`iucn_sim` - Improved predictions of future extinctions using IUCN status assessments

Tobias Andermann[1,2], Søren Faurby[1,2], Robert Cooke[1,2], Daniele Silvestro[1,2,*] and Alexandre Antonelli[1,2,3,*]

[1]Department of Biological and Environmental Sciences, Univeristy of Gothenburg, Göteborg, Sweden

[2]Gothenburg Global Biodiversity Centre, Göteborg, Sweden

[3]Royal Botanic Gardens, Kew, Richmond, UK

[*]Joint last authorship

December 16, 2019

Summary

1. The on-going environmental crisis poses an urgent need for predicting future extinction events, which can aid with targeting conservation efforts. Commonly, such predictions are made based on conservation status assessments produced by the International Union for Conservation of Nature (IUCN). However, when researchers apply these conservation status data for predicting future extinctions, important information is often omitted, which can majorly impact the accuracy of these predictions.

2. Here we present `iucn_sim`, a command line program, which implements an improved approach for simulating future extinctions based on IUCN status data. In contrast to previous approaches, `iucn_sim` explicitly models future changes in conservation status for each species, based on information derived from the IUCN assessment history of the last decades. Additionally the program considers generation length information when translating status information into extinction probabilities, as intended per IUCN definition.

3. The program implements a Markov-chain Monte Carlo estimation of extinction rates for each species, based on the simulated extinctions. These estimates inherently contain the chances of conservation status changes and the generation length of each given species.

4. Based on an empirical data example including all birds (class Aves), we find that our improved approach has a strong effect on the estimated species-specific extinction rates as well as on the overall number of predicted extinctions. Using simulated data we show that `iucn_sim` reliably estimates extinction rates with high accuracy if run for a sufficient number of simulations.

Keywords

Aves, Bayesian, Death process, Extinction rate, Extinction risk, Generation length, IUCN, MCMC, Protection status.

Introduction

We are in the middle of a massive biodiversity crisis (Barnosky et al., 2011; Davis et al., 2018; Díaz et al., 2019). Extinction risks have been steadily increasing for as long as we have been keeping record (Ceballos et al., 2015), with no indications of a slow down. It remains therefore vital to predict the number of future extinctions, whether in terms of species, phylogenetic, or functional diversity (Davis et al., 2018; Cooke et al., 2019). An important use of such predictions is to aid conservation prioritization (Mooers et al., 2008). However, all predictions require reliable estimates of extinction risk.

One of the most authoritative global initiatives to quantify extinction risks across animal and plant species is the IUCN Red List (IUCN, 2019), which categorizes the conservation status of organisms based on expert assessments. Since 2001, IUCN has adopted the IUCN v3.1 evaluation system for determining species' conservation statuses (IUCN, 2001). By this standard, extant species are assessed as Least Concern (LC), Near Threatened (NT), Vulnerable (VU), Endangered (EN), or Critically Endangered (CR). If there is not sufficient information available for a species to enable a proper status assessment, the species is categorized as Data Deficient (DD). Species that have not yet been reviewed by IUCN are categorized as Non Evaluated (NE).

IUCN conservation status assessments have been used in numerous scientific studies to infer future biodiversity loss (Cooke et al., 2019; Davis et al., 2018; Faith, 2015; Mooers et al., 2008; Oliveira et al., 2019; Veron et al., 2016). The challenge in this approach is to meaningfully transform the IUCN-defined conservation status categories into explicit extinction probabilities. In these previous studies, researchers have used specific extinction risks, which per IUCN definition are associated with the threatened statuses VU, EN, and CR. Sometimes these risks are also extrapolated to species of the statuses LC and NT (e.g. Davis et al., 2018; Mooers et al., 2008; Veron et al., 2016).

In order for IUCN to decide on assigning a species to one of the threatened categories VU, EN, or CR, this species must meet at least one of five assessment criteria (A-E). One of those criteria (E) is associated with a specific extinction probability, while the other criteria (A-D) mostly encompass estimates of decreasing population trends and fragmentation. The IUCN extinction probability thresholds defined in criterion E are as follows:

- VU: 10% extinction probability within 100 years
- EN: 20% extinction probability within 20 years or 5 generations, whichever is longer (maximum 100 years)
- CR: 50% extinction probability within 10 years or 3 generations, whichever is longer (maximum 100 years)

Even though these extinction probabilities only apply to species assessed under criterion E, they are commonly applied equally to all species sharing the same conservation status (e.g. Davis et al., 2018; Mooers et al., 2008). The underlying assumption that the minimum extinction risks defined for criterion E can be

meaningfully transferred to species listed under one of the other four criteria (A-D) is difficult to test empirically, but is a necessary simplification in order to model the extinction probabilities for the majority of species. However, there are several other important aspects that can be easily incorporated but are commonly neglected when translating IUCN conservation statuses into extinction probabilities.

Neglected information

To the best of our knowledge, there are two key elements that are usually not incorporated when using IUCN data for future extinction predictions: generation length (GL) and expected future conservation status changes.

Generation length is defined as the average turnover rate of breeding individuals in a population (IUCN Standards and Petitions Committee, 2019) and therefore reflects the turnover between generations. Generation length should not be confused with age of sexual maturity, which can be used in the calculation of generation length, but is not equivalent. As per the IUCN definition that we stated above, the extinction probability for the categories EN and CR is to be understood in context of the GL of the given species, if $5 \times$ GL exceeds 20 years for EN species, or if $3 \times$ GL exceeds 10 years for CR species. We argue that including GL data should be the standard practice when modelling extinction risks based on IUCN data, particularly because GL data is readily available for many species (e.g. BirdLife International, 2019; IUCN, 2019; Pacifici et al., 2013) and can be meaningfully approximated through phylogenetic or body mass correlation (Cooke et al., 2018) for species missing GL data.

