giraud et al. 2016; Bowler et al. 2019), with a certain bias towards the species.

How to model

As discussed above, choices on the spatial distribution of samples (where to model) eventually lead to three types of datasets, spatially representative, spatially biased independent from the species, or spatially biased towards given species' presences. From a modelling perspective, these results in a trade-off between the number of presences (i.e. zero inflated models when there are few presences) and the need to account for spatial auto-correlation in the data. Similarly, the sampling protocols in selected sampling locations (how to sample) affect the type and quality of inference we get from each location. From a modeling perspective, this affects the type of data we need to deal with, be it, presence-only, presence/absence, or presence/absence with detectability or estimates of abundances. Put together, depending on the type of rarity, and 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 allow pseudo-absences to be generated based on external, additional sources of 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, as they do not imply absences but rather samples of the available environment, where presences are compared against background locations that were unsampled (Philips et al 2009, Merow et al 2013). They do not produce probability of occurrence but relative occurrence rates (Guillera-Arroita et al 2015) and can be appropriate to rare species modelling if proper bias correction is applied (**Table 3**; **Figure 1d**).

In cases where presence/absence data are available, several developments in SDMs allow handling data overdispersion (e.g. negative-binomial and mixed effect models; Molenberghs et al. 2007, O'Hara and Kotze 2014, Harrison 2014), spatial-autocorrelation (e.g. Dormann et al. 2007; Marcer 2013), uncertainty in predictions (e.g., ensemble forecasting; Araújo and New 2007, Guisan et al. 2017, Thuiller et al. 2019), and biases due to sampling scales (Keil et al. 2013, Keil and Chase 2019). In this respect, hierarchical models (HM)* become especially helpful due to their flexibility. Indeed, HMs aim to describe, on the one hand, the true state of nature that is not or only partly observable, and, on the other hand, the measurement error (Kéry and Royle 2015). Consequently, HMs are highly valuable for rare species modeling in that they allow modeling both the process error (e.g. variations in the occurrence probability potentially due to variation in available resources), and the observation error (e.g. variations in the detection probability potentially due to the variability in observer's skills, or variations in the occurrence probability due to a poor choice in the habitat type to sample), which result in the so-called true vs. false absences, respectively (Zuur et al. 2009). For instance, multiscale hierarchical SDMs allow accounting for the fact that increasing the sampling extent increases the probability of detecting rare species (Rocchini et al. 2017). As such, HMs allow imperfect detectability to be considered in the modeling procedure (**Table 3**). Furthermore, by integrating prior knowledge, Bayesian Belief Networks allow explicitly decomposing causal pathways involved in the capture rate of species, including the respective influences of detection and occupancy while handling small or incomplete datasets (Ussitalo 2007). For instance, capture can be considered as dependent on detectability, which is influenced by date and trapping effort, and by occupancy, which may be influenced by suitability of local habitat conditions (Marcot et al. 2006). Such methods have already proved relevant for modelling species distribution (e.g. Van Echelpoel et al. 2015), and responses of rare and endangered species (e.g. Smith et al. 2007; Hamilton et al. 2015) (**Table 3**).

When abundance data from standardized survey or monitoring protocols are available, these can be of great interest to fit rare species distribution models and track distribution changes (e.g. Howard et al. 2014). However, because such protocols usually do not allow detecting most of the rare species, especially the clumped and low local density species (see ‘how to sample’ section), abundance-based SDMs are rarely accessible for rare species.

If marked data are available, species occurrence and distribution modeling can be done using classical site-occupancy models and the different methods developed under the field of mark-release-recapture* analyses (Pollock et al. 1990, MacKenzie et al. 20017) (**Table 3**). However, these data are usually hardly available over large spatial scales (see ‘how to sample’ section).

When unmarked occurrence data are available from spatio-temporally replicated measurements of presences/absences, and 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 (Royle and Nichols 2003, Kéry and Royle 2015) allows estimating occurrence probability and can accommodate detection heterogeneity (**Table 3**; **Figure 1d**). When unmarked abundance data are available, N-mixture models are a good solution to estimate both detectability and abundances and have also proved their usefulness in large-scale species distribution modelling (Jakob et al. 2014, Guélat and Kéry 2018, Kéry 2018) (**Table 3; Figure 1d**). When some 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 et Lindenmayer 2005). When data are zero-inflated, one can apply variants of Royle-Nichols model or N-mixture models that allow extra parameters and account for the overdispersion of the data. Several variants of N-mixture models have also been developed to address different situations related to spatial biases 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 assumptions allowing the application of such models are quite restrictive in the context of species distribution modelling and further simulation studies are needed to assess their performance on rare species when assumptions are not met. In addition, obtaining the abundance data needed by some of these models can be particularly costly and this approach is 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 sources of data are available (presences, presence/absence, abundance), recent works have shown that their combination within single modelling frameworks provides valuable insights into predicting species occupancy, abundance and/or distribution (**Table 3**). Even if available over restricted spatial extent, multiple sources of abundance data can be used in combination with other more extensive data such as occupancy surveys or opportunistic observations (e.g. Atlas or citizen science data). One can build HMs that 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 promising methods allow extrapolation and even comparison of rare species distribution across large spatial scales (e.g. Giraud et al. 2016; Bowler et al. 2019) and can potentially apply to all categories of rarity providing that relevant data sources are available and model is well built (**Figure 1d**).

To summarize, model choice will mainly depend on the nature of the data, their overdispersion, and the spatial biases involved. Moving from presence only to presence/absence up to abundance in **Figure 1d**, there is a change in the temporal comparability of SDMs, and thus in our ability to track distributional changes. In the top row, the output is relative likelihood, which is non-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 is constant across time and space, the resulting probability map is comparable for a given species over time. Finally, the population size row allows the separate estimation of detectability and probability of occurrence, which is comparable over time (and over species and space). This comparability is of high importance as it enables conservationists to assess change in the distribution of rare species and to detect any distribution shrinkage that could lead to revising/updating the species status and protection needs.

