Automated assessments based on species distribution models can support regional Red Listing but need to be applied with caution: A case study from central Germany

Alexander Zizka[1]*, Indra Starke-Ottich[2], David Eichenberg[3], Dirk Bönsel[2], Georg Zizka[3,4]

1. Biodiversity of Plants Lab, Department of Biology, Philipps-University Marburg, Karl-von-Frisch-Straße 8, Marburg, Germany
2. Abteilung Botanik und Molekulare Evolutionsforschung, Senckenberg Forschungsinstitut und Naturmuseum Senckenberg, Frankfurt, Germany
3. Independent Researcher, Leipzig, Germany
4. Institut für Ökologie, Evolution und Diversität, Fachbereich Biowissenschaften, Goethe Universität Frankfurt

* Corresponding author: alexander.zizka@biologie.uni-marburg.de

ORCIDs

Alexander Zizka: 0000-0002-1680-9192
David Eichenberg: 0000-0001-5740-5621
Georg Zizka: 0000-0002-5030-7049

Abstract

Regional Red Lists (RL) are a cornerstone for biodiversity monitoring and conservation legislation in many countries, including Germany. Yet, the effort to keep RL up-to-date by regularly re-assessing species often is a substantial burden for regional authorities and volunteers. Automation methods using species distribution models have proven promising to speed up assessments on the global RL, but their value for regional Red Lists remains unclear. Here, we use the central German state of Hesse as a model to test how an automated RL assessment based on modelled species distribution ranges at three time slices compares to the latest expert-based RL for > 1,100 plant species. We find the resulting assessments of extinction risk, current population situation and population trends to agree in roughly 50% of the cases. While the model-based assessments were simplistic in some cases, they more more adequate in others. In particular, the assessment of moderately common species was a particular strength of the model-based approach. By examining wrongly assessed species in detail, we identify six themes in which automation may be particularly useful to support expert based regional red listing in the future, including: population trend data, cultivated habitat, taxonomically problematic groups, rare and under collected species, and the quantification of uncertainty in the assessments.

Key words

Conservation assessment, Distribution model, Expert-based, Hesse, Long-term monitoring, Time slices

Declarations

Funding

None

Competing interest

AZ is a member of the GfBS board. AZ develops methods for automated Red Listing, DB and ISO are authors of the latest Red List for Hesse.

Availability of Data and code

All data and analysis scripts will be available from a zenodo repository upon publication.

Authors' contributions

AZ and GZ conceived of this study. AZ, GZ, DE and ISO designed the analysis, AZ analyzed the data. ISO, DE, and DB provided data and expertise, AZ and GZ wrote the manuscript with contributions from all authors.

Introduction

Red Lists (RL) quantify species extinction risk using easily accessible threat categories, and are a cornerstone of conservation research, policy and practice (Bennun et al., 2018; Betts et al., 2020; Rodrigues et al., 2006). Internationally, RL are the basis for environmental legislation (e.g. CBD and CITES), and in Germany the national and state RL are major tools to monitor biodiversity in the framework of the National Strategy for Biodiversity. Yet, due to the work-intense assessment process, often done by volunteers, RL are taxonomically, temporally and geographically biased, which tilts conservation efforts towards charismatic taxa and well-studied regions (Cowie et al., 2022). Furthermore, keeping RL updated by repeated assessments every ten years is a major challenge (Rondinini et al., 2014), especially in the face of dwindling expert availability (Frobel & Schlumprecht, 2013).

RL exist for different taxa and on different geographic scales, all RL follow the same principle: Specialist(s) with expertise in the biota of a specific region assess the extinction risk of species following defined criteria, and, as a result, suggest a documented extinction risk classification. The specific assessment criteria and assessment rigour differ among RL and individual assessments, yet they are virtually always based on indicators related to range size, habitat availability, number of individuals, and population trends through time. While the RL of the International Union for the Conservation of Nature assesses the global extinction risk for c. 147,500 species across the Tree of Life (at the time of writing, www.iucnredlist.org), it is often the national and regional RL (which measure local extirpation) that drive specific conservation legislation and action (Azam et al., 2016; Jiang et al., 2020). The RL of Germany and the German state of Hesse classify species into one of seven extinction risk categories based on three indicators: current populations status, long-term population trend (app. 50-150 years) and short-term population trend (10-25 years; Starke-Ottich et al., 2019).

Many RL are incomplete and (at least partly) outdated (Cazalis et al., 2022; Cowie et al., 2022; www.iucnredlist.org), because the collection of data to estimate extinction risk requires substantial effort, specific skills, and expertise (IUCN Standards and Petitions Subcommittee, 2017). For instance, the quantification of population trends requires repeated and spatially comprehensive monitoring of individuals or populations over prolonged time-periods. As a result of these high requirements, charismatic and easy to collect taxa such as mammals, amphibians, birds, and, to a lesser extent flowering plants and specific insect groups are better assessed (Cowie et al., 2022). Furthermore, most expert-based assessments rely on distribution data and indicators of range size alone (Cazalis et al., 2022, www.iucnredlist.org). The lack of coverage and the resulting taxonomic bias is exacerbated by the need to repeat

red listing in regular time intervals to account for changes in species status in a changing world: new threats and the effectiveness of conservation measures need to be quantified, to keep legislation up to date and ensure effective species conservation (Rondinini et al., 2014), while at the same time the number of experts available is decreasing (Frobel & Schlumprecht, 2013; Starke-Ottich et al., 2019).

In the face of an accelerating global (Antonelli et al., 2020) and national biodiversity loss (Eichenberg et al., 2021; Habel et al., 2016; Jandt et al., 2022; Kamp et al., 2021), as well as the limited availability of experts for RL assessments, species distribution models (SDM) and automated RL assessment methods (AA) are two promising approaches to speed up RL assessments and overcome biases (Cazalis et al., 2022). SDM can be used to estimate species geographic ranges from known geo-referenced occurrence records and environmental variables (Franklin, 2010) and thereby can provide estimates of a species' probable range in the absence of complete sampling. SDM-based range estimates can then inform the RL assessment process, by providing estimates of current population status, and its change over time. SDM can vary in complexity; in a simple form (Area of Habitat) they are widely used in global RL assessments (IUCN Standards and Petitions Subcommittee, 2017). In addition to SDM, AA can be applied to batch estimate the extinction risk of many species. AA exist in varying complexity, from algorithms to batch estimate range size indices from species occurrence records (Bachman et al., 2011, 2020; Dauby et al., 2017) to artificial intelligence models using a broad range of data to estimate RL extinction risk (Walker et al., 2022; Zizka et al., 2021, 2022). SDM and AA may be applied individually or in combination to assist Red Listing. Yet, both have the caveat that they add complexity to the assessment process which may pose methodological hurdles for assessors (which are often taxonomic specialists) and can generate cumbersome black boxes difficult to document and interpret for RL. These reasons have so far limited the application of SDM and AA for regional RL, with few academic exceptions (Ralimanana et al., 2022; Schmidt et al., 2017).

Recently, modelled distribution ranges for more than 2,100 plant species in Germany, over three time slices have become available (Eichenberg et al., 2021). Here, we test how reliably an accessible automation algorithm can be used to harness this resource to support regional red listing. Specifically, we batch apply RL thresholds to the modelled distributions to estimate the current population status, the long-term and short-term population trends as well as the RL status of >1,100 plant species from the state of Hesse in central Germany. We ask the following questions:

1. How well does the automated assessment reproduce the expert-based regional Red List extinction risk assessments?

2. How well does the automated method approximate expert estimates of the individual indices composing the extinction risk assessment: species current population status, long-term population trend and short-term population trend?
3. For which species is the discrepancy between automated and expert-based RL particularly high? We anticipated that such species highlight potential strengths of either approach and thereby help to identify ways how automated assessments can be used to optimize expert RLs in the future.