A second missing element in future predictions, which has not previously been addressed, relates to the fact that IUCN categories are generally treated as static

entities that do not change over time. However, almost two decades of IUCN re-assessments of species (IUCN, 2019), using the IUCN v3.1 standard, have shown that the conservation status of species can change significantly in a relatively short time span, for instance as a result of the effectiveness of conservation efforts. For a species classified as LC, the immediate extinction risk is negligibly small, while for a species classified as CR, the immediate extinction risk is very high. But if we simulate for example 100 years into the future, categories may change due to new or intensified risks or thanks to conservation efforts, which inadvertently will affect the extinction probabilities.

An example of a change in IUCN status is the Pink Pigeon (*Nesoenas mayeri*), which was listed as CR in the 1990's, with only 9 birds remaining, due to habitat loss and predation by introduced species (IUCN, 2019; Swinnerton, 2001). However, following an intensive conservation recovery program, the Pink Pigeon is now listed as VU, with around 470 wild birds (IUCN, 2019). Yet, most species show changes with the opposite trend, for example several species of vultures, which are declining due to poisoning and persecution (Green et al., 2007). There are 22 species of vulture according to the IUCN Red List, 12 of these are classified as threatened (VU, EN or CR), including 9 CR (IUCN, 2019), with sharp declines in population sizes. For instance, the White-headed Vulture (*Trigonoceps occipitalis*), White-backed Vulture (*Gyps africanus*), Hooded Vulture (*Necrosyrtes monachus*) and Rüppell's Vulture (*Gyps rueppelli*) were all listed as LC in 2004 but are now all classified as CR. Information about these changes can be accessed through the IUCN history record and can then be used to inform future simulations.

The `iucn_sim` program

Here we introduce `iucn_sim`, a command-line program that uses available IUCN status assessments of species and generation lengths to simulate 1) future changes in IUCN status, 2) possible times of extinction across species, 3) estimates of species-specific extinction rates for any given set of extant species over a user-defined time span (Fig. 1).

The program, including all software dependencies, is easy to install with a single command (Supplementary Code Sample 1), using the conda package manager (`https://docs.conda.io/en/latest/`). Future simulations are based on extinction risks associated with the current IUCN statuses of the target species, while modeling the possibility of status change informed by the IUCN history of a specified taxonomic group (reference group). The simulator accounts for generation length of the target species, if these data are provided by the user, to properly model the extinction probabilities associated with the IUCN statuses EN and CR. Our simulation approach further allows for modeling DD and NE species for which `iucn_sim` draws new statuses based on the IUCN history data and the current status distribution of the reference group.

The program produces a future diversity trajectory of all input species as well as an overview of the simulated future status distribution (Fig. 2). Further the user can choose to plot individual histograms of simulated extinction times for each species within a specified time frame. Finally `iucn_sim` estimates species-specific extinction rates from the simulated extinction times, using a Markov-chain Monte Carlo algorithm (MCMC).

The `get_rates` function

The purpose of the `get_rates` function (Supplementary Code Sample 2) is to estimate the rates at which species are changing their IUCN status. It incorporates user-provided GL data to calculate the extinction risk for statuses EN and CR, as intended by IUCN definition. These rates are then applied in a subsequent step to simulate future extinctions, while simultaneously modeling potential changes in the IUCN status of species. We note that generation length is only directly involved in the extinction risk for species with statuses EN and CR by IUCN definition, but since our simulation approach incorporates the possibility of changes in IUCN status there will also be a marginal effect of generation length in the extinction risk for species currently assigned to other IUCN categories (see Fig 2a).

There are two main input types the user needs to provide for this function: A) the name of a reference group which will be used to calculate status transition rates and B) the list of target species names for which to simulate future extinctions, including estimates of GL (if available).

Reference group

We model the changes in IUCN status as a stochastic process defined by transition probabilities that quantify the expected number of transitions between any pair of IUCN statuses. The status-transition events are derived from empirical IUCN history data of a user-defined reference group. From these data we estimate annual transition rates between all pairs of IUCN statuses, and use them in simulations to predict future changes.

In order to estimate the status transition rates, the `get_rates` function down-

loads the complete IUCN history (starting at year 2001, to ensure compatibility with the IUCN v3.1 standard) of all species belonging to the reference group, using the `rl_history()` function of the R-package rredlist (Chamberlain, 2017). As reference group, the user can either choose a single taxonomic group, such as the class 'Aves', or a list of taxonomic groups, such as the orders 'Passeriformes' (passerines) and 'Psittaciformes' (parrots), or a list of species names.

Based on the fetched IUCN history data, the `get_rates` function counts all types of status changes that have occurred in the history of the specified group as well as the cumulative amount of time spent in each status across all species (Table 1). The program then estimate the rates of transitions between pairs of statuses using Bayesian sampling. For example, if N_{ij} transitions were observed from status i to status j and the cumulative time spent in i across all species in the reference group is t_i, the program applies a MCMC to sample the annual transition rate q_{ij} from the following posterior:

$$P(q_{ij}|N_{ij}, t_i) \propto P(N_{ij}, t_i|q_{ij}) \times P(q_{ij}) \tag{1}$$

where the log likelihood function is that of a Poisson process describing status change

$$\log P(N_{ij}, t_i|q_{ij}) \propto N_{ij} \log(q_{ij}) - q_{ij} t_i$$

and $P(q_{ij}) \sim \mathcal{U}[0, \infty]$ is a uniform prior on the the transition rate. Posterior samples of the transition rates are then used in the subsequent simulations to predict future status changes while incorporating uncertainties in the rates.