Conclusion and future perspectives

Protecting species diversity implies protecting rare species. However, surveying and modelling rarity also implies considerable methodological difficulties. In this paper, we have identified how the main decisions on sampling strategy condition the properties of the data, and how these properties in turn condition the range of appropriate modelling methods. An exhaustive list of the multiple ways by which rare species can be sampled and modelled is beyond the scope of this paper and has been done elsewhere (e.g., Kenkel et al. 1990, Milner-Gulland and Rowcliffe 2007, Thompson 2013b). Instead our focus is the neglected issue of how to identify the main trade-offs we face when modelling the distribution of rare species, the decision path linking the form of rarity with the sampling and modelling strategies, and to summarize the main strategies that account for these trade-offs. We provide some guidelines to optimize monitoring and modeling of rare species depending on the characteristics of their rarity and that ensure the consistency between sampling methods and modeling approaches – ensuring the link in these steps of the same endeavor (**Figure 1**).

Significant data on the occurrence of species is collected by numerous people, e.g. by citizen scientists (Chandler et al. 2017, Amano et al. 2016). Such data is highly valuable for monitoring biodiversity at different scales, but often biased and limited to specific areas. While there are ways to correct biases in citizen science data (Robinson et al. 2017, Bird et al. 2014), for monitoring "rarest" species (i.e. narrow distributional range, clumped population, low local density), systematic species-targeted sampling design may be preferred. Significant advances are also expected to emerge from advanced remote sensing techniques, genetic tools and the use of detection dogs. Such approaches have the potential to significantly increase the detection rate of rare species at comparatively low costs, with more or less bias towards the species. Above all, future research is still needed to integrate the type of rarity more systematically, how and where to sample with the selection and computational advances and the availability of appropriate models (**Figure 1**).

Considering most forms of rarity, our synthesis highlights the particular potential of HMs as a flexible tool to improve rarity modelling while accounting for spatial, observer, and species-specific biases. In particular, advances in zero-inflation modelling clearly have to be better integrated into rare species distribution modelling as both the conceptual and technical foundations of these approaches are relevant to 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, especially in the current context promoting open data, citizen science, and the rise of biodiversity synthesis science (**Figure 1**).

Other promising perspectives have recently emerged, such as functional rarity modelling (Violle et al. 2017, Carmona et al. 2017) and the use of co-occurring species information (a.k.a. the “neighbourly advice”, McInerny and Purves 2011) and of positive associations among rare species (Calatayud et al. 2019, Hines and Keil 2020) as potentially valuable information to model rarity distribution. Other model developments include harnessing information from other sources that either directly inform a species’ distribution at larger scales, such as incorporating expert-drawn range maps (Merow et al., 2017) or elevational ranges (Ellis-Soto et al., n.d.) 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 incorporating ancillary information such as trait (Ovaskainen et al. 2011; Pollock et al., 2012) or phylogenetic similarity (Ovaskainen et al., 2017), also are a favorable place of further developments for rare species modelling. Finally, machine-learning based methods, including non-parametric methods, and methods tolerant to unstructured data, have shown promises for modelling and mapping rarity with strong predictive ability (e.g. Pouteau et al. 2012, Robinson et al. 2018). Further research and sensitivity analyses are needed to assess the appropriateness of all these methods in the workflow of rarity sampling and modelling, depending on the rarity type of the species.

References

Aizpurua, O., J.-Y. Paquet, L. Brotons, and N. Titeux. 2015. Optimising long-term monitoring projects for species distribution modelling: how atlas data may help. Ecography 38:29–40.

Araújo, M. B., and M. New. 2007. Ensemble forecasting of species distributions. Trends in Ecology & Evolution 22:42–47.

Araújo, M. B., P. H. Williams, and R. J. Fuller. 2002. Dynamics of extinction and the selection of nature reserves. Proceedings of the Royal Society of London. Series B: Biological Sciences 269:1971–1980.

Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution.

Benito, B. M., M. M. Martínez-Ortega, L. M. Muñoz, J. Lorite, and J. Peñas. 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 18:2509–2520.

Brummitt, N.A., S.P. Bachman, J. Griffiths-Lee, M. Lutz, J.F. Moat, A. Farjon, J.S. Donaldson, C. Hilton-Taylor, T.R. Meagher, S. Albuquerque, E. Aletrari, K. Andrews, G. Atchison, E. Baloch, B. Barlozzini, A. Brunazzi, J. Carretero, M. Celesti, H. Chadburn, E. Cianfoni, C. Cockel, V. Coldwell, B. Concetti, S. Contu, V. Crook, P. Dyson, L. Gardiner, N. Ghanim, H. Greene, A. Groom, R. Harker, D. Hopkins, S. Khela, P. Lakeman-Fraser, H. Lindon, H. Lockwood, C. Loftus, D. Lombrici, L. Lopez-Poveda, J. Lyon, P. Malcolm-Tompkins,

K. McGregor, L. Moreno, L. Murray, K. Nazar, E. Power, M. Quiton Tuijtelaars, R. Salter, R. Segrott, H. Thacker, L.J. Thomas, S. Tingvoll, G. Watkinson, K. Wojtaszekova & E.M. Nic Lughadha, 2015. Green Plants in the Red: a baseline global assessment for the IUCN Sampled Red List Index for Plants. PLoS ONE 10(8): e0135152 doi:10.1371/journal.pone.0135152

Buckland, S.T., Rexstad, E.A., Marques, T.A., Oedekoven, C.S (2015). Distance Sampling: Methods and Applications. Springer

Butchart, S. H. M., M. D. Marco, and J. E. M. Watson. 2016. Formulating Smart Commitments on Biodiversity: Lessons from the Aichi Targets. Conservation Letters 9:457–468.