Methods

Hesse is a state in central Germany with an area of c. 21,100 km²; mid-elevation mountains, and the floodplains of the Rhine and Main rivers are the main habitat types. We chose Hesse as model region for three reasons: (1) it has a long tradition of scientific floristic exploration (since the 1600s) and in formal Red Listing of flowering plants (since 1976) and many data are available on species populations trends and threat status (five version of the regional Red List have been published since 1976; Buttler et al., 1996; Hemm et al., 2008; Hessische Landesanstalt für Umwelt, 1976; Kalheber et al., 1980; Starke-Ottich et al., 2019); (2) the regional Red List of the state is compiled for four separate regions each with individual assessments of population trends, further increasing data availability; (3) up-to-date models of the distribution of the regional flora exist in high resolution (25km^2) and at three time slices (1960-1987, 1988-1996, 1997-2017; Eichenberg et al., 2021).

Expert conservation assessment - The Hessian Red List (HRL)

We based our comparison on the 5th version of the Hessian Red List (Starke-Ottich et al., 2019; HRL hereafter). The HRL followed the methodology provided by the German Federal Ministry for Nature Conservation for Red Listing (Ludwig et al., 2006) and was compiled using the methodology briefly summarized in the next paragraph.

For HRL, twelve expert assessors scored the extinction risk of each species known from Hesse into one of seven categories: extinct or lost (0), threatened with extinction (1), highly threatened (2), threatened (3), threat of unknown extent (G), extremely rare (R), near threatened (V), and data deficient (D). The assessments were based on three indicators: 1) current population status (extinct to very common) 2) short-term population trend (very strong decrease to strong increase), 3) long-term population trend (very strong decrease to strong increase; Tab. 1). Estimates on the individual indicators were based on expert judgement and were then combined into a Red List assessment for each species, using a standardized semi-automated pipeline. While the methodology (Ludwig et al., 2006) allowed the inclusion of

specific risk factors, these were not used for the HRL. The relevant time-span for the long-term population trend was app. 50-150 years. Assessments for this period were based on information from herbarium collections and literature. The availability of information differed between the four regions. For the short-term trend and the current population status, the recommended time span to consider was 10-25 years. Here, the assessment was based on the situation or its changes since the publication of the 4th RL in 2008 (Hemm et al., 2008). Population trend assessments for extinct species were not recorded in the assessment cycle of HRL v5, but instead translated from v4 (Hemm et al., 2008). The HRL included all native species known from Hesse, but neither neophytes nor established archeophytes (mostly plants from cultivation, e.g. in orchards or on fields). The extinction risk of species was assessed in four parts of Hesse separately (NW = Northwest, NO = Northeast, SW = Southwest, SO = Southeast,"regions" hereafter; Fig. 1; Buttler et al., 1996). The regions were divided following dominant landscape types in the state (Klausing, 1988), and therefore the regions were of different size; NO covered 52% of the state area while SO covered 9%. While regions NW, NO, and SO principally comprised mid-elevation mountain habitats, region SW mostly consisted of the lowlands of the Rhine and Main river floodplains.

Automated conservation assessment (ARL)

We generated an automated Red List for the Flora of Hessen (ARL) applying RL thresholds to modelled species distribution as follows.

First, we obtained modelled distributions for all vascular plant species in Germany from Eichenberg et al. (2021). This dataset provides individual occurrence probabilities for 2,136 plant species in a 5x5 km^2 grid at three time slices: 1960-1987, 1988-1996, 1997-2017 across Germany. The occurrence probabilities were modelled based on 29 million occurrence records from a variety of sources equally distributed among the three time-slices using spatiotemporal models based on environmental data and accounting for sampling bias. Hence, the dataset represents the probability of occurrence of each species in a respective grid-cell at a given time slice. The underlying method may be applied to any geographic area, taxon and time slice, given enough data are available. See Eichenberg et al. (2021) for details of the species distribution models.

Next, we calculated species range sizes in each time slice and their change among time slices based on the occurrence probabilities. To do so, we sampled 200 replicates of the occurrence of each species in each grid cell with an independent binomial draw from the occurrence probability of each grid cell in each time-slice. Next, we summarized the mean number of grid cells occupied by each species in each time-slice and calculated the proportion of grid cells

occupied by each species by dividing by the total number of cells in Hesse. Then we calculated changes in distribution through time and used them as proxies for population trend by using the distribution at time slice 1997-2017 to approximate the current population status, the difference between time slice 1960-1987 and 1997-2017 to approximate long term population trend and the difference between time slice 1988-1996 and 1997-2017 to approximate short term population trend. Finally, we converted the changes into RL categories using thresholds which we selected following guidelines by the German Federal Ministry for Nature Conservation (Ludwig et al., 2006; illustrated in Table 1) and the methodology of HRL (Starke-Ottich et al., 2019). Based on the individual indicators we classified species into extinction risk categories: as a combination of the categorical values estimated above following the conversion table provided by the German Ministry for Nature Conservation (Ludwig et al., 2006) into the same categories as used for HRL (0, 1, 2, 3, G, R, V, D). The conversion table gives resulting threat categories for each combination of criteria values, e.g. ss=v results in "threatened" (category 3) and sh<^ results in "not threatened" (category *).

Comparing Expert (HRL) and Automated Conservation Assessment (ARL)

We first harmonized the taxonomy since the data underlying the ARL used a different taxonomic backbone than HRL (HRL: Jäger, 2017; ARL: GermanSL; Jansen & Dengler, 2008 version 1.4, https://germansl.infinitenature.org/). Next, we removed species which for various reason were not comparable between HRL and ARL including (1) entries representing subspecies or varieties; (2) hybrids; (3) species which names we could not harmonize between HRL and ARL because one name in one list corresponded to multiple accepted names in the other list; (4) entries which belong to taxonomically difficult groups including, among others, apomictic species in *Rubus* and *Taraxacum*; (5) species in HRL with no modelled distribution available from Eichenberg et al. (2021); (6) species which the model predicted as present, but were never considered present in Hesse by the HRL; (7) species that were considered present in Hesse by HRL but never by ARL.

To test the agreement between HRL and ARL, we generated confusion matrices and calculated accuracy and per-class specificity, and sensitivity for the overall extinction risk category as well as the individual indicators. For calculation of accuracy, we excluded species classified as "?" in HRL or ARL.

To identify systematic reasons for mismatch between HRL and ARL related to current population status, we calculated confusion matrices for the Red List status per current population status as of HRL, and compared the mean error (the magnitude of mismatch, as the difference in ordinal classes, e. g. HRL: * and ARL: V amounts to an error of 1, HRL: * and

ARL: 1 to an error of 4) as well as overall accuracy among the distribution classes. Furthermore, we compared the raw modelled values for current population status with the categorized estimates from HRL.

To identify drivers of HRL ARL mismatch and identify how automated assessments may support regional red listing in the future, we qualitatively inspected the most prominent outliers in each category, and group them into categories based on species traits and expert knowledge. Furthermore, we provide and expert opinion (ISO and DB) if ARL or HRL provide a more comprehensive estimate of the current population status for 829 cases in which both approaches disagreed most severely.

Software used

We did all statistical analyses in R (R Core Team, 2022) and used the caret (Kuhn, 2022), cowplot (Wilke, 2020), plotly (Sievert, 2020), tidyverse v1.3.0 (Wickham et al., 2019), readxl v1.3.1 (Wickham & Bryan, 2022), viridis v0.5.1 (Garnier et al., 2021), and writexl (Ooms, 2021) packages.

Results

We obtained a modelled presence for 1,255 species in Hesse (as basis for the automated Red Listing, ARL; Fig 1A) and information for 1,933 taxa on the official Red List of Hesse (HRL, Fig 1B). During matching we removed 147 taxa (730 instances, since species may occur in up to four regions) because they represented subspecies or varieties; 39 species (195 entities) because they were hybrids; 38 species (190 entities) because we could not harmonize their names between HRL and ARL because one name in one list corresponded to multiple names in the other list; 236 species (1,180 instance) because they belonged to taxonomically difficult groups*;* 259 species (1,295 instances) because no modelled distribution was available; 296 species (457 instances) because, while the model predicted them as present, they were never considered present in Hesse by the HRL (species formerly present and now considered extinct-“0”-in HRL are included in the analyses); and 52 species (62 instances) because they were considered present by HRL but never by ARL. In the end we retained 1,139 species with 5,023 instances (Hesse 1,139 species, Region NO: 1,004 species, NW: 951 species, SO: 893 species SW: 1,037 species; Fig 1C) for the comparison of HRL and ARL, corresponding to 40.8% of the 2,790 taxa included in HRL (see Supplementary material S1 and S2 for a full list of matched species and S3 for a list of removed species).