The choice of the reference group is important, because the precision of the

estimated transition rates depends on the available number of empirical transitions (Supplementary Fig. S1). There are two main considerations to make when choosing a reference group: 1) Is the chosen reference group expected to reflect the trends of status change for the species that are being simulated? 2) Does the reference group contain a sufficient number of species so that stochastic effects do not overrule the actual trends for that group?

These two objectives can conflict, for example if the objective is to simulate future extinctions for vultures. In that case using all birds (class Aves) as reference group ($\sim$ 11,000 species) provides a large enough group where several occurrences of each type of status change are being observed in the IUCN history. However, given the notable recent worsening of almost all vulture species‘ conservation status, the trends observed over all birds may not be representative of this group. It is not an analytical requirement to choose a monophyletic clade as a reference group.

As a general guideline we recommend to choose sufficiently large reference groups of more than 1000 species to minimize stochastic effects (see Fig. S1). In the best case (but not necessarily) this group should contain all of the target species.

Target species list and GL data

Besides the reference group that is used for status transition rate estimation, the user also provides a list of target species, which are the species whose future extinctions are being simulated. For all these species, `get_rates` fetches the current IUCN protection status, if available. To translate these categories into explicit extinction probabilities to be used for future simulations, we transformed the ex-

tinction probabilities (E_t) associated with threatened IUCN statuses (see Introduction), defined over specific time frames (t), into annual extinction probabilities (E_1), using the formula provided by (Kindvall & Gärdenfors, 2003):

$$E_1 = 1 - \sqrt[t]{1 - E_t}$$

From these annual extinction probabilities for threatened categories, we extrapolated the annual extinction probabilities for statuses LC and NT by fitting a power function to these points (Appendix 1), estimating the parameters a and b:

$$E_1 = a \times x^b$$

with x representing the index of the IUCN category, sorted by increasing severity (i.e. $x_{LC} = 1$, $x_{NT} = 2$, ..., $x_{CR} = 5$).

To properly model the extinction probabilities linked to the IUCN categories EN and CR for individual species, we strongly encourage users to provide GL estimates for all target species. For species that are lacking GL information, this aspect is disregarded. When ignoring GL information, the extinction risk for species with moderate or long generation times (>3.33 years) will be overestimated (Fig 2), based on the IUCN extinction risk assumptions outlined in the introduction.

The user may provide multiple GL estimates for each species, representing the uncertainty around the GL estimate of each species, in which case `get_rates` will calculate separate extinction probabilities for the statuses EN and CR for each provided GL estimate. In that case each simulation replicate will draw randomly from the produced EN and CR associated extinction probabilities, in order to incorporate the uncertainty surrounding these estimates into the simulations.

The final status transition rates and the species-specific extinction probabilities are exported as text files and are used in the next step to generate q-matrices containing all transition rates and probabilities of extinction (separate q-matrix for each species and simulation replicate). These q-matrices are then used to simulate future extinctions, while simultaneously evolving the IUCN status of all species.

The `run_sim` function

For running the `run_sim` function, the user provides the output of the `get_rates` function and sets the number of years to be simulated into the future, as well as the number of simulation replicates (Supplementary Code Samples 1 and 3). The function will simulate future extinction dates, which are then used to infer averaged extinction rates.

Treating non-assessed species

Before simulating into the future, each species is assigned its current IUCN status as starting status. For all species currently assigned as DD, the function randomly draws a new status at the beginning of each simulation replicate, based on the empirical frequency of the estimated transition rates leading from DD to the statuses LC, NT, VU, EN, or CR. All user-provided species names that cannot be found in the IUCN taxonomy are modeled as NE. For these species a new valid status is randomly drawn based on the frequencies of known IUCN statuses across all species in the list. In each simulation, the function re-initializes the IUCN status of DD and NE species, thus incorporating this uncertainty in the simulation.

Future simulations

The `run_sim` function performs time-forward simulations in which each species can stochastically change based on the following transition matrix, which is populated with the rates obtained from the `get_rates`:

$$Q = \begin{pmatrix} & \textbf{LC} & \textbf{NT} & \textbf{VU} & \textbf{EN} & \textbf{CR} & \textbf{EX} \\ \textbf{LC} & - & q_{LC\to NT} & q_{LC\to VU} & q_{LC\to EN} & q_{LC\to CR} & q_{LC\to EX} \\ \textbf{NT} & q_{NT\to LC} & - & q_{NT\to VU} & q_{NT\to EN} & q_{NT\to CR} & q_{NT\to EX} \\ \textbf{VU} & q_{VU\to LC} & q_{VU\to NT} & - & q_{VU\to EN} & q_{VU\to CR} & q_{VU\to EX} \\ \textbf{EN} & q_{EN\to LC} & q_{EN\to NT} & q_{EN\to VU} & - & q_{EN\to CR} & q_{EN\to EX}(GL) \\ \textbf{CR} & q_{CR\to LC} & q_{CR\to NT} & q_{CR\to VU} & q_{CR\to EN} & - & q_{CR\to EX}(GL) \\ \textbf{EX} & 0 & 0 & 0 & 0 & 0 & - \end{pmatrix}$$

The transitions rates between statuses are sampled from their posterior distribution based on the reference group (Eqn. 1), whereas the extinction rates for each status are assigned based on the IUCN guidelines and using GL information for statuses EN and CR for each species. Rates from EX to any other class are necessarily set to 0, as once extinct species are not allowed to switch back to any of the other IUCN categories.