Calatayud, J., E. Andivia, A. Escudero, C. J. Melián, R. Bernardo-Madrid, M. Stoffel, C. Aponte, N. G. Medina, R. Molina-Venegas, X. Arnan, M. Rosvall, M. Neuman, J. A. Noriega, F. Alves-Martins, I. Draper, A. Luzuriaga, J. A. Ballesteros-Cánovas, C. Morales-Molino, P. Ferrandis, A. Herrero, L. Pataro, L. Juen, A. Cea, and J. Madrigal-González. 2019. Positive associations among rare species and their persistence in ecological assemblages. Nature Ecology & Evolution 4:40–45.

Carmona, C. P., F. de Bello, T. Sasaki, K. Uchida, and M. Pärtel. 2017. Towards a Common Toolbox for Rarity: A Response to Violle et al. Trends in Ecology & Evolution 32:889–891.

Chiffard, J., C. Marciau, N. G. Yoccoz, F. Mouillot, S. Duchateau, I. Nadeau, P. Fontanilles, and A. Besnard. 2020. Adaptive niche-based sampling to improve ability to find rare and elusive species: simulations and field tests. Methods in Ecology and Evolution n/a.

Courchamp, F., E. Angulo, P. Rivalan, R. J. Hall, L. Signoret, L. Bull, and Y. Meinard. 2006. Rarity Value and Species Extinction: The Anthropogenic Allee Effect. PLoS Biology 4:e415.

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

Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population. Journal of Animal Ecology 12:42–58.

Gärdenfors, U., C. Hilton-Taylor, G. M. Mace, and J. P. Rodríguez. 2001. The Application of IUCN Red List Criteria at Regional Levels. Conservation Biology 15:1206–1212.

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 353:219–230.

Gelman, A., P. in the D. of S. A. Gelman, and J. Hill. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Griggs, D., M. Stafford-Smith, O. Gaffney, J. Rockström, M. C. Öhman, P. Shyamsundar, W. Steffen, G. Glaser, N. Kanie, and I. Noble. 2013. Sustainable development goals for people and planet. Nature 495:305–307.

Guélat, J., and M. Kéry. 2018. Effects of spatial autocorrelation and imperfect detection on species distribution models. Methods in Ecology and Evolution 9:1614–1625.

Guisan, A., W. Thuiller, and N. E. Zimmermann. 2017. Habitat Suitability and Distribution Models: with Applications in R. Cambridge University Press.

Harrison, X. A. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ 2:e616.

Harte, J., and A. P. Kinzig. 1997. On the Implications of Species-Area Relationships for Endemism, Spatial Turnover, and Food Web Patterns. Oikos 80:417–427.

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

Hefley, T. J., D. M. Baasch, A. J. Tyre, and E. E. Blankenship. 2014. Correction of location errors for presence-only species distribution models. Methods in Ecology and Evolution 5:207–214.

Henle, K., K. F. Davies, M. Kleyer, C. Margules, and J. Settele. 2004. Predictors of Species Sensitivity to Fragmentation. Biodiversity & Conservation 13:207–251.

Henle, K., W. Kunin, O. Schweiger, D. S. Schmeller, V. Grobelnik, Y. Matsinos, J. Pantis, L. Penev, S. G. Potts, I. Ring, J. Similä, J. Tzanopoulos, S. van den Hove, M. Baguette, J. Clobert, L. Excoffier, E. Framstad, M. Grodzińska-Jurczak, S. Lengyel, P. Marty, A. Moilanen, E. Porcher, D. Storch, I. Steffan-Dewenter, M. T. Sykes, M. Zobel, and J. Settele. 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 19:187–193.

Hines, J., and P. Keil. 2020. Common competitors and rare friends. Nature Ecology & Evolution 4:8–9.

Hollerbach L, Heurich M, Reiners TE, Nowak C: Detection dogs allow for systematic non-invasive collection of DNA samples from Eurasian lynx. Mammalian Biology, 90: 42-46.

Howard, C., P. A. Stephens, J. W. Pearce-Higgins, R. D. Gregory, and S. G. Willis. 2014. Improving species distribution models: the value of data on abundance. Methods in Ecology and Evolution 5:506–513.

IPBES. 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). IPBES Secretariat, Bonn, Germany.

Işik, K. 2011. Rare and endemic species: why are they prone to extinction? TURKISH JOURNAL OF BOTANY 35:411–417.

Jakob, C., F. Ponce-Boutin, and A. Besnard. 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 78:540–549.

Joseph, L. N., S. A. Field, C. Wilcox, and H. P. Possingham. 2006. Presence–Absence versus Abundance Data for Monitoring Threatened Species. Conservation Biology 20:1679–1687.

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

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

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

Kenkel, N. C., P. Juhász-Nagy, and J. Podani. 1990. On sampling procedures in population and community ecology. Pages 195–207 *in* G. Grabherr, L. Mucina, M. B. Dale, and C. J. F. T. Braak, editors. Progress in theoretical vegetation science. Springer Netherlands.

Kéry, M. 2018. Identifiability in N-mixture models: a large-scale screening test with bird data. Ecology 99:281–288.

Kéry, M., and J. A. Royle. 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.

Kunin, W. E. 1998. Extrapolating Species Abundance Across Spatial Scales. Science 281:1513–1515.

Kunin, W. E., and K. J. Gaston. 1993. The biology of rarity: Patterns, causes and consequences. Trends in Ecology & Evolution 8:298–301.

Leese, F., M. Sander, D. Buchner, V. Elbrecht, P. Haase, Zizka, V. (2020) Improved freshwater macroinvertebrate detection from eDNA through minimized non-target amplification. Bio-Rxiv: https://doi.org/10.1101/2020.04.27.063545.