Overall, the HRL considered 288 of the retained species as threatened (3: 103, 2:114, 1: 69) and 737 not threatened (*: 644, V: 93), 33 species as rare (R), 2 species as data deficient (?), and 79 species as extinct (ex). In contrast, ARL considered 415 species threatened (3: 136, 2: 116, 1: 163), 676 species not threatened (*: 585, V: 91), 44 species as rare, and 4 species as extinct for Hesse.

Considering the extinction risk categories, for Hessen, HRL and ARL agreed on 229 species as threatened and 618 as not threatened (Accuracy of ARL: 0.84). Concerning the detailed extinction risk category (Fig. 2), the overall accuracy of ARL was 0.59 with agreement on 643 species (of 1,135; 56,7%); the per class specificity of ARL was much higher than the sensitivity. For 226 spp. (20%), HRL differs from ARL towards higher extinction risk (difference of 1 category: 120 spp. (10,6%), 2: 63 (5,6%), 3: 26 (2,3%), 4: 10 (0,9%), 5: 6 (0,5%), and 6 categories: 1 sp. (0,09%)). ARL differs from HRL towards higher extinction risk in 266 spp. (23,4%) (difference of 1 category: 146 spp. (12,9%), 2: 77 (6,8%), 3: 27 (2,4%), 4: 11 (0,9%), 5: 5 (0,4%). Notably, ARL considered more than double the number of species in the highest threat category than HRL. Per class sensitivity varied between 0.26 and 0.79 and was much higher for the extreme classes (* = 0.79, 1 = 0.71) as compared to the intermediate classes (V: 0.24, 3:0.32, 2:0.26). The accuracy in RL status was similar among regions NO: 0.60, NW: 0.60, SO: 0.66, SW:0.63.

Considering individual indicators (which were only available on the region level not for the entire state), accuracy was highest for the long-term trend (0.64) followed by current population status (accuracy 0.47), and lowest for short-term trend (0.45). Per class specificity (current population status: 0.72-0.94, short-term trend: 0.73-0.94, long-term trend:0.60-1) was higher for all indicators than sensitivity (0.26-0.82, 0.09-0.55, 0-0.74; Fig. 3).

For the current population status in total, 3,848 cases (species x region) were included. ARL and HRL extinction risk assessments were identical for 1,420 cases. However, because ARL included the category “h” (abundant) for 729 cases, which was lacking in HRL, a detailed analysis was of little value. Lumping the “h” category with the categories “sh” or “mh” was not advisable, because “h” included records that can either be closer to “sh” or “mh”. When excluding the 729 records with assessment “h”, a total of 3,119 records remained, with 1,420 records of identical assessment (44,3%). In contrast, 170 records (5,5%) were assessed less abundant in ARL than in HRL (1 category difference in 150/4,8% of cases, 2: 16/0,5% and 3: 4/0,1%. The number of records that were assessed less abundant in HRL than in ARL is strikingly higher: 1,539 records (49,3%; 1 category difference in 1,226/39,3% cases, 2: 285/9,1%, 3: 14/0,4%, 4: 4/0,1%, Fig. 4).

The comparison of the short-term trend was based on 3,463 records. Among these, 16 records were assorted to the category “?” (insufficient information) in HRL and we therefore excluded them from the further analysis. Of the remaining 3,447 records, 1,723 received the same evaluation in HRL and ARL (nearly 50%, Fig. 5). In 1,253 records, the ARL suggested a more negative trend than the HRL (1 category difference in 865/25,1% cases, 2: 286/8,3%, 3: 100/2,9%, 4: 2/0,05%). We found the opposite–HRL more negative than ARL–in 471 cases (1 category difference in 404/11,7% of cases, 2: 57/1,65%, 3: 9/0,3%). The differences in assessment were particularly numerous in the HRL categories (V) and =, for which modelled assessment span from ∧ to VVV.

The comparison of the long-term trend included 3,463 cases (Fig. 5). Of those, we excluded 21 cases of category (<), because it was not available in the ARL. Of the remaining 3,442 cases, 2,237 (65%) were assessed identically by HRL and ARL, with 1,902 of them assorted to category “=” in both assessments. HRL was more optimistic in 655 records (19%; 1 category difference in 609/17,7% of the cases, 2: 43/1,2%, 3: 3/0,1%, while ARL was more optimistic in 550/17,7% of the cases; 1 category difference in 462/13,4%, 2: 75/2,2%, 3: 13/0,4% of the cases.

Discussion

Here we compared an accessible Automated Red Listing approach (ARL) based on species distribution models with a traditional expert-based Red List (HRL) for a well-documented area in central Europe. We found the RL extinction risk categories to be consistent in 57% of cases and the estimates for individual RL indices consistent in around half of the cases (current population status: 44,3%, short-term trend: 50%, long-term trend: 65%). In cases of mismatch the error was small in many cases (difference of 1 category in 23,5% of the instances) but could be considerable in some cases (up to 5 categories). We found the discrepancies between ARL and HRL to increase with species rarity, and to be particularly severe for species of azonal habitats, species of taxonomic difficult groups, cultivated species and species of cultivated landscapes, and species that are moderately common (categories “mh” and “h”).

The in-detail assessment of the severely mismatched species’ extinction risk and current populations status estimates suggests that ARL has the potential to assist and improve expert-based regional Red List assessments in the future in many cases (Supplementary material S4) and reveals routes forward for the improvement of ARL. We will discuss both issues in the

next paragraphs. The interactive graphs in Supplementary material S5-S11, allow to explore mismatches between HRL and ARL in detail.

Rare species and species with small-ranges. We found a tendency for ARL to both under- or overestimate extinction risk and current population status compared to HRL in species that are either rare (categories “s”, “ss”, and “es” in HRL) or have a small known range, often from forests or dry lawns. Examples from this group are *Vicia sylvatica* (HRL: *; ARL: 2), *Iris sibirica* (HRL: 2; ARL: *), *Moenchia erecta* (HRL: 2; ARL: *), and *Hepatica nobilis* (HRL: *, ARL: 2). There are many possible, likely interacting, reasons for mismatches in this group of species. A major reason may be the approximation population status and trend by modelled species distributions. While such a use of model species distributions is encouraged by HRL (Starke-Ottich et al., 2019), in some cases it may be overly simplistic without case-by-case judgement. Species may occur over a relatively large range but in low densities – or in limited areas but in large populations. Similarly, a constant geographic range through time may conceal a population decline within this range. While expert-based estimates in HRL are more likely to include true population trends for well known species, it should be noted, that the modelled distribution in ARL are based on habitat suitability and the resulting probabilities therefore can be expected to be proportional to density and may be more accurate for less well-known species. Furthermore, in many cases assessments in HRL will use a similar “range-size-for-population-size” proxy (without explicitly defining it) since informative population data through time is only available for few species. Additional likely causes for a mismatch in this category are: First, reduced availability of documented occurrence data. Documented occurrence records are often scarce for rare species (which by definition are found less frequently and may be protected and hence not be documented) and the performance of SDMs is inertly linked to the amount of data available for model training (Wisz et al., 2008). Hence, a lower performance of ARL is expected for rare species. Second, the modelled distributions in ARL were based on distribution data across Germany, so the range size of species rare in Hessen, but common in other parts of Germany may be overestimated. Furthermore, edge effects may cause the model to predict species as present in Hessen, that only occur in neighbouring states. Third, experts usually know rare species particularly well, increasing the discrepancy across ARL and HRL further. As a route forward we anticipate the adaption of more sophisticated predictive automated assessment methods available for the global RL to regional RL which can process incomplete information on population size and trends together with species functional traits will remove the need to use distribution range as proxy for population status and substantially improve the performance of ARL in the future (Cazalis et al., 2022). In addition, the expected increase in data availability of species occurrence data for rare species due to ongoing digitalization efforts and the spreading of data collection by citizen

scientists (for instance via smartphone apps such as Flora Incognita, iNaturalist, and Naturgucker) will continuously improve ARL performance for rare species. In general, we think that HRL is particularly accurate for rare species due to the experts focussed attention on them, so that we see the highest benefit of ARL for the rare to common species of the extinction risk categories *, V, 3, and 2.