As we model transitions as a Poisson process, the `run_sim` function generates time-forward simulations for each species based on exponentially distributed waiting times between transition events. For a given current status i the waiting time until the next event is

$$\Delta t \sim \text{Exp}\left(\sum_{j \in S \backslash i} q_{ij}\right)$$

where $S \backslash i$ is the set of statuses excluding the current status i. The type of transition after the waiting time Δt is then sampled randomly with probabilities proportional to the rates in $S \backslash i$. The time-forward simulations are run up to a pre-defined time $t_{\max}$, e.g. 100 years after the starting point.

The function allows the user to simulate different future conservation scenarios. For example one can simulate an increase of conservation efforts by a specific factor. This factor is then applied to all rates in the q-matrix leading to an improvement in conservation status for each species. The user can disable future status changes, which simulates extinctions only based on the current conservation status of each species, equivalent to the approach of Mooers et al. (2008) (Fig. 3).

As output, the function provides a summary of the sampled extinction dates for each taxon and the probability of extinction by the user-provided date. After replicating the simulations multiple times, the function collects for each species a vector of extinction times t_{EX} if $t_{EX} < t_{\max}$ or waiting times of size $t_{\max}$ during which the IUCN status might change without resulting in extinction. These extinction and waiting times are then applied to estimate species-specific annual extinction rates averaged across the time window considered for the simulations.

We note that the actual annual extinction rates can vary over time as a function of changes in the IUCN status, so the extinction rates inferred here are a time-averaged proxy of the process. However, since we do not expect the extinction rate for a given taxon to stay constant over long time periods, particularly when modeling changes in conservation status, we do not advice using `iucn_sim` to

estimate extinction rates spanning across several hundred years or more.

For a given set of extinction times and waiting times simulated for species i, the `run_sim` function uses MCMC to obtain posterior samples of the extinction rate μ_i using the likelihood function of a death process (Silvestro et al., 2019):

$$P(\mathbf{w}|\mu_i) \propto \mu_i^D \times \exp(-\mu_i \sum_{j \in \mathbf{w}} (w_j)) \tag{2}$$

where D is the number of instances in which $w \leq t_{\max}$, i.e. the number of species predicted to go extinct within the considered time window. Posterior estimates of the extinction rates are obtained through MCMC sampling from the posterior distribution:

$$P(\mu_i|\mathbf{w}) \propto P(\mathbf{w}|\mu_i) \times P(\mu_i) \tag{3}$$

where $P(\mu_i)$ is a uniform prior distribution set on the extinction rate $\mathcal{U}[0, \infty]$.

Testing accuracy of rate estimation

Status transition rates

We simulated IUCN status transitions under known rates, in order to test how accurately our program estimates transition rates and what effect the size of the chosen reference group has. Mimicking the empirical IUCN history data, we simulated IUCN status changes over a time period of 20 years for reference groups of 100, 1,000, and 10,000 species. The starting status for each species was drawn randomly, based on the empirical frequencies of the current IUCN status distribution across all birds. To produce realistic transition rates to use for our simulations, we

randomly drew these rates from a uniform range in log-space, ranging between the minimum to the maximum empirical rate estimated for birds. We drew 30 rates to reflect the 30 possible transition types between the six valid IUCN statuses LC, NT, VU, EN, CR, and DD. We then simulated the change of IUCN statuses through time in the same manner as described above for the future simulations for the empirical bird data, with the difference that no extinction events are being modeled.

After the IUCN history for all species was simulated in this manner, we counted the occurrences of each status transition type and estimated the transition rates from these counts, using the `get_rates` function. For comparison we plotted the resulting rate estimates against the true rates that were used to simulate the data (Fig. S1). Based on the results we recommend choosing reference groups of preferably more than 1,000 species, because stochastic fluctuations of status counts below that threshold preclude the estimation of transition rates with any meaningful accuracy, particularly so for low rates.

Extinction rates

We simulated extinction times for 1000 species under known extinction rates, to evaluate the accuracy of the estimated extinction rates produced by the `run_sim` function. The extinction rates (μ) that were used for these simulations were randomly drawn from a uniform range (in log-space) with a minimum and maximum value derived from the annual IUCN extinction risks of the statuses LC and CR, respectively, as modeled in this study. Based on the chosen number of simulation replicates, N extinction time replicates (t_e) were drawn randomly from an exponential distribution with mean $1/\mu$ for each species:

$$t_{EX} \sim \text{Exp}\left(\frac{1}{\mu}\right)$$

This simulation was repeated for 100, 1,000, and 10,000 simulation replicates, in order to test how many replicates are necessary for an accurate rate estimation. The results show that `iucn_sim` estimates extinction rates with high accuracy, yet it requires around 10,000 simulation replicates to ensure this accuracy also for very low rates, as those for species starting as LC (Fig. 4).

Empirical data example

We ran `iucn_sim` to estimate future extinction events for all birds over the next 100 years.