Lin, Y.-P., W.-C. Lin, Y.-C. Wang, W.-Y. Lien, T.-S. Ding, P.-F. Lee, T.-Y. Wu, A. Klenke, D. S. Schmeller, and K. Henle. 2014. An optimal spatial sampling approach for modelling the distribution of species. Page Scaling in Ecology and Biodiversity Conservation. Pensoft Publishers. Pensoft.

Liu, C., M. White, and G. Newell. 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of Biogeography 40:778–789.

MacKenzie DI, Royle JA (2005) Designing occupancy studies: general advice and allocating survey effort. Journal of Applied Ecology 42: 1105-1114

MacKenzie, D. I., J. Nichols, J. Royle, K. Pollock, L. Bailey, and J. Hines 2017. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Academic Press.

Magurran, A. E., S. R. Baillie, S. T. Buckland, J. McP. Dick, D. A. Elston, E. M. Scott, R. I. Smith, P. J. Somerfield, and A. D. Watt. 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends in Ecology & Evolution 25:574–582.

Magurran, A. E., and P. A. Henderson. 2003. Explaining the excess of rare species in natural species abundance distributions. Nature 422:714–716.

Maxwell, S. L., R. A. Fuller, T. M. Brooks, and J. E. M. Watson. 2016. Biodiversity: The ravages of guns, nets and bulldozers. Nature News 536:143.

McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, and E. P. White. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology Letters 10:995–1015.

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

McKinney, M. L. 1997. How do rare species avoid extinction? A paleontological view. Pages 110–129 *in* W. E. Kunin and K. J. Gaston, editors. The Biology of Rarity: Causes and consequences of rare—common differences. Springer Netherlands, Dordrecht.

McRae, L., S. Deinet, and R. Freeman. 2017. The Diversity-Weighted Living Planet Index: Controlling for Taxonomic Bias in a Global Biodiversity Indicator. PLoS ONE 12:1–20.

Milner-Gulland, E. J., and J. M. Rowcliffe. 2007. Conservation and Sustainable Use: A Handbook of Techniques. Oxford University Press.

Molenberghs, G., G. Verbeke, and C. G. B. Demétrio. 2007. An extended random-effects approach to modeling repeated, overdispersed count data. Lifetime Data Analysis 13:513–531.

O’Hara, R. B., and D. J. Kotze. 2014. Do not log-transform count data. Methods in Ecology and Evolution:118–122.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19:181–197.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical Inference for Capture-Recapture Experiments. Wildlife Monographs:3–97.

Pouteau, R., J.-Y. Meyer, R. Taputuarai, and B. Stoll. 2012. Support vector machines to map rare and endangered native plants in Pacific islands forests. Ecological Informatics 9:37–46.

Preston, F. W. 1948. The Commonness, And Rarity, of Species. Ecology 29:254–283.

Rabinowitz, D. 1981. Seven forms of rarity. Pages 205–217 *in* H. Synge, editor. The Biological Aspects of Rare Plant Conservation. Riley.

Raes, N., and H. ter Steege. 2007. A null-model for significance testing of presence-only species distribution models. Ecography 30:727–736.

Rands, M. R. W., W. M. Adams, L. Bennun, S. H. M. Butchart, A. Clements, D. Coomes, A. Entwistle, I. Hodge, V. Kapos, J. P. W. Scharlemann, W. J. Sutherland, and B. Vira. 2010. Biodiversity Conservation: Challenges Beyond 2010. Science 329:1298–1303.

Raudenbush, S. W., and A. S. Bryk. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. SAGE.

Roberts, D. L., L. Taylor, and L. N. Joppa. 2016. Threatened or Data Deficient: assessing the conservation status of poorly known species. Diversity and Distributions 22:558–565.

Robinson, O. J., V. Ruiz-Gutierrez, and D. Fink. 2018. Correcting for bias in distribution modelling for rare species using citizen science data. Diversity and Distributions 24:460–472.

Rocchini, D., C. X. Garzon-Lopez, M. Marcantonio, V. Amici, G. Bacaro, L. Bastin, N. Brummitt, A. Chiarucci, G. M. Foody, H. C. Hauffe, K. S. He, C. Ricotta, A. Rizzoli, and R. Rosà. 2017. Anticipating species distributions: Handling sampling effort bias under a Bayesian framework. Science of The Total Environment 584–585:282–290.

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology 84:777–790.

Schüttler, E., R. Klenke, S. Galuppo, R.A. Castro, C. Bonacic, J. Laker & K. Henle (2017): Habitat use and sensitivity to fragmentation in America's smallest wildcat. Mamm. Biol. 86: 1-8.

Southwood, T. R. E., and P. A. Henderson. 2009. Ecological Methods. John Wiley & Sons.

Steyer, K., O. Simon, R.H.S. Kraus, P. Haase & C. Nowak (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: 39–46.

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

Thompson, S. K. 2013a. Adaptive web sampling in ecology. Statistical Methods & Applications 22:33–43.

Thompson, W. 2013b. Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters. Island Press.

Thuiller, W., M. Guéguen, J. Renaud, D. N. Karger, and N. E. Zimmermann. 2019. Uncertainty in ensembles of global biodiversity scenarios. Nature Communications 10:1–9.

Violle, C., W. Thuiller, N. Mouquet, F. Munoz, N. J. B. Kraft, M. W. Cadotte, S. W. Livingstone, and D. Mouillot. 2017. Functional Rarity: The Ecology of Outliers. Trends in Ecology & Evolution 32:356–367.

Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. Trends in Ecology & Evolution 16:446–453.

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

## Other refs to integrate into Zotero

Cunningham, R.B. and Lindenmayer, D.B. (2005) Modeling count data of rare species: some statistical issues. *Ecology* 86, 1135–1142

Gaston, K. J. 1998. Species-range size distributions: products of speciation, extinction and transformation. Philosophical Transactions of the Royal Society B-Biological Sciences 353:219-230.