Rare to common species. Although the agreement of extinction risk estimates between HRL and ARL decreased from common to rare species (Fig. 5) we found particularly relevant mismatches in the current distribution assessment for the intermediate categories of rare to common species (s, mh, h). Examples for this category are *Bistorta officinalis, Salvia pratensis, Populus alba* (HRL: ss, ARL: h), *Arum maculatum, Dipsacus fullonum, Echium vulgare, Primula veris* (HRL: s, ARL: h), *Digitaria ischaemum, Saponaria officinalis, Sedum album* (HRL: ss, ARL: mh). We hypothesize that a major reason is limited data availability for experts in HRL and a relatively strong performance of ARL in this category. Experts can usually assign very rare and very common species confidently to extinction risk categories and approximate trends in their population status. In contrast, recognizing trends in rare to common species is more challenging, since less information is available to experts, because they usually receive less attention (Starke-Ottich et al., 2019). Hence, these species may increase or decrease substantially before changes are picked up by experts, losing valuable time for conservation.

Species of azonal (small-scale) habitats. We found that HRL often considers species of specific azonal habitats as more threatened and as having a lower current population status than ARL, including most prominently species from wet meadows (following www.floraweb.de), bogs, flooding meadows, dry lawns. Examples in these categories are *Equisetum pratense* (HRL: 3, ARL: *), *Carex flava, Iris sibirica, Moenchia erecta* (HRL: 2, ARL: *), *Crepis pulchra* (HRL: R, ARL: *). There may be multiple non-mutually exclusive explanations for this mismatch. First, these habitats have received considerable conservation and research attention by specialists in the last years making HRL particularly precise and accurate for these habitats. Second, these habitats are often protected and, in some cases, hard to access, so that particularly few collection and observation records as basis for the modelled distributions exist. Third, projecting ranges for species of azonal, small-scale habitats is particularly challenging for species distribution models. Our modelling approach is comprehensive, as it includes 76 variables to measure ecological similarity among grid cells, but due to the 25 km^2 resolution may fail to capture microhabitat conditions and therefore systematic underestimate the occurrence probability of species of azonal habitats. In the future

higher resolution SDMs may improve the match between ARL and HRL, as could the relaxation of the neighbourhood criterion in the modelling algorithm.

Species in taxonomically difficult groups. We found ARL to often assess the current distribution of aggregate species more optimistically than HRL. Examples are *Alchemilla vulgaris* (NW: ARL: sh, HRL: s)*, Bistorta officinalis* (SW: ARL: h, HRL: ss)*, Carex flava* (SO: ARL: h, HRL: ss; SW: ARL: mh, HRL: es)*,* and *Galium mollugo* (SW: ARL: sh, HRL: ss). The reason is likely an unsatisfactory taxonomic matching between HRL and ARL. Although we have taken great care to resolve the taxonomies between the two datasets and have excluded particularly complex groups from the analyses, the correct assessment of these groups is an expected challenge. Another interesting example is *Nymphaea alba*, where a high percentage of recorded occurrences was of hybrid origin or even misidentified, which is correctly reflected in the HRL but not ARL (Starke-Ottich et al., 2019). As a route forward, further effort for taxonomic harmonisation before analyses may remedy this problem, yet it will likely persist, since a correct identification of species in taxonomic difficult groups cannot be guaranteed when using data from multiple sources including for instance citizen science, which is a challenge for ARL and HRL alike. Hence, automated assessment methods are currently best not used for groups that are taxonomically complex or particularly difficult to differentiate, and we urge expert assessors to judge carefully on these groups. This is also true for taxa below species levels (e.g. subspecies and varieties) which are often listed on regional RL, but where taxonomic definition and identification is difficult and the value for conservation in our eyes is sometimes challenging to identify.

Species related to cultivation or anthropogenic habitats. We found some of the most severe mismatches between HRL and ARL for cultivated species also occurring outside cultivation and species commonly occurring in anthropogenic habitats. Examples in this category are *Juglans regia*, *Setaria viridis*, *Setaria pumila*, and *Solanum nigrum,* for which ARL overestimates extinction risk. One reason for this systematic error was the low number of occurrence records included as basis for the modelling in ARL for these species, likely caused by the disregard of data collectors due to an ascribed low value or relevance for RL. These species are a good example of the value of expert judgement in the RL context. As a route forward information on cultivation or habitat could easily be excluded a priori from the simple ARL algorithm used here, or be incorporated in more sophisticated approaches. On the other hand, for species occurring naturally in agricultural fields or ruderal or urban habitats, HRL often underestimated current distribution range, e.g. in *Anchusa officinalis, Barbarea stricta, Bryonia dioica, Filago germanica, Bupleurum rotundifolium, Galeopsis ladanum, Geranium*

molle, Odontites vernus. An explanation could be, that especially urban habitats and their flora might receive less attention from experts in the course of Red list assessments.

Quantifying uncertainty. All Red Listing approaches rely on arbitrarily chosen thresholds, either to convert continuous modelled distributions (in the case of ARL) or abstract expert knowledge (in the case of HRL) into the distinct categories required for Red Lists. While we have carefully chosen the thresholds for ARL following the HRL guidelines (Ludwig et al., 2006), the guidelines are necessarily vague (since they apply to many taxa) and different thresholds would likely change the resulting assessments. Compared to HRL, ARL has the advantage, that the thresholds are clearly documented and therefore the assessments are reproducible and it is possible to quantify this uncertainty, for instance by integrating assessments over multiple thresholds. Additionally, the time intervals chosen were on the limit of the ranges acceptable for short term and long-term trends. There are several reasons for this, which affect HRL and ARL equally. First, less data is available further back in time. The recommendation as baseline for the long-term trend is 50-150 years (Ludwig et al., 2006). This is already conceptually challenging for ARL (where theoretically some, but considerably less data exist from this time as basis for modelling, and environmental variables used for modelling may have changed) and HRL (where experts may have some experience on the upper bound of this interval and general information from (often grey) literature may be accessible, see Händler, 2019 for an example). In practice, for ARL we had to rely on the upper limit of this time span, since the first time slice with modelled species distributions available was from 1960-1987 baseline for the long-term trend (on average 50 years ago at the time of the analyses). Furthermore, positive trends are included in a limited way in the HRL, since there is only one positive trend category for positive short- and long-term trends, and only one expert-assessor of HRL made use of positive trends. Despite, and partly due to these issues, we think that trend estimation is the second issue where ARL can significantly contribute to improving the red listing process by making best use of scarce data and providing standardized, reproducible parameter estimates that can help to ensure standardization of expert-based estimates.

Conclusion

Overall, we found the automated approach thresholding modelled species distributions to estimate species Red Lists categories and indicators simplistic and expert RL assessments more reliable, especially in the group of rare and endangered species, which is most important. The reasons are manifold and include, among others, the failure of the SDM to capture azonal habitats satisfactory, the difficulty to correctly represent taxonomically difficult groups in

particular species aggregates correctly, and the necessary arbitrary decisions in automatically applying Red List thresholds to the modelled distributions. In contrast, we identify two major strength of the automated approach in (1) providing a data-driven reference for rare to abundant species which are often difficult to estimate for experts, and (2) in offering options to quantify uncertainty. Therefore, we advocate the automated assessment as axuillary tool during the assessment process of these species to identify species warranting further expert attention. To facilitate such an approach for Hessen we here provide interactive graphs in the supplementary material, allowing to identify species with conflicting automated and expert based assessments easily. We are confident, that despite the demonstrated shortcomings of the simple automated approach the rapidly increasing data availability due to digitalization and citizen science-based data collection will substantially improve automated approaches in the near future. Our Study illustrates the potential of automated approaches to support expert-based regional Red Lists.