Generating GL estimates

As an underlying taxonomy we downloaded species lists of all extant bird species from IUCN v2019-2 (IUCN, 2019), with the R-package rredlist (Chamberlain, 2017). Generation length data for the majority of these species was provided by BirdLife International (http://www.birdlife.org). For all remaining species we modeled GL estimates using multivariate phylogenetic imputation under the assumption that GL has a phylogenetic correlation and is also correlated with body mass. Body mass data was downloaded from Cooke et al. (2019), which is based on data from the databases EltonTraits (Wilman et al., 2014) and the Amniote Life History Database (Myhrvold et al., 2015). To obtain phylogenies we downloaded 1000 samples of the posterior species trees distribution produced by Jetz et al. (2012), based on the Ericson backbone ("EricsonStage2_0001_1000.zip").

A fraction of 90% of bird species names listed in IUCN v2019-2 were also present in the phylogenies. After taxonomic revision we matched 96% of all IUCN bird species with the tips in the phylogenies.

To estimate GL values for all species lacking such data, we ran a phylogenetic imputation, using the R-package rphylopars (Goolsby et al., 2017). To determine the best model we calculated the AIC score for all available models (Supplementary Fig. S2) and chose 'EB' as the best model based on the AIC results. In order to incorporate the uncertainty of the phylogenetic estimates, we ran separate imputations for 100 randomly selected trees from the downloaded species tree distribution. We exported the 100 resulting mean values of the GL estimates for each species.

For all remaining species that were not present in the phylogeny we modelled the GL value to be the mean of the encompassing genus, calculated separately for each of the 100 GL data replicates. This resulted in our final dataset containing GL estimates for all bird species listed by IUCN v2019-2. The GL estimates for all birds as well as those for other groups are available on the project's GitHub page.

Running `iucn_sim`

We provided the list of IUCN bird species names and the 100 GL estimates for each species as input for `get_rates` (Supplementary Code sample 1). As reference group we used the whole class Aves ($\sim$ 11,000 species). Table 1 shows the counted empirical occurrences of each status transition type within the IUCN history of birds. The transition rates estimated from these counts can be found in the Supplementary Data.

We used these transition rate estimates and the GL-informed extinction probabilities calculated by the `get_rates` function to run 10,000 future simulations for the next 100 years for all birds, using the `run_sim` function (Supplementary Code sample 1). Figure 2 shows the resulting simulated diversity trajectory and status distribution for the next 100 years, with a predicted mean of 737 bird species losses (95% credibility interval: 680 to 799 species). The resulting simulated extinction probabilities and estimated extinction rates for all bird species can be found in the Supplementary Data.

Our empirical results show that accounting for GL decreases the resulting extinction rate estimates (Fig. 3). As an example we highlight this effect for the Red-headed vulture (*Sarcogyps calvus*), which is categorized as CR and has a relatively long generation length of 15 years (Fig. 3b). This effect on CR species with long GL times is expected since the immediate extinction probability applied in the simulations for EN and CR species decreases when incorporating the GL information, according to IUCN definition (see Introduction). But also for LC species, as highlighted for the Turkey vulture (*Cathartes aura*, GL = 9.9 years), a small decreasing effect of GL data incorporation can be seen in the extinction rate estimates, since occasionally these species will change to the categories EN or CR in the future simulations, when allowing for future status changes (Fig. 3a). Overall, accounting for GL data leads to a decrease in the number of predicted extinctions across the whole target group (birds in our example, see Supplementary Fig. S3).

The effect of modeling future status changes can vary and can lead to an increase or decrease in the estimated extinction rates for a given species. The strength and direction of this effect depends on the estimated status transition

rates and is therefore expected to change depending on the chosen reference group. However, for LC species this usually leads to an increase in the estimated extinction rates (Fig. 3c), because these species can only change to a more threatened status (LC being the least threatened status). Similarly for CR species the effect of modeling future status changes typically leads to a decrease in extinction rates (Fig. 3d), since species can only switch to less threatened categories in the future (CR being the most threatened status). Overall, modeling future status changes leads to a sharp increase in the number of predicted extinctions across the whole target group (Fig. S3).

Conclusions

To summarize, the incorporation of both GL and future status changes increases the biological credibility of the resulting extinction rate estimates for individual species, as well as that of the estimated number of species extinctions for the whole target group. It is therefore strongly advisable to include these two factors when producing future extinction predictions based on IUCN status information and it should be adopted as standard practice, particularly for groups with a well covered IUCN record and with available GL data. The source code of our program `iucn_sim` is available on GitHub (`https://github.com/tobiashofmann88/iucn_extinction_simulator`) and is open for contributions and feedback from users, leading to the incorporation of further improvements for predicting future extinctions. Future additions to the program could for example include more specific future modeling of species based on similarities in biological traits, geographic location, or niche space.

Table 1: Status transitions counted in the IUCN history of birds (class Aves). For example, the empirical count of transitions from status LC to NT is 176, while the count of transitions from NT to LC is 100.

	LC	NT	VU	EN	CR	DD
LC	0	176	74	18	3	1
NT	100	0	71	22	3	1
VU	14	76	0	95	13	1
EN	1	10	63	0	47	0
CR	0	2	9	41	0	0
DD	9	10	5	2	0	0

References

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O.U., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., Mersey, B. & Ferrer, E.A. (2011) Has the Earth's sixth mass extinction already arrived? Nature, **471**, 51–57.

BirdLife International (2019) BirdLife Data Zone. Available at http://datazone.birdlife.org/home.

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M. & Palmer, T.M. (2015) Accelerated modern human – induced species losses: entering the sixth mass extinction. Sciences Advances, **1**, 1–5.