Guillera-Arroita, Gurutzeta, José J. Lahoz-Monfort, Jane Elith, Ascelin Gordon, Heini Kujala, Pia E. Lentini, Michael A. McCarthy, Reid Tingley, and Brendan A. Wintle. ‘Is My

Species Distribution Model Fit for Purpose? Matching Data and Models to Applications'. *Global Ecology and Biogeography* 24, no. 3 (March 2015): 276–92. https://doi.org/10.1111/geb.12268.

Joseph, L.N. *et al.* (2009) Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. *Ecological Applications* 19, 3

Kendall, W. (2010) The 'Robust Design'. Chapter 15. Program MARK: a gentle introduction. Eighth edition. Colorado State University, Fort Collins, USA.< http://www. phidot. org/software/mark/docs/book/>. Accessed 31.10.2017

Kéry, M. and 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.

MacKenzie, D.I. et al. (2017) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*, Elsevier.

Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A., Scott, E.M., et al. (2010). Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends in Ecology & Evolution, Special Issue: Long-term ecological research, 25, 574–582.

McGill, B. J., R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. L. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, and E. P. White. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecology letters 10:995-1015.

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, 1–20.

Merow, Cory, Matthew J. Smith, and John A. Silander. 'A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter'. *Ecography* 36, no. 10 (October 2013): 1058–69. https://doi.org/10.1111/j.1600-0587.2013.07872.x.

Nichols, J.D. *et al.* (2008) Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology* 45, 1321–1329

Pollock, K.H. *et al.* (1990) Statistical Inference for Capture-Recapture Experiments. *Wildlife Monographs*

Rabinowitz, D. 1981. Seven forms of rarity. Pages 205-217 *in* H. Synge, editor. The Biological Aspects of Rare Plant Conservation. John Wiley & Sons, Chichester.

Royle, J.A. (2004) N-Mixture Models for Estimating Population Size from Spatially

Replicated Counts. *Biometrics* 60, 108–115

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

Proosdij, André S. J. van, Marc S. M. Sosef, Jan J. Wieringa, and Niels Raes. 'Minimum Required Number of Specimen Records to Develop Accurate Species Distribution Models'. Ecography 39, no. 6 (June 2016): 542–52. https://doi.org/10.1111/ecog.01509.

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

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

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., Blanchet, F. G., 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

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

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.

Figures

Currently: 5 items = 1 figure + 3 tables + 1 box

Tables 1, 2, 3

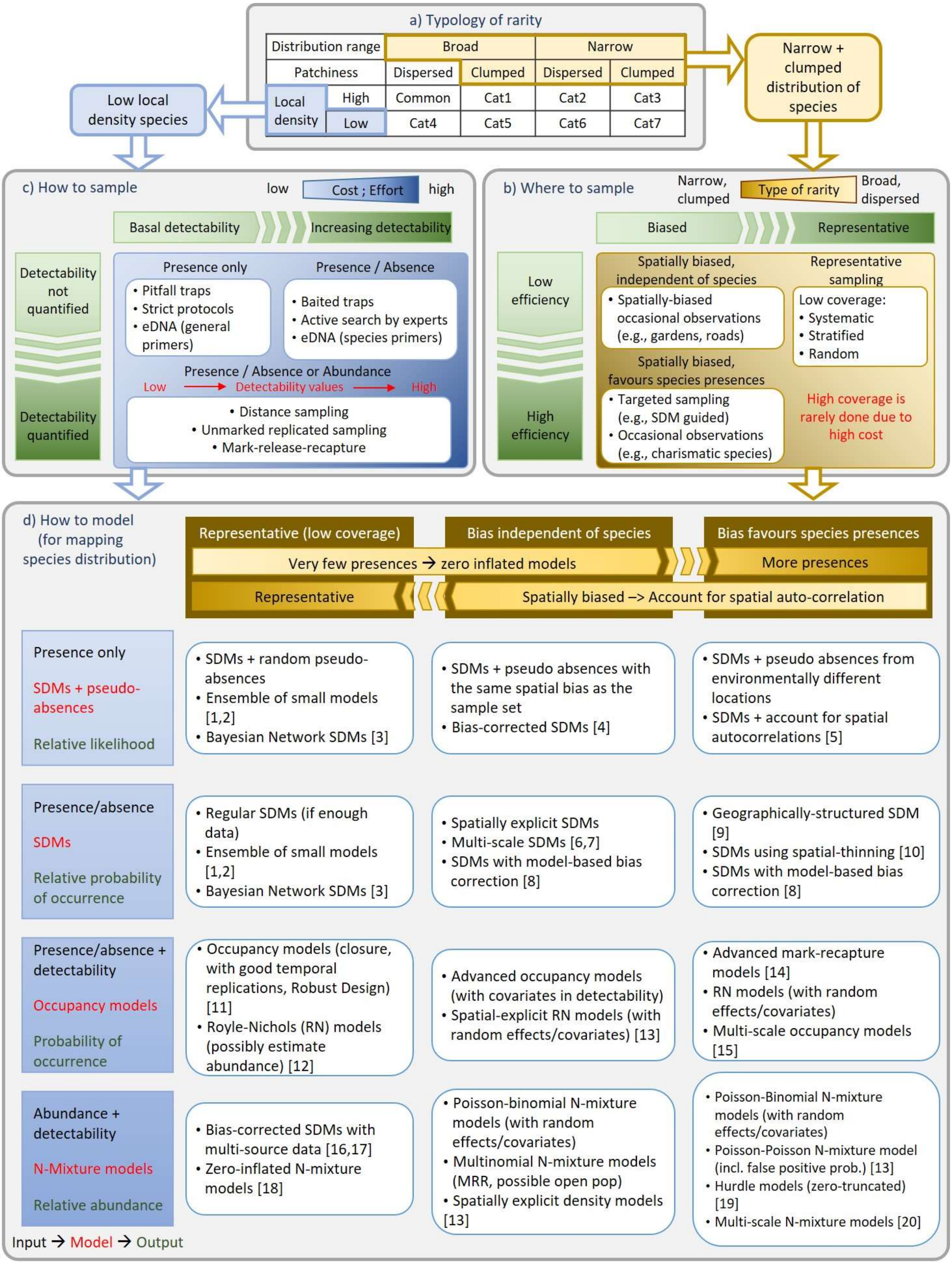
Non-exhaustive list of methods to assess (1) where to sample, (2) how to sample, and (3) how to model rare species data with their brief description, advantages and limits, the type of rarity for which they appear as most appropriate, and examples of references related. Inputs/outputs of modelling methods can be P (presences only), lik (presence likelihood), PA (Presences/Absences), ab (abundance), det (detectability information), pocc (probability of occurrence). Words with "*" refer to the Glossary (**Box 1**)