References

Antonelli, A., Fry, C., Smith, R. J., Simmonds, M. S. J., Kersey, P. J., Pritchard, H. W., Abbo, M. S., Acedo, C., Adams, J., Ainsworth, A. M., Allkin, B., Annecke, W., Bachman, S. P., Bacon, K., Bárrios, S., Barstow, C., Battison, A., Bell, E., Bensusan, K., … Zhang, B. G. (2020). *State of the World's Plants and Fungi 2020. Kew Gardens, available at https://doi.org/10.34885/172*. Royal Botanic Gardens, Kew. https://doi.org/10.34885/172

Azam, C., Gigot, G., Witte, I., & Schatz, B. (2016). National and subnational Red Lists in European and Mediterranean countries: Current state and use for conservation. *Endangered Species Research*, *30*, 255–266. https://doi.org/10.3354/esr00740

Bachman, S., Moat, J., Hill, A., Torre, J. de la, & Scott, B. (2011). Supporting Red List threat assessments with GeoCAT: Geospatial conservation assessment tool. *ZooKeys*, *150*, 117–126. https://doi.org/10.3897/zookeys.150.2109

Bachman, S., Walker, B., Barrios, S., Copeland, A., & Moat, J. (2020). Rapid Least Concern: Towards automating Red List assessments. *Biodiversity Data Journal*, *8*, e47018. https://doi.org/10.3897/BDJ.8.e47018

Bennun, L., Regan, E. C., Bird, J., van Bochove, J.-W., Katariya, V., Livingstone, S., Mitchell, R., Savy, C., Starkey, M., Temple, H., & Pilgrim, J. D. (2018). The Value of the IUCN Red List for Business Decision-Making. *Conservation Letters*, *11*(1), e12353. https://doi.org/10.1111/conl.12353

Betts, J., Young, R. P., Hilton-Taylor, C., Hoffmann, M., Rodríguez, J. P., Stuart, S. N., & Milner-Gulland, E. j. (2020). A framework for evaluating the impact of the IUCN Red List of threatened species. *Conservation Biology*, *34*(3), 632–643. https://doi.org/10.1111/cobi.13454

Buttler, K., Frede, A., Kubosch, R., Gregor, T., Hand, R., Cezanne, R., & Hodvina, S. (1996). *Rote Liste der Farn- und Samenpflanzen Hessens, 3. Fassung*. Hessisches Ministerium des Innern und für Landwirtschaft, Forsten und Naturschutz.

Cazalis, V., Di Marco, M., Butchart, S. H. M., Akçakaya, H. R., González-Suárez, M., Meyer, C., Clausnitzer, V., Böhm, M., Zizka, A., Cardoso, P., Schipper, A. M., Bachman, S. P., Young, B. E., Hoffmann, M., Benítez-López, A., Lucas, P. M., Pettorelli, N., Patoine, G., Pacifici, M., … Santini, L. (2022). Bridging the research-implementation gap in IUCN Red List assessments. *Trends in Ecology & Evolution*, *37*(4), 359–370. https://doi.org/10.1016/j.tree.2021.12.002

Cowie, R., Bouchet, P., & Fontaine, B. (2022). The Sixth Mass Extinction: Fact, fiction or speculation? *Biological reviews of the Cambridge Philosophical Society*, *97*, 0–000. https://doi.org/10.1111/brv.12816

Dauby, G., Stévart, T., Droissart, V., Cosiaux, A., Deblauwe, V., Simo-Droissart, M., Sosef, M. S. M., Lowry, P. P., Schatz, G. E., Gereau, R. E., & Couvreur, T. L. P. (2017). ConR: An R package to assist large-scale multispecies preliminary conservation assessments using distribution data. *Ecology and Evolution*, *7*(24), 11292–11303. https://doi.org/10.1002/ece3.3704

Eichenberg, D., Bowler, D. E., Bonn, A., Bruelheide, H., Grescho, V., Harter, D., Jandt, U., May, R., Winter, M., & Jansen, F. (2021). Widespread decline in Central European

plant diversity across six decades. *Global Change Biology*, *27*(5), 1097–1110. https://doi.org/10.1111/gcb.15447

Franklin, J. (2010). *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press. https://doi.org/10.1017/CBO9780511810602

Frobel, K., & Schlumprecht, H. (2013). Erosion der Artenkenner—Ergebnise einer Befragung und notwendige Reaktionen. *Naturschutz und Landschaftsplanung*, *48*(4), 105–113.

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro, A., Sciaini, Marco, Scherer, & Cédric. (2021). *viridis—Colorblind-Friendly Color Maps for R*. https://doi.org/10.5281/zenodo.4679424

Habel, J. C., Segerer, A., Ulrich, W., Torchyk, O., Weisser, W. W., & Schmitt, T. (2016). Butterfly community shifts over two centuries. *Conservation Biology*, *30*(4), 754–762. https://doi.org/10.1111/cobi.12656

Händler, A. (2019). *Die aktuelle Verbreitung und Bestandssituation von Pflanzen der Roten Liste Hessens 2008, im Vergleich mit den Angaben von Dillenius (1719), Walther (1802) sowie Heyer & Rossmann (1860-1863) in Gießen und Umgebung* [Master Thesis]. GEB-IDN/14834. http://geb.uni-giessen.de/geb/volltexte/2019/14834/

Hemm, K., Barth, U., Buttler, K., Frede, A., Kubasch, T., Gregor, T., Hand, R., Cezanne, R., Hodvina, S., Mahn, D., Nawrath, S., Huck, S., & Uebeler, M. (2008). *Rote Liste der Farn- und Samenpflanzen Hessens, 4. Fassung*. Hessisches Ministerium für Umwelt, ländlichen Raum und Verbraucherschutz.

Hessische Landesanstalt für Umwelt. (1976). *Rote Liste der bestandsgefährdeten Farn- und Blütenpflanzen Hessens (Stand: März 1976)*. Hessische Landesanstalt für Umwelt.

IUCN Standards and Petitions Subcommittee. (2017). *Guidelines for Using the IUCN Red List—Categories and Criteria. Version 13. Prepared by the Standards and Petitions Subcommittee. Http://www.iucnredlist.org/documents/RedListGuidelines.pdf* (S. 1–60).

Jäger, E. (Hrsg). (2017). *Rothmaler – Exkursionsflora von Deutschland, Gefäßpflanzen: Grundband, 21. Aufl.* Spektrum.

Jandt, U., Bruelheide, H., Jansen, F., Bonn, A., Grescho, V., Klenke, R. A., Sabatini, F. M., Bernhardt-Römermann, M., Blüml, V., Dengler, J., Diekmann, M., Doerfler, I., Döring, U., Dullinger, S., Haider, S., Heinken, T., Horchler, P., Kuhn, G., Lindner, M., … Wulf, M. (2022). More losses than gains during one century of plant biodiversity change in Germany. *Nature*, *611*(7936), Article 7936. https://doi.org/10.1038/s41586-022-05320-w

Jansen, F., & Dengler, J. (2008). GermanSL – Eine universelle taxonomische Referenzliste für Vegetationsdatenbanken. *Tuexenia*, 239–253.

Jiang, Z., Jiang, J., Wang, Y., Zhang, E., Zhang, Y., & Cai, B. (2020). Significance of country red lists of endangered species for biodiversity conservation. *Biodiversity Science*, *28*(5), 558. https://doi.org/10.17520/biods.2020149

Kalheber, H., Korneck, D., Müller, R., Nieschalk, A., Nieschalk, C., Sauer, H., & Seibig, A. (1980). *Rote Liste der in Hessen ausgestorbenen, verschollenen und gefährdeten Farn- und Blütenpflanzen. 2. Fassung, Stand 31. 12. 1979*. Hess. Landesanstalt für Umwelt.

Kamp, J., Frank, C., Trautmann, S., Busch, M., Dröschmeister, R., Flade, M., Gerlach, B., Karthäuser, J., Kunz, F., Mitschke, A., Schwarz, J., & Sudfeldt, C. (2021). Population trends of common breeding birds in Germany 1990–2018. *Journal of Ornithology*, *162*(1), 1–15. https://doi.org/10.1007/s10336-020-01830-4

Klausing, O. (1988). Die Naturräume Hessens mit einer Karte der naturräumlichen Gliederung 1:200 000. [2. Aufl.]. *Umweltplanung, Arbeits- Umweltschutz*, *67*, 1–43.

Kuhn, M. (2022). *caret: Classification and Regression Training*. https://CRAN.R-project.org/package=caret

Ludwig, G., Haupt, H., Gruttke, H., & Binot-Hafke, M. (2006). Methodische Anleitung zur Erstellung Roter Listen gefährdeter Tiere, Pflanzen und Pilze. *BfN Skripten*, *191*, 1–97.