Chamberlain, S. (2017) Rredlist: 'IUCN' Red List Client.

Cooke, R.S.C., Eigenbrod, F. & Bates, A.E. (2019) Projected losses of global mammal and bird ecological strategies. Nature Communications, **10**, 2279.

Cooke, R.S.C., Gilbert, T.C., Riordan, P. & Mallon, D. (2018) Improving generation length estimates for the IUCN Red List. PLOS ONE, **13**, e0191770.

Davis, M., Faurby, S. & Svenning, J.C. (2018) Mammal diversity will take millions of years to recover from the current biodiversity crisis. Proceedings of the National Academy of Sciences of the United States of America, **115**, 11262–11267.

Díaz, S., Settele, J., Brondízio, E., Ngo, H.T., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K.A., Butchart, S.H.M., Chan, K.M.A., Garibaldi, L.A., Liu, K.I.J., Subramanian, S.M., Midgley, G.F., Miloslavich, P., Molnár,

Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., Chowdhury, R.R., Shin, Y.J., Visseren-Hamakers, I.J., Willis, K.J. & Zayas, C.N. (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services.

Faith, D.P. (2015) Phylogenetic diversity, functional trait diversity and extinction: avoiding tipping points and worst-case losses. Philosophical Transactions of the Royal Society B: Biological Sciences, **370**, 20140011.

Goolsby, E.W., Bruggeman, J. & Ané, C. (2017) Rphylopars : fast multivariate phylogenetic comparative methods for missing data and within-species variation. Methods in Ecology and Evolution, **8**, 22–27.

Green, R.E., Taggart, M.A., Senacha, K.R., Raghavan, B., Pain, D.J., Jhala, Y. & Cuthbert, R. (2007) Rate of Decline of the Oriental White-Backed Vulture Population in India Estimated from a Survey of Diclofenac Residues in Carcasses of Ungulates. PLoS ONE, **2**, e686.

IUCN (2001) IUCN Red List Categories and Criteria: Version 3.1. Available at https://www.iucn.org/content/iucn-red-list-categories-and-criteria-version-31.

IUCN (2019) IUCN Red List of threatened species v2019-2. Available at http://www.iucnredlist.org.

IUCN Standards and Petitions Committee (2019) Guidelines for Using the IUCN Red List Categories and Criteria. Version 14. Available at http://www.iucnredlist.org/documents/RedListGuidelines.pdf.

Jetz, W., Thomas, G., Joy, J., Hartmann, K. & Mooers, A. (2012) The global diversity of birds in space and time. Nature, **491**, 444–448.

Kindvall, O. & Gärdenfors, U. (2003) Temporal Extrapolation of PVA Results in Relation to the IUCN Red List Criterion E. Conservation Biology, **17**, 316–321.

Mooers, A.Ø., Faith, D.P. & Maddison, W.P. (2008) Converting endangered species categories to probabilities of extinction for phylogenetic conservation prioritization. PLoS ONE, **3**, 1–5.

Myhrvold, N.P., Baldridge, E., Chan, B., Sivam, D., Freeman, D.L. & Ernest, S.K.M. (2015) An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. Ecology, **96**, 3109–000.

Oliveira, B.F., Sheffers, B.R. & Costa, G.C. (2019) Decoupled erosion of amphibians' phylogenetic and functional diversity due to extinction. Global Ecology and Biogeography, p. geb.13031.

Pacifici, M., Santini, L., Di Marco, M., Baisero, D., Francucci, L., Grottolo Marasini, G., Visconti, P. & Rondinini, C. (2013) Generation length for mammals. Nature Conservation, **5**, 89–94.

Silvestro, D., Salamin, N., Antonelli, A. & Meyer, X. (2019) Improved estimation of macroevolutionary rates from fossil data using a Bayesian framework. Paleobiology, **45**, 546–570.

Swinnerton, K.J. (2001) The ecology and conservation of the pink pigeon Columba mayeri in Mauritius. Ph.D. thesis, University of Kent at Canterbury.

Veron, S., Penone, C., Clergeau, P., Costa, G.C., Oliveira, B.F., São-Pedro, V.A. & Pavoine, S. (2016) Integrating data-deficient species in analyses of evolutionary history loss. Ecology and Evolution, **6**, 8502–8514.

Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M.M. & Jetz, W. (2014) EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. Ecology, **95**, 2027–2027.