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.

References: [1] Breiner 2014, [2] Lomba 2010, [3] Chen 2012, [4] Fithian 2014, [5] Marcer 2013, [6] Keil 2013, [7] Rocchini et al. 2017, [8] El-Gabbas 2017, [9] Radosavljevic 2014, [10] Boria 2014, [11] McKenzie 2017, [12] Royle & Nichols 2003, [13] Kéry & Royle 2015, [14] Willson et al. 2011, [15] Nichols et al. 2008, [16] Giraud 2016, [17] Bowler et al. 2019, [18] Joseph et al. 2009, [19] Cunningham & Lindenmayer 2005, [20] Chandler et al. 2011

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

Basal 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 biased, independent of species
- Spatially-biased occasional observations (e.g., gardens, roads)

Spatially biased, favours species presences
- Targeted sampling (e.g., SDM guided)
- Occasional observations (e.g., charismatic species)

Representative sampling

Low coverage:
- Systematic
- Stratified
- Random

High coverage is rarely done due to high cost

d) How to model (for mapping species distribution)

Very few presences → zero inflated models → More presences

Representative ← Spatially biased –> Account for spatial auto-correlation

	Representative (low coverage)	Bias independent of species	Bias favours species presences
Presence only SDMs + pseudo-absences Relative likelihood	• SDMs + random pseudo-absences • Ensemble of small models [1,2] • Bayesian Network SDMs [3]	• SDMs + pseudo absences with the same spatial bias as the sample set • Bias-corrected SDMs [4]	• SDMs + pseudo absences from environmentally different locations • SDMs + account for spatial autocorrelations [5]
Presence/absence SDMs Relative probability of occurrence	• Regular SDMs (if enough data) • Ensemble of small models [1,2] • Bayesian Network SDMs [3]	• Spatially explicit SDMs • Multi-scale SDMs [6,7] • SDMs with model-based bias correction [8]	• Geographically-structured SDM [9] • SDMs using spatial-thinning [10] • SDMs with model-based bias correction [8]
Presence/absence + detectability Occupancy models Probability of occurrence	• Occupancy models (closure, with good temporal replications, Robust Design) [11] • Royle-Nichols (RN) models (possibly estimate abundance) [12]	• Advanced occupancy models (with covariates in detectability) • Spatial-explicit RN models (with random effects/covariates) [13]	• Advanced mark-recapture models [14] • RN models (with random effects/covariates) • Multi-scale occupancy models [15]
Abundance + detectability N-Mixture models Relative abundance	• Bias-corrected SDMs with multi-source data [16,17] • Zero-inflated N-mixture models [18]	• Poisson-binomial N-mixture models (with random effects/covariates) • Multinomial N-mixture models (MRR, possible open pop) • Spatially explicit density models [13]	• Poisson-Binomial N-mixture models (with random effects/covariates) • Poisson-Poisson N-mixture model (incl. false positive prob.) [13] • Hurdle models (zero-truncated) [19] • Multi-scale N-mixture models [20]

Input → Model → Output

Tables caption

Non-exhaustive list of methods to assess (Table 1) where to sample, (Table 2) how to sample, and (Table 3) 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. In Table 3: inputs/outputs can be P (presences only), lik (presence likelihood), PA (Presences/Absences), ab (abundance), det (detectability information), pocc (probability of occurrence). Words with "*" refer to the Glossary (Box 1). References are listed at the end.

Table 1. Where to sample?

Method	**Brief description**	**Pros**	**Cons**	**Suitable for which rarity categories?**	**References**

Accumulated opportunistic observations	Sampling locations are not chosen but emerge from 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
			- risk of misidentification in the case of non-expert observations (particularly critical as even a small fraction of miss-IDed 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	Greg-Smith 1964;
					Diekmann et al. 2007;
					Hedgren & Weslien 2008

Systematic sampling	Sampling according to a fixed spatial interval(s) that depends on the predefined total number of locations to be	- simple to implement, no need of external information nor a priori species-specific knowledge - more cost-efficient than simple random sampling as it guarantees	- needs prior information on 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	Cat4 (and Cat5 if habitats are organised randomly)	Madow 1953; Fortin et al. 1989

	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)	even distribution of sites and good coverage of the study area - temporally comparable samples - no target species, multi-species sample	the sampling, e.g. in clumped populations species, if 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)		
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 W.L. 2013

Adaptive (cluster) sampling / prior-informed sampling	Sampling design where the sites selection depends on previous sampling raw outcomes, either from the overall survey, e.g. adaptive *cluster* sampling consists in searching for a species in a given location and if the species is found, searches continue nearby (neighborhood shape can vary according to the study needs), or from other surveys, i.e. the sites 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 complexified
			- 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 et al. 1989;
					Yoccoz et al. 2001;
					Thompson S.K. 1990; 2013;
					Thompson W.L. 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. SDM profiling, adaptive niche-based sampling	- sampling coverage optimisation - allows a systematic and exhaustive pre-selection of suitable locations	- time-consuming process - requires predictor layers (with good spatial and thematical 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)	Le Lay et al. 2010; Aizpurua et al. 2015; Chiffard et al. 2020

Table 2. How to sample?