Ooms, J. (2021). *writexl: Export Data Frames to Excel „xlsx“ Format*. https://CRAN.R-project.org/package=writexl

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ralimanana, H., Perrigo, A. L., Smith, R. J., Borrell, J. S., Faurby, S., Rajaonah, M. T., Randriamboavonjy, T., Vorontsova, M. S., Cooke, R. S. C., Phelps, L. N., Sayol, F., Andela, N., Andermann, T., Andriamanohera, A. M., Andriambololonera, S., Bachman, S. P., Bacon, C. D., Baker, W. J., Belluardo, F., … Antonelli, A. (2022). Madagascar's extraordinary biodiversity: Threats and opportunities. *Science*, *378*(6623), eadf1466. https://doi.org/10.1126/science.adf1466

Rodrigues, A. S. L., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M., & Brooks, T. M. (2006). The value of the IUCN Red List for conservation. *Trends in Ecology & Evolution*, *21*(2), 71–76. https://doi.org/10.1016/j.tree.2005.10.010

Rondinini, C., Marco, M. D., Visconti, P., Butchart, S. H. M., & Boitani, L. (2014). Update or Outdate: Long-Term Viability of the IUCN Red List. *Conservation Letters*, *7*(2), 126–130. https://doi.org/10.1111/conl.12040

Schmidt, M., Zizka, A., Traoré, S., Ataholo, M., Chatelain, C., Daget, P., Dressler, S., Hahn, K., Kirchmair, I., Krohmer, J., Mbayngone, E., Müller, J. V., Nacoulma, B., Ouédraogo, A., Ouédraogo, O., Sambaré, O., Schuman, K., Wieringa, J. J., Zizka, G., & Thiombiano, A. (2017). Diversity, distribution and preliminary conservation status of the flora of Burkina Faso. *Phytotaxa Monographs*, *304*(1), 1–215.

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC. https://plotly-r.com

Starke-Ottich, I., Gregor, T., Barth, U., Blatt, H., Böger, K., Bönsel, D., Cezanne, R., Frede, A., Gottschlich, G., Hemm, K., Hodvina, S., Jansen, W., Kubosch, R., Mahn, D., & Uebeler, M. (2019). *Rote Liste der Farn- und Samenpflanzen Hessens—5. Fassung*. Hessisches Landesamt für Naturschutz, Umwelt und Gerologie. https://www.hlnug.de/fileadmin/dokumente/naturschutz/Rote_Listen/RL_HE_Farn-_und_Samenpflanzen_Fassung5_2019.pdf

Walker, B. E., Leão, T. C. C., Bachman, S. P., Lucas, E., & Nic Lughadha, E. (2022). Evidence-based guidelines for automated conservation assessments of plant species. *Conservation Biology*, *37*(1), e13992. https://doi.org/10.1111/cobi.13992

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2022). *readxl: Read Excel Files*. https://CRAN.R-project.org/package=readxl

Wilke, C. O. (2020). *cowplot: Streamlined Plot Theme and Plot Annotations for „ggplot2“*. https://CRAN.R-project.org/package=cowplot

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., Guisan, A., & Group, N. P. S. D. W. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, *14*(5), 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x

Zizka, A., Andermann, T., & Silvestro, D. (2022). IUCNN – Deep learning approaches to approximate species' extinction risk. *Diversity and Distributions*, *28*(2), 227–241. https://doi.org/10.1111/ddi.13450

Zizka, A., Silvestro, D., Vitt, P., & Knight, T. M. (2021). Automated conservation assessment of the orchid family with deep learning. *Conservation Biology*, *35*(3), 897–908. https://doi.org/10.1111/cobi.13616

Tables

Table. 1. Categories of current population status, long-term trend and short-term trend in the Hessian Red List HRL (Ludwig et al., 2006) and used in the automated assessment in this study (ARL). Thresholds indicate the percentage of cells in Hesse where a species was modelled as present (current population status) or the percent change in cells with modelled presence in 1997-2017 compared to the reference time interval (1988-1996 for the short-term trend and 1960-1987 for the long-term trend).

Current population status			**Long-term trend**			**Short-term trend**		
Sym	**Category**	**Threshold used in ARL**	**Sym**	**Category**	**Threshold used in ARL**	**Sym**	**Category**	**Threshold used in ARL**
Ex	Extinct	0	<<<	Very strong decrease	-100- -80%	vvv	Very strong decrease	-100 - -35%
Es	Extremely rare	0-1%	<<	Strong decrease	-80 - -55%	vv	Strong decrease	-35 - -25%
Ss	Very rare	1-5%	<	Moderate decrease	-55 - -30%			-
S	Rare	5-15%	(<)	Decrease of unknown mangitude	-	(v)	Moderate decrease, or decrease of unknown magnitude	-25 - -10%
Mh	Moderately common	15-35%	=	Stagnant	-30 - +30%	=	Stagnant	-10% - +10%
^h	Common	35-70%	-	-	-	-	-	-
Sh	Very common	75-100%	>	Distinct increase	> +30%	^	Distinct increase	> +10%
?	Unknown	-	?	Data deficient	-	?	Data deficient	-

Figures

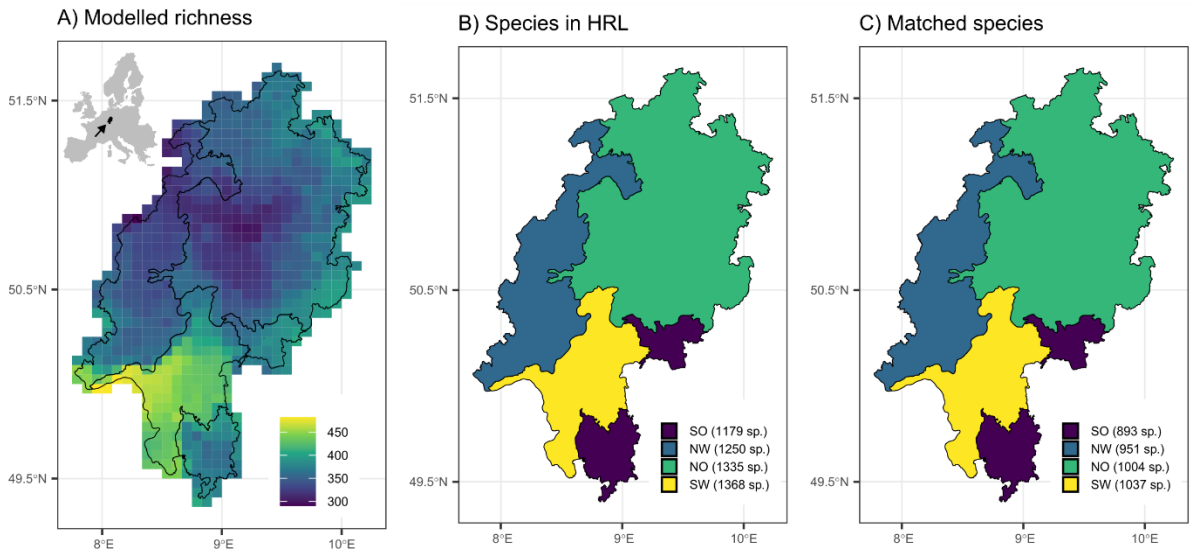

Figure 1. Map of the study area the state of Hesse in central Germany. **A)** Modelled richness of all species with distribution models available, as the cumulative occurrence probability in each 25 km^2 grid cell. **B)** The four regions used for separate Red List assessments in Hesse, colours indicate the species number in each region on the Hessian Red List (HRL). **C)** The number of species matched between modelled assessment (ARL) and Hessian Red List (HRL) in each region.