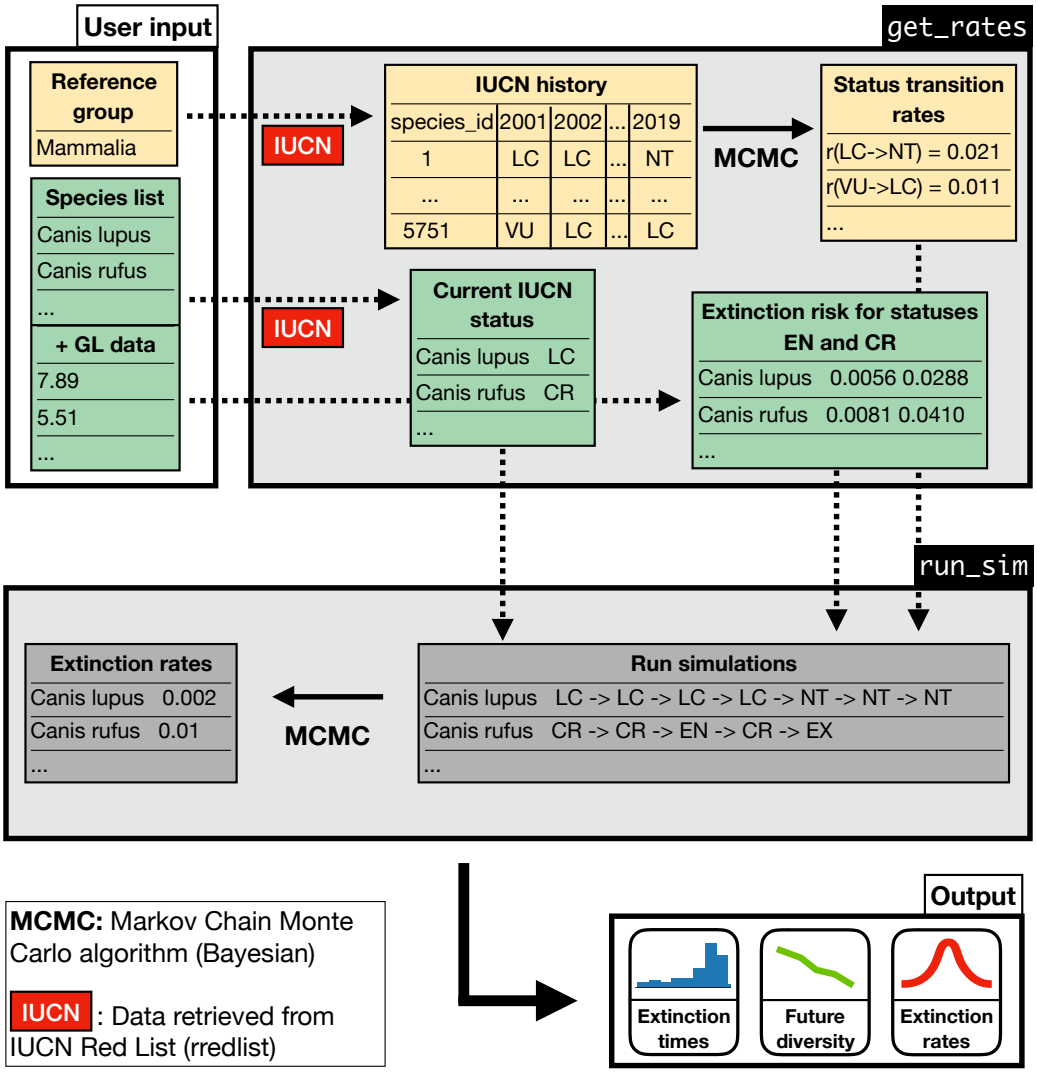


Figure 1: Workflow of `iucn_sim`. The user defines a reference group for status transition rate estimation, as well as a list of target species for which future extinctions and status changes will be simulated. Optionally the user is encouraged to also provide GL estimates for each target species, which are applied in calculating the extinction risks associated with the statuses EN and CR. The current conservation status of all species is determined, using available IUCN information. All of these steps take place within the `get_rates` function, as indicated by the grey box in the top right of the figure. The estimated transition rates, calculated extinction risks, and current status distribution of all target species is parsed on into the `run_sim` function. Next, these data are applied to simulate future status changes and extinctions. Finally extinction rates are estimated from the simulation output and various summary statistics and plots are being produced as output.

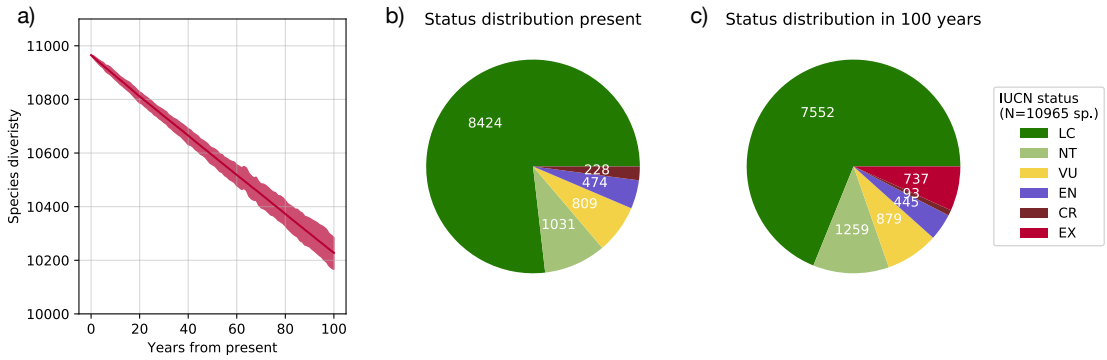


Figure 2: Future diversity trajectory and status distribution for birds. Panel a) shows the future diversity trajectory of the next 100 years for birds, based on future extinctions simulated with `iucn_sim`. The pie-charts show the IUCN status distribution at the beginning (b) and the end (c) of the simulations. The simulations included body mass data for all species and we allowed for future status changes.