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 low local density	Cat1, Cat2, Cat3	Enquist et al. 2016 (BIEN);
					Bruelheide et al. 2019 (sPlot);
					Risely et al. 2010 (Britsh Trust for Ornithology);
					Jiguet et al. 2012 (French Breeding Bird Survey)

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 preence/absence data	- effort required is high unless detection probability is high - may require survey methods targeted to particular rare species, such as lures	All	MacKenzie & 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 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 & 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 extent - 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)	Schüttlera et al. 2016 (camera trapping) 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, 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)	England et al. 2005;
					Taberlet 2012;
					Bohmann et al. 2014;
					Rees et al. 2014;
					Jerde et al. 2011;
					Pilliod et al. 2014;
					Wilcox et al. 2013;
					Beng & Corlett 2020

Table 3. How to model?

Method		**Brief description**	**Pros**	**Cons**	**Suitable for which rarity categories ?**	**Examples / references**	**Input data -> Output calculated/estimated* ***
Data processing	Data processing	Different processing strategies can be applied on 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	- 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
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	- requires absence data - often too simplistic, resulting in strongly biased results - can suffer overfitting if the number of predictors is too high compared to	Common species	Guisan & Zimmermann 2000	PA -> relative pocc

				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			
	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	- 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
	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	- accurate - particularly appropriate and flexible for rare species modelling - hypothesis-driven	- interpretation sometimes difficult - hypothesis-driven - requires information on observational conditions	All	Dormann et al. 2007, Marcer 2013 (models accounting for spatial auto-correlation); Fithian 2014 (mixed effect models with an observer random effect); El-Gabbas 2017 (SDMs with model-based bias correction); Zuur et al. 2009 (zero-inflated models)	P -> relative lik PA -> relative pocc ab+det -> relative ab

		allow modelling true and false absences separately					
	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	- complicated fitting frameworks	Common species	Keil 2013 (hierarchical models incorporating distribution information at multiple grain sizes); Rocchini et al. 2017	PA -> relative pocc P -> relative lik
	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	- 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 & Anderson 2014; Philips et al. 2009	PA -> relative pocc P -> relative lik
	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	- assumes that habitat suitability is the most limiting driver of species distribution - reduces quantity of modelling data	Common species	Boria et al. 2014	PA -> relative pocc P -> relative lik

Ensemble of multiple SDMs	Ensemble SDMs	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 single best model - ensemble predictions perform better compared to single modelling techniques - can use variance between models as estimate of uncertainty	- all the cons of SDM approaches above - model averaging also has limitations (e.g. sensitivity to performance score and thresholds used)	Common species	Araújo & New 2006	PA -> relative pocc P -> relative lik
	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 low number of occurrences - allows structuring the modelling framework according to different scales of drivers of species distribution (e.g. local vs. climatic predictors)	- 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

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 to 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
				- 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 & Pollino 2012;
						MacCracken et al. 2012;
						Hamilton et al. 2015;
						Van Echelpoel et al. 2015
							P -> relative lik
							PA -> relative pocc
							ab -> relative ab

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.	- by aggregating data at large scales, overcomes sampling gaps (false abences in atlas data) and effects of sampling biases
			- no need for covariates
				- 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
					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)

				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			
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 fulfill 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	- 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

			as fulfilling the assumption of population closure				

	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 use 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 colonizations and extinctions, but simply presence or absence at the sample unit.	- accounts for the scale-dependence of occupancy estimation	- 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

N-mixture models	Royle-Nichols models (RN) or	HM* that estimate species occurrence	- provides two useful estimates :	- requires a sufficient amount of	All, especially	Royle & Nichols 2003; Kéry & Royle 2015	PA+det -> pocc

	Bernoulli-Poisson N-mixture models (for occurrences)	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 modeled data are identical. Can account for spatial autocorrelation using covariates as random or fixed effects	one for the detection probability and one for the occurrence probability	spatio-temporal replications in the data - requires good sets of predictors for both the detection and the occurrence parts of the model	for Cat4, Cat5, Cat6, Cat7 (low local density)		

	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) while 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-
			- provides two useful estimates : one for the detectability and one for the relative abundances
			- provides fine estimation of species relative abundances
			- with a sufficient amount of data and in some circumstances, some of these models can be used relaxing the population closure assumption
			- zero-inflated and hurdle models are particularly interesting for rare species (due to high risk of data overdispersion), quite intuitive to use and relatively easy to apply even in a likelihood framework
				- most of these models require good quality and 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, Denes et al. 2015 ("The N-mix" model);
						Kéry & Royle 2015 (Poisson-binomial/Poisson-Poisson/multinomial/density models);
						Cunningham & Lindemayer 2005, Fletcher et al. 2005, Zuur et al. 2009 (hurdle models);
						Chandler & Hepinstall-Cymerman 2016 (multiscale N-mixture models)
							ab+det -> relative ab

		explicit density models					

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 use 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)	- 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

References cited in the Tables

Aizpurua, O., J.-Y. Paquet, L. Brotons, and N. Titeux. 2015. Optimising long-term monitoring projects for species distribution modelling: how atlas data may help. Ecography 38:29–40.

Araújo, M. B., and M. New. 2006. Ensemble Forecasting of Species Distributions. Trends in Ecology & Evolution 22(1):42–47.

Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution.