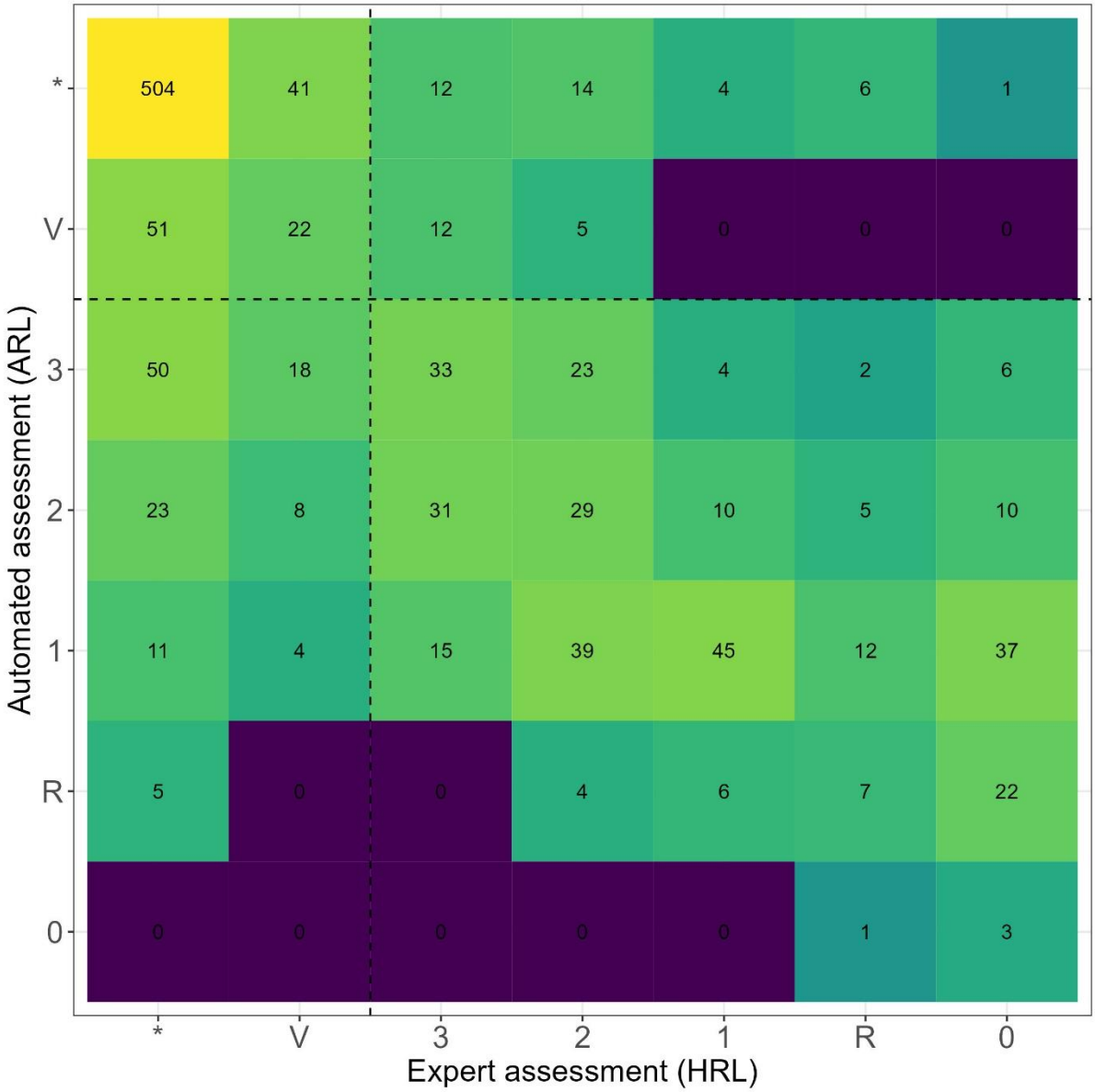

Figure 2. Confusion matrix comparing the extinction risk estimates from an automated, model-based assessment with the official expert-based Red List of Hesse for 1,135 species in Hesse. Numbers and colours indicate the number of species in each cell. The models agree for most not threatened (*) species and most critically endangered species (1), but tended to differ for intermediate categories (V,3,2). Supplementary material S5 contains an interactive version of this figure indicating the species names per tile when hovering over each tile.

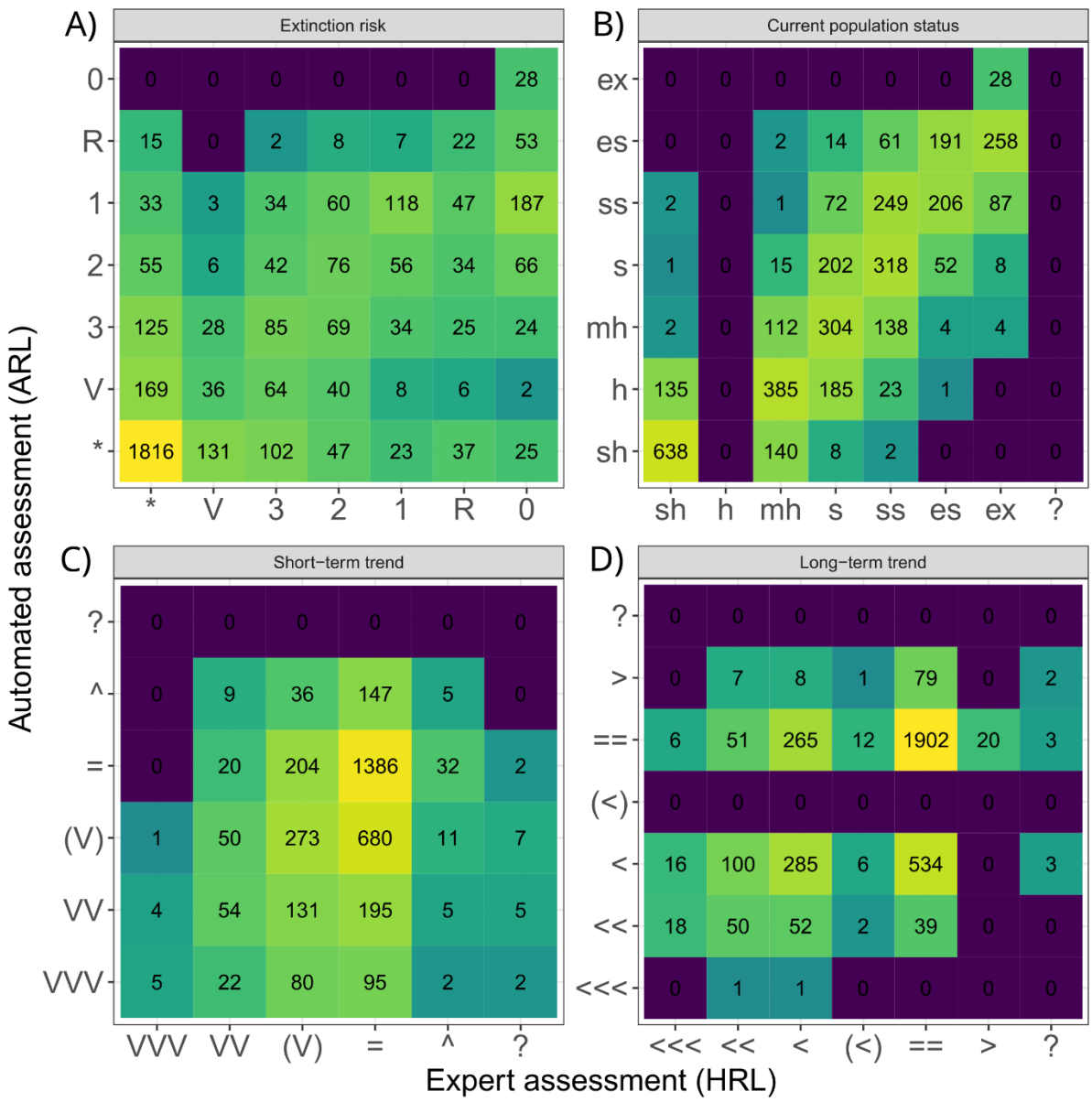

Figure 3. The agreement between a model-based Red list assessment (ARL) and the Hessian Red list (HRL) across the four regions of the study area. **A)** Extinction risk status, **B)** Current population status **C)** Short-term trend **D)** Long-term trend. Species may be counted multiple times, if they occur in multiple of the RL regions in Hesse. Supplementary material S6-S8 contain interactive version of this figure indicating the species names and region per grid-cell.

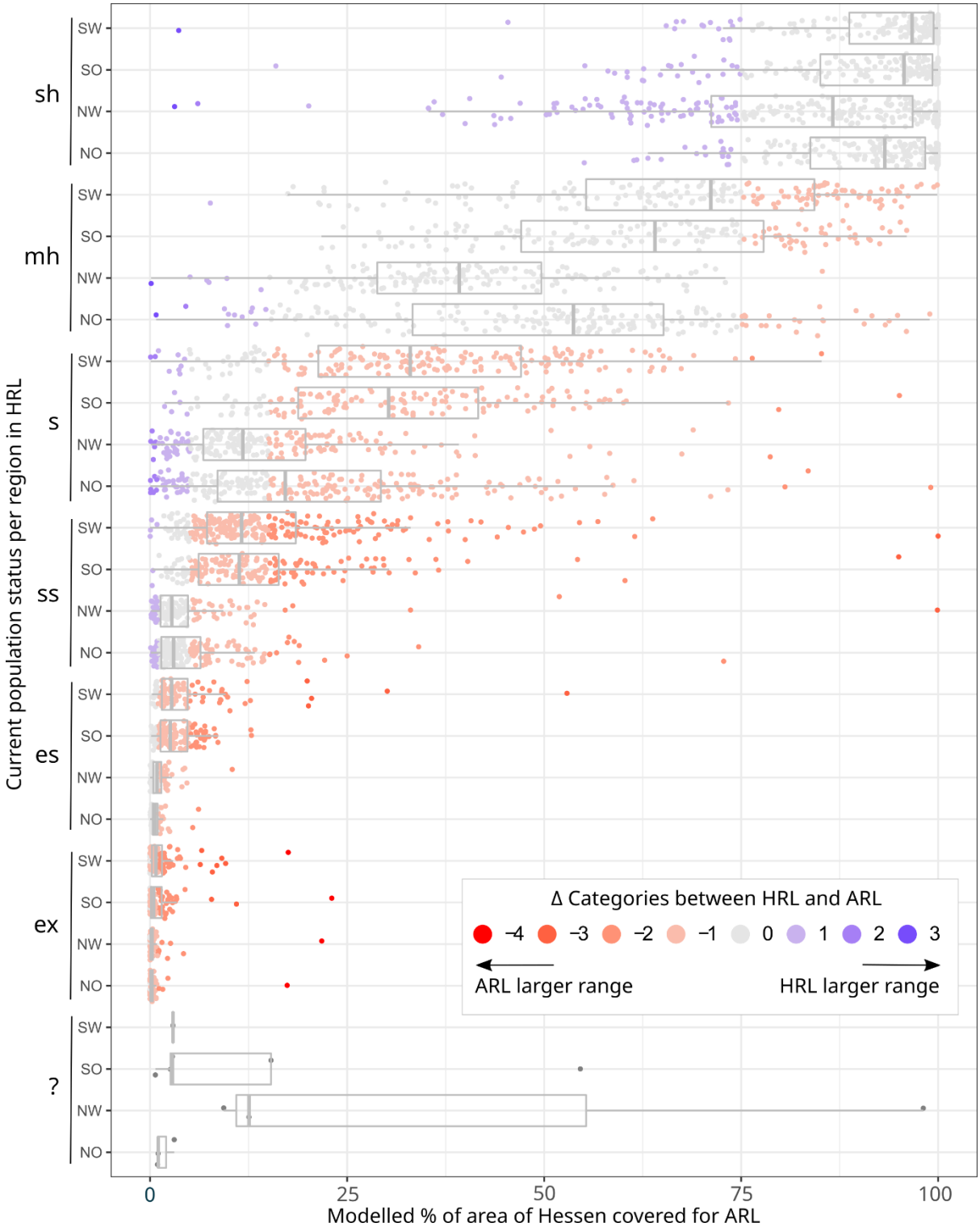

Figure 4. Comparison of species current population status between modelled assessment (ARL) and Hessian Red List (HRL). Each dot represents a species, the colour indicates the difference in current population status categories between ARL and HRL. Supplementary Material S9 contains an interactive Version of this figure, indicating the species names and modelled distribution for each dot. SW = south-west, SO = south-east, NW = north-west, NO = north-east

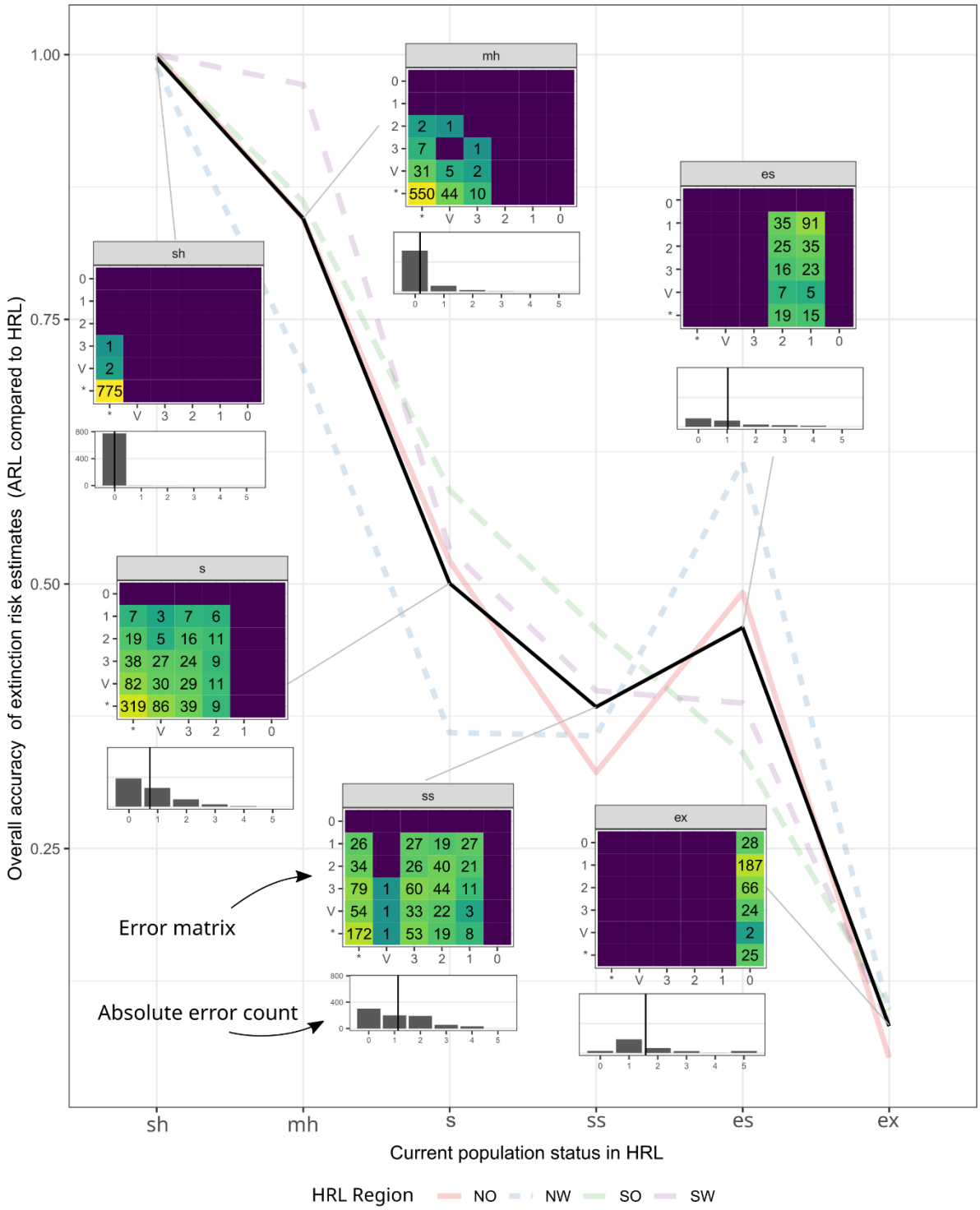

Figure 5. The accuracy of ARL compared to HRL depending on the current population status. Inlets show the confusion matrix for each current population status class (HRL on the x-axis and ARL on the y-axis) and the distribution of the error (by how many categories the assessments disagree in how many cases). Accuracy is high for common species and generally decreases the rare a species becomes.