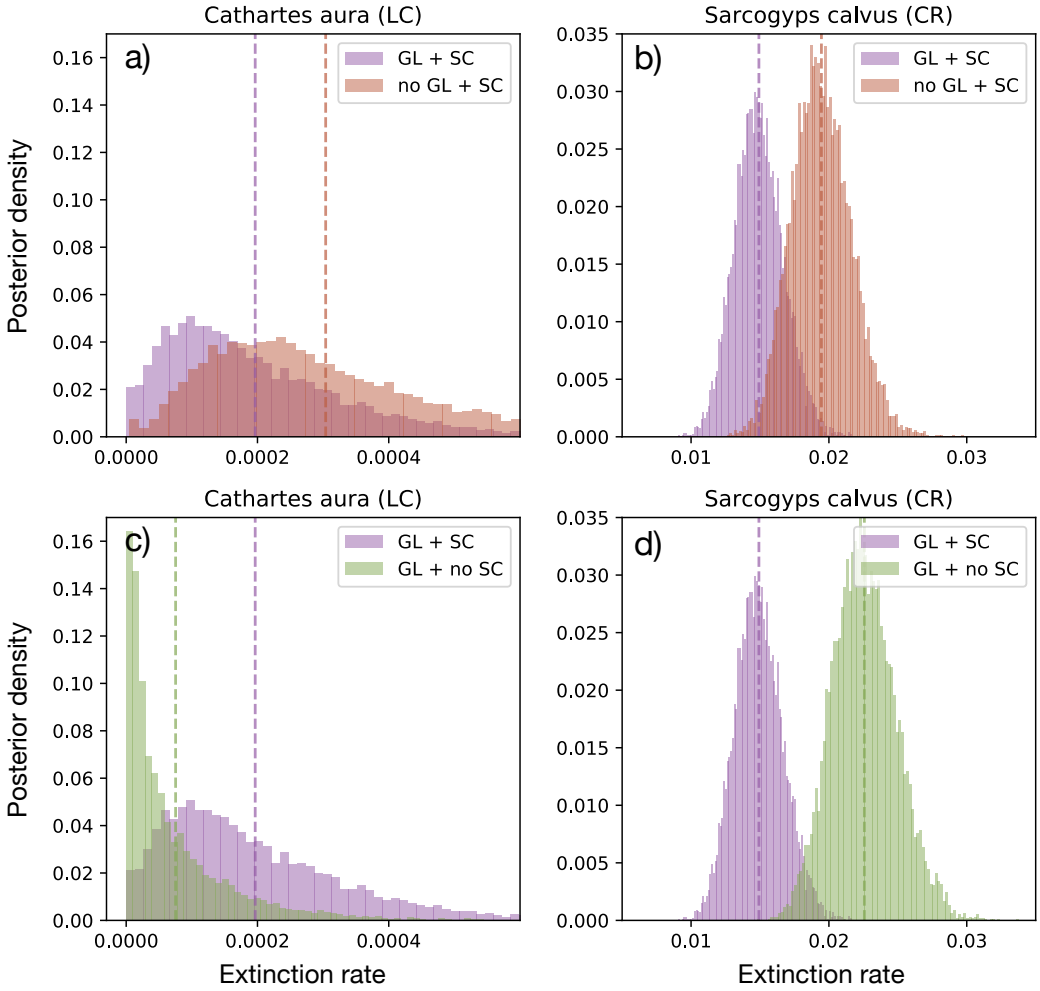
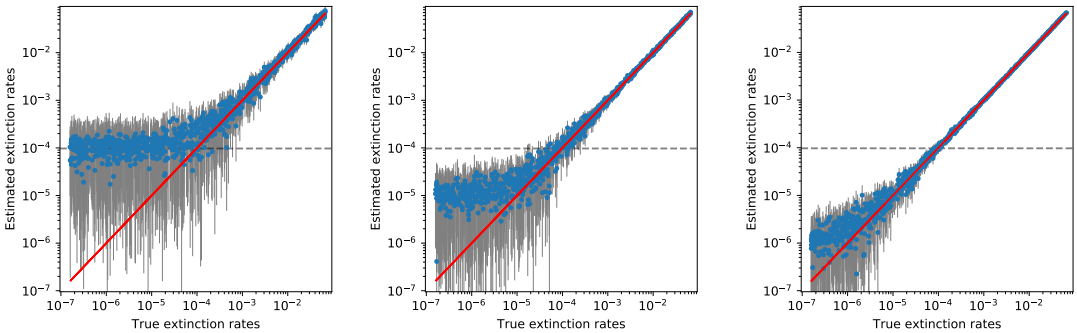


Figure 3: The effect of generation length (GL) and status-change (SC) on estimated extinction rates. The plots show histograms of the posterior density of extinction rates estimated with `iucn_sim` for two different species: the Turkey vulture (*Cathartes aura*, GL = 9.9 years, Least Concern), panels a) and c); and the Red-headed vulture (*Sarcogyps calvus*, GL = 15 years, Critically Endangered), panels b) and d). Upper panels show that the extinction rate estimates slightly decrease when including GL data into the simulations (purple) compared to ignoring GL data (red) for both LC and CR species. Bottom panels show that modeling future status changes slightly increases the extinction rate of LC species, but leads to a decrease for CR species (d). Note that the effect of future status changes on extinction rates depends on the estimated status transition rates and is therefore expected to change depending on the chosen reference group.



(a) N=100 (b) N=1,000 (c) N=10,000

Figure 4: Increasing precision and accuracy of extinction rate estimates with more simulation replicates. We plotted the true extinction rates that were used to simulate extinction times for 1000 putative species (x-axis) against the extinction rates estimated with the `run_sim` function (y-axis). We then ran three analyses with (a) 100, (b) 1,000, and (c) 10,000 simulation replicates. The plots show the mean values (blue dots) and the 95% credible interval (grey vertical lines). The dotted horizontal line shows the minimum extinction rate estimate based on the empirical dataset for all birds (10,000 simulation replicates). Extinction rates below this line are therefore unlikely to occur in empirical data sets. The diagonal red line shows a theoretical perfect correlation for reference.