Breiner, F. T., Guisan, A., Bergamini, A. & Nobis, M. P. Overcoming limitations of modelling rare species by using ensembles of small models. Methods in Ecology and Evolution 6, 1210–1218 (2015).

Chandler, R. B., J. A. Royle, and D. I. King. 2011. Inference about density and temporary emigration in unmarked populations. Ecology 92:1429–1435.

Chiffard, J., C. Marciau, N. G. Yoccoz, F. Mouillot, S. Duchateau, I. Nadeau, P. Fontanilles, and A. Besnard. 2020. Adaptive niche-based sampling to improve ability to find rare and elusive species: simulations and field tests. Methods in Ecology and Evolution n/a.

Cunningham, R. B., and D. B. Lindenmayer. 2005. Modeling count data of rare species: some statistical issues. Ecology 86:1135–1142.

Dénes, F. V., L. F. Silveira, and S. R. Beissinger. 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.

Dormann, F. C., J. M. McPherson, M. B. Araújo, R. Bivand, J. Bolliger, G. Carl, R. G. Davies, A. Hirzel, W. Jetz, W. Daniel Kissling, I. Kühn, R. Ohlemüller, P. R. Peres-Neto, B. Reineking, B. Schröder, F. M. Schurr, and R. Wilson. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30:609–628.

Fithian, W., J. Elith, T. Hastie, and D. A. Keith. 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. Methods in Ecology and Evolution 6:424–438.

Fletcher, D., D. MacKenzie, and E. Villouta. 2005. Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. Environmental and ecological statistics 12:45–54.

Fortin, M.-J., P. Drapeau, and P. Legendre. 1989. Spatial Autocorrelation and Sampling Design in Plant Ecology. Vegetatio 83:209–222.

Giraud, C., C. Calenge, C. Coron, and R. Julliard. 2016. Capitalizing on opportunistic data for monitoring relative abundances of species. Biometrics 72:649–658.

Guisan, A., and N. E. Zimmermann. 2000. Predictive Habitat Distribution Models in Ecology. Ecological Modelling 135(2):147–86

Jeliazkov, A., Y. Bas, C. Kerbiriou, J.-F. Julien, C. Penone, and I. Le Viol. 2016. Large-scale semi-automated acoustic monitoring allows to detect temporal decline of bush-crickets. Global Ecology and Conservation 6:208–218.

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

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

Kéry, M., and J. A. Royle. 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., and others. 1989. Ecological methodology. Harper & Row New York.

MacKenzie, D. I. 2006. Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Elsevier.

MacKenzie, D., J. Nichols, G. Lachman, S. Droege, J. Royle, and C. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248–2255.

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.

Marsh, C. J., Y. Gavish, W. E. Kunin, and N. A. Brummitt. 2019. Mind the Gap: Can Downscaling Area of Occupancy Overcome Sampling Gaps When Assessing IUCN Red List Status? Diversity and Distributions. 25:1832–45.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations: Modelling excess zeros in ecology. Ecology Letters 8:1235–1246.

Nekola, J. C., A. L. Šizling, A. G. Boyer, and D. Storch. 2008. Artifactions in the Log-Transformation of Species Abundance Distributions. Folia Geobotanica 43:259–268.

Nichols, J. D., L. L. Bailey, A. F. O’Connell Jr., N. W. Talancy, E. H. Campbell Grant, A. T. Gilbert, E. M. Annand, T. P. Husband, and J. E. Hines. 2008. Multi-scale occupancy estimation and modelling using multiple detection methods. Journal of Applied Ecology 45:1321–1329.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19:181–197.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical Inference for Capture-Recapture Experiments. Wildlife Monographs:3–97.
Radosavljevic, A., and R. P. Anderson. 2014. Making Better MAXENT Models of Species Distributions: Complexity, Overfitting and Evaluation. Journal of Biogeography 41(4):629–43.
Rocchini, D., C. X. Garzon-Lopez, M. Marcantonio, V. Amici, G. Bacaro, L. Bastin, N. Brummitt, A. Chiarucci, G. M. Foody, H. C. Hauffe, K. S. He, C. Ricotta, A. Rizzoli, and R. Rosà. 2017. Anticipating species distributions: Handling sampling effort bias under a Bayesian framework. Science of The Total Environment 584–585:282–290.

Rosenstock, S. S., D. R. Anderson, K. M. Giesen, T. Leukering, M. F. Carter, and F. Thompson III. 2002. Landbird counting techniques: current practices and an alternative. The Auk 119:46–53.

Ross, B. E., M. B. Hooten, and D. N. Koons. 2012. An Accessible Method for Implementing Hierarchical Models with Spatio-Temporal Abundance Data. PLOS ONE 7:e49395.

Royle, J. A. 2004. N-Mixture Models for Estimating Population Size from Spatially Replicated Counts. Biometrics 60:108–115.

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology 84:777–790.
Taberlet, P., E. Coissac, F. Pompanon, C. Brochmann, & E. Willerslev, 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21: 2045–2050.
Thompson, S. K. 1990. Adaptive Cluster Sampling. Journal of the American Statistical Association 85:1050–1059.

Thompson, W. 2013. Sampling Rare or Elusive Species: Concepts, Designs, and Techniques for Estimating Population Parameters. Island Press.

Thompson, W. L. 2002. Towards reliable bird surveys: accounting for individuals present but not detected. The Auk 119:18–25.

Welsh, A. H., R. B. Cunningham, and R. L. Chambers. 2000. Methodology for estimating the abundance of rare animals: seabird nesting on North East Herald Cay. Biometrics 56:22–30.

Willson, J. D., C. T. Winne, and B. D. Todd. 2011. Ecological and methodological factors affecting detectability and population estimation in elusive species. The Journal of Wildlife Management 75:36–45.

Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. Trends in Ecology & Evolution 16:446–453.